Impact of Transportation on Air Quality at Elementary and Middle Schools in South Carolina

Final Report

by

Gurcan Comert, Benedict College
Samuel Darko, Benedict College
Nathan Huynh, University of South Carolina
Judith Mwakalonge, South Carolina State University
Bright Elijah, Benedict College
Quentin Eloise, Benedict College

**Contact information**
Gurcan Comert, Ph.D.
1600 Harden St., Columbia, SC 29204
Benedict College
Phone: (803) 705-4803; _E-mail:_ gurcan.comert@benedict.edu

November 2019

Center for Connected Multimodal Mobility (C2M2)

200 Lowry Hall, Clemson University
Clemson, SC 29634

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by the Center for Connected Multimodal Mobility (C^2M^2) (Tier 1 University Transportation Center) Grant, which is headquartered at Clemson University, Clemson, South Carolina, USA, from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Non-exclusive rights are retained by the U.S. DOT.

ACKNOWLEDGMENT

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
Impact of Transportation on Air Quality at Elementary and Middle Schools in South Carolina	November 2019
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Gurcan Comert, Ph.D.; ORCID: 0000-0002-2373-5013 Samuel Darko, Ph.D.: ORCID: 0000-0002-3923-2020 Nathan Huynh, Ph.D.; ORCID: 0000-0002-4605-5651 Judith Mwakalonge, Ph.D.; ORCID: 0000-0002-7497-6829 Bright Elijah; ORCID: 0000-0002-2035-3589 Quentin Eloise; ORCID: 0000-0003-0939-0159	

9. Performing Organization Name and Address	10. Work Unit No.
Department of Computer Science, Physics, and Engineering Benedict College 1600 Harden St., Columbia, SC 29206	
	11. Contract or Grant No.
	69A3551747117

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Center for Connected Multimodal Mobility (C²M²) USDOT Tier 1 University Transportation Center Clemson University 200 Lowry Hall, Clemson Clemson, SC 29634	Final Report (09/20/2017 - 6/31/2019)
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Many studies have reported associations between respiratory symptoms and resident proximity to roadway traffic. However, only a few have documented the relationship between traffic volume and air quality in local areas. This study investigates the impact of traffic volume on air quality at different geographical locations in the state of South Carolina using multilevel linear mixed models and Grey Systems. Historical traffic volume and air quality data between 2006 and 2016 are obtained from the South Carolina Department of Transportation (SCDOT) and the United States Environmental Protection Agency (EPA) monitoring stations. These data are used to develop prediction models that relate Air Quality Index (AQI) to traffic volume for selected counties and schools. For the selected counties, two models were developed: one with Ozone (O_3) and one with particulate matter ($PM_{2.5}$) as the dependent variable. For the schools, one model is developed with Ozone (O_3) as the dependent variable. The number of counties and schools studied are limited by the availability of air monitoring stations dedicated to measuring O_3 and $PM_{2.5}$. Several types of models are investigated. They include linear regression model (LM), linear mixed-effect regression model (LMER), Grey Systems (GM), error-corrected GM (EGM), Grey Verhulst (GV), error-corrected GV (EGV), and LMER combined with EGM. The LM model produced the least accurate estimate while the LMER combined with the EGM model produced the most accurate estimate (average RMSE is less than 5%). The models' estimates suggest that air quality in South Carolina will continue to get worse in the coming years due to increasing annual average daily traffic (AADT). An interesting finding of this study is that some counties and schools will have higher levels of O_3 or $PM_{2.5}$ when AADT decreases, which suggests that there are additional factors other than AADT that influence the air quality in these counties and schools.

17. Keywords	18. Distribution Statement
Air Quality; Grey Systems; Air Quality Index (AQI); linear mixed-effect regression model.	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	34	NA

Center for Connected Multimodal Mobility (C²M²)

Clemson University, Benedict College, The Citadel, South Carolina State University, University of South Carolina
Page | iv

Table of Contents

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY[1]

Studies have indicated vehicle emissions as a primary source of ambient air pollutants in urban areas. Over the past decade, traffic volume has been observed to be steadily rising without any sign of decline. Previous studies have established associations between respiratory diseases and/or symptoms, such as asthma with residential proximity to major roads with high traffic volume. Many studies have reported associations between respiratory symptoms and resident proximity to traffic. However, only a few have documented information about the relationship between traffic volume and air quality in local areas. Therefore, real-time monitoring of traffic generated air pollution is necessary to evaluate its impact. With the advent of connected and autonomous vehicles, real-time data and computationally efficient statistical models, better decision-making tools can be developed to minimize exposure risks. Connected vehicle mobility applications go hand in hand with environmental applications as improving travel time and reliability for multiple modes will lead to fuel efficiency and reduce vehicle delay/idle time.

This project aimed to develop models that address transportation impact on air quality at school zones from connected sensors (roadside or onboard). Experiments and data collected from connected systems would be used in intelligent transportation system applications, such as signal control and freeway traffic monitoring. The objectives were to (1) develop a system that can provide real-time emissions as well as traffic data, (2) develop models that accurately explains air quality versus traffic for different locations including schools, and (3) develop methods for estimating emissions in arterials and freeway corridors through connected vehicles and infrastructures.

The project investigated the effectiveness of low-cost air quality sensors in traffic applications. The study included set up, calibration, and data collection at an elementary school and a college campus. Based on the experiments and collected data, the following observations can be made for the usage of sensors and the impact of traffic on air quality:

1. Outdoor equipment's operation range should be carefully selected for reliable data collection.
2. Sensors may fail and may not record data for a period of time.
3. Calibration should be done for a longer period considering data may not be reported from an EPA monitoring site.
4. Equipment can be used with portable power sources and simply installed on a pole. Humidity and temperature levels should be carefully considered. The performance of the outdoor version of the sensors is better.
5. Sensors are inexpensive and are able to provide various data related to pollutants at every 5 seconds, thus, they can be used within various intelligent transportation systems real-time applications.
6. Collected data and analysis suggest that air quality impacted more by temperature. The effect of traffic may not have been captured due to fixed sensor location (dissipation impact) and the sensors' lack of sensitivity.

The project also investigated the impact of traffic volume on air quality at different geographical locations in the state of South Carolina using multilevel linear mixed models and Grey Systems.

Historical traffic volume and air quality data between 2006 and 2016 were obtained from the South Carolina Department of Transportation (SCDOT) and the United States Environmental Protection Agency (EPA) monitoring stations. The data was used to develop prediction models that relate Air Quality Index (AQI) to traffic volume for selected counties and schools. For the counties, two models are developed: one with Ozone (O_3) and one with $PM_{2.5}$ as the dependent variable. For the schools, only one model is developed with Ozone (O_3) as the dependent variable. The number of counties and schools studied are limited by the availability of air monitoring stations dedicated to measuring O_3 and $PM_{2.5}$.

Several types of models were investigated. They include linear regression model (LM), linear mixed-effect regression model (LMER), Grey Systems (GM), error-corrected GM (EGM), Grey Verhulst (GV), error-corrected GV (EGV), and LMER combined with EGM. The LM model produced the least accurate estimate while the LMER combined with the EGM model produced the most accurate estimate (average RMSE is less than 5%). The models' estimates suggest that air quality in South Carolina will continue to get worse in the coming years due to increasing annual average daily traffic (AADT). An interesting finding of this study is that some counties and schools will have higher levels of O_3 or $PM_{2.5}$ when AADT decreases. This finding suggests that there are additional factors, other than AADT, which influence the air quality in these counties and schools.

CHAPTER 1

Introduction and Background

Studies have indicated vehicle emissions as a primary source of ambient air pollutants in urban areas. Over the past decade, traffic volume has been observed to be steadily rising without any sign of decline. Therefore, the need for deliberate, continuous monitoring and systematic studies on the effects of traffic generated pollution at schools is needed. With the advent of connected and autonomous vehicles, real-time data, and computationally efficient statistical models, better decision-making tools can be developed to minimize exposure risks.

Previous studies have established associations between respiratory diseases and/or symptoms such as the prevalence of asthma in residential areas close to major roads with high traffic volume (Gauderman et al. (2005); McConnell et al. (2010)). Studies have also shown higher rates of morbidity and mortality for drivers, commuters as well as individuals living near major roadways (e.g., Wjst et al. (1993); Zhang and Batterman (2013)). Exposure to traffic-related air pollution has been linked to a variety of short-term and long-term health effects, including asthma, reduced lung function, impaired lung development in children, and negative cardiovascular effects in adults, as well as negative influence on academic performance (Brunekreef et al. (1997); Rakowska et al. (2014)). The exposure of children to traffic-related air pollution while at school is a growing concern because many schools are located near heavily traveled roadways (e.g., Janssen et al. (2003, 2001); Mohai et al. (2011); Adams and Requia (2017); Mohammadyan et al. (2017)). Pollutants such as ozone (O_3) and $PM_{2.5}$ are known to cause serious respiratory defects (Guarnieri and Balmes (2014)). Ground ozone (O_3) is formed when Nitrogen Dioxide (NO_2) reacts with Volatile organic compounds (*VOC)* in the presence of heat from sunlight. $PM_{2.5}$ is composed of particulate matter with a diameter of 2.5 micrometers (*µms*) or smaller.

To date, only a few studies have investigated the relationship between traffic volume and AQI. To this end, this study aims to develop predictive models that relate air quality in the form of Air Quality Index (AQI) to traffic volume, specifically, the annual average daily traffic (AADT). AQI is a numeric value ranging from 0 to 500 used for reporting daily air quality. An AQI value of 50 or below represents good air quality. It should be noted that in this study we are assessing the impact of traffic volume on air quality at a macroscopic level. This approach is similar to the work by de Miranda et al. (2017) who studied the relationship between black carbon and heavy traffic in Sao Paulo, Brazil and Hao et al. (2018) who evaluated the environmental impact of traffic congestion. Alternatively, air quality or emissions can at a microscopic level by using a traffic microsimulation software such as VISSIM and the U.S. EPA MOVES model. Examples of such studies include the work of Xie et al. (2012) who used PARAMICS and MOVES to develop an integrated model for reliable estimation of daily vehicle fuel savings and emissions. Similarly, AbouSenna et al. (2013) used VISSIM and MOVES to predict emissions from vehicles on a limited-access highway, Xu et al. (2016) who developed a tool to combine VISSIM and MOVES to estimate vehicle emissions for a corridor or network and Shaaban et al. (2019) who used VISSIM and MOVES to assess the impact of converting roundabouts to traffic signals on vehicle emissions along an urban arterial. The EPA MOVES model uses the Vehicle Specific Power (VSP) framework to characterize modal emission rates. VSP was first developed by Jimenez-Palacios (1998). This framework allows MOVES to be applied to any transportation network (as long as VSP data are available), including those outside the U.S. The MOVES model has been used in other countries such as China, India, Mexico, Qatar, and Brazil. The models are developed using the traffic data from 19 South Carolina counties that are selected based upon the availability of EPA air monitoring stations.

Connected Vehicles (CV) application of "low emission zones" would present unique opportunities for better decision making for medium to long term planning. This study will provide crucial real-time emissions data at different locations in South Carolina. This project studies the correlation of the impact of traffic in terms of AADTs on AQIs in South Carolina. Such models can be used by various agencies, urban planners, and developers to identify suitable locations for K-12 schools and hospitals and to generate environmental policies. For example, in Atlanta Georgia, the Clean Air Act requires areas with poor air quality (non-attainment areas) to have transportation plans that are consistent with air quality goals and standards (Howitt and Moore (1999); Hallmark et al. (2000)).

In this project, the Grey models based on the Grey System theory are utilized and they are compared against regression models. This approach is utilized because it is known to be capable of handling datasets with missing independent variables (Liu et al. (2010)). Additionally, Grey models can be used to model systems that are non-stationary and nonlinear. The performance of Grey models against back propagation neural network (NN) and radial basis function was evaluated by An et al. (2012), and the authors found that the Grey model performed better in predicting monthly average daily traffic volume. Similarly, Gao et al. (2010) found that Grey models outperformed support vector machine (SVM) and artificial NN models in predicting average hourly volumes. Compared to NN and SVM, Grey models can handle a low sample size and do not require as much computational power. This study is the first to apply Grey models to predict emissions.

The remainder of this report is organized as follows. Chapter 2 provides a description of the data. Chapters 3 and 4 discuss the modeling techniques used in the study, which are: multiple linear regression, multilevel linear regression, and Grey Systems. Chapter 5 presents numerical experiments for model validation. Lastly, Chapter 6 provides concluding remarks and future research directions.

CHAPTER 2

Data Analysis

2.1 Macro Data

The data used in this study are obtained from the South Carolina Department of Transportation (SCDOT) and the United States Environmental Protection Agency (USEPA) websites. Figure 1 shows the locations of EPA monitoring stations located throughout the state of South Carolina. In developing the county-level models, data from all monitoring stations are used. For the school-level models, only those schools with nearby EPA monitoring stations and those that are adjacent to major roadways with high traffic volume are considered. Only 7 schools in South Carolina met these criteria.
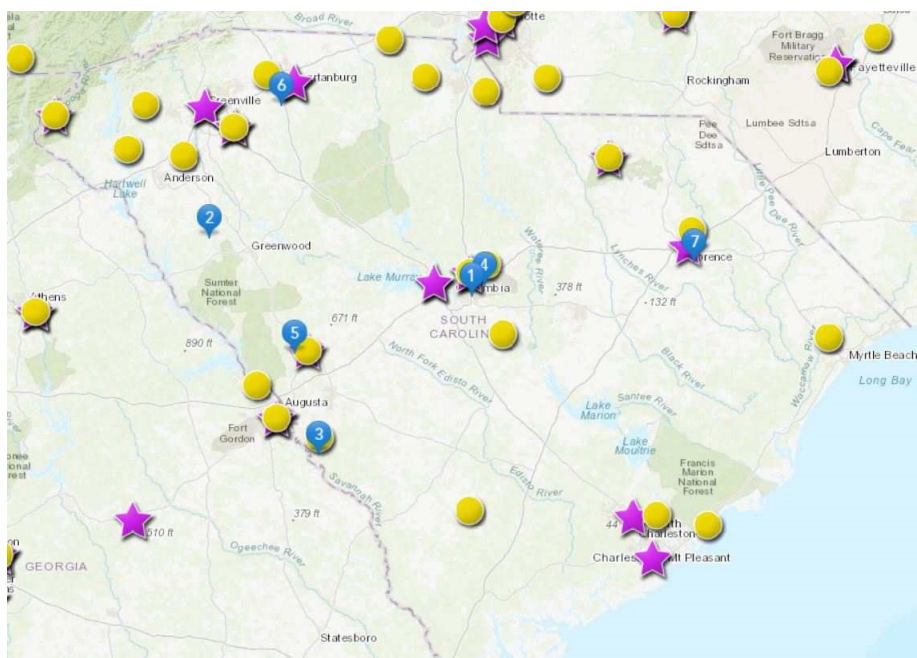
Figure 1 EPA Stations (star = PM$_{2.5}$, circle = O$_3$) and nearby schools (numbered)

Table 1 shows the emissions and AADT data obtained for 19 South Carolina counties and selected schools in 2006. Similar data were obtained up to 2016, for a total of 11 years. The datasets from EPA tends to contain missing data. To deal with this issue, missing data are imputed with approximate Bayesian inference using R-package (Gelman et al. 2015). Figure 2 shows the utilized dataset before and after the missing data imputation. The black regions represent missing data that were subsequently imputed. Figure 2 presents standardized values via the transformation of $((x - \mu_x)/2\sigma_x)$ of the observations.

Table 1 O_3 and $PM_{2.5}$ AQIs with traffic counts for different counties and schools

Year.		O_3	$PM_{2.5}$	Avg. AADT	County	Missing O_3	Missing $PM_2.5$
2006		56	40	1561	Abbeville	FALSE	TRUE
2006		53	52	5573	Aiken	FALSE	FALSE
2006		42	43	5280	Anderson	FALSE	TRUE
2006		48	53	2247	Barnwell	FALSE	TRUE
2006		41	42	7429	Beaufort	TRUE	FALSE
2006		38	59	8959	Berkeley	FALSE	TRUE
2006		48	48	14842	Charleston	FALSE	FALSE
2006		45	50	5108	Cherokee	FALSE	TRUE
2006		49	30	3448	Chester	FALSE	TRUE
2006		45	51	2226	Chesterfield	FALSE	FALSE
2006		45	53	3823	Colleton	FALSE	TRUE
2006		54	57	3017	Darlington	FALSE	TRUE
2006		39	49	1768	Edgefield	FALSE	FALSE
2006		65	47	6550	Florence	TRUE	FALSE
2006		45	47	4632	Georgetown	TRUE	FALSE
2006		45	59	9759	Greenville	TRUE	FALSE
2006		37	44	9382	Horry	TRUE	FALSE
2006		48	55	8946	Lexington	TRUE	FALSE
2006		43	41	3836	Oconee	FALSE	FALSE
2006		53	38	5023	Pickens	FALSE	TRUE
2006		52	53	11772	Richland	FALSE	FALSE
2006		63	52	7168	Spartanburg	FALSE	FALSE
2006		53	22	2218	Union	FALSE	TRUE
2006		45	43	1763	Williamsburg	FALSE	TRUE
2006		47	52	7237	York	FALSE	TRUE

Year	School	O_3	Avg. AADT
2006	(1)Dent Middle School	48	16900
2006	(2)Dixie High School	56	550
2006	(3)Jackson Middle School	53	4283
2006	(4)Spring Valley High School	59	22150
2006	(5)WE Parker Elementary School	39	900
2006	(6)Westgate Christian School	63	3525
2006	(7)Wilson High School	54	4267
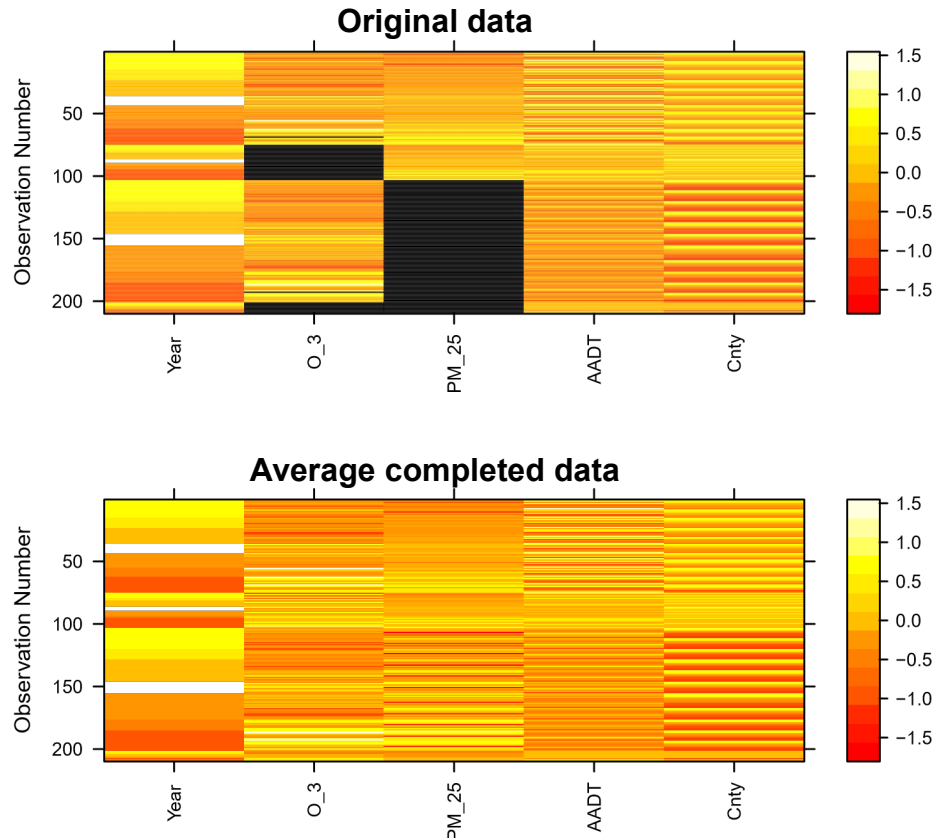
Figure 2 Imputation of missing data using Gaussian Mixtures

In Figure 3, average O_3 and $PM_{2.5}$ measurements for multiple years are shown. It can be seen that ozone levels can be expressed as a multilevel model with different coefficients for each county, and it can also be expressed as a single-parameter model with a covariance matrix of counties. Note that these emission values are averaged annually and they are assumed to be representative of the air quality level over entire counties and schools.
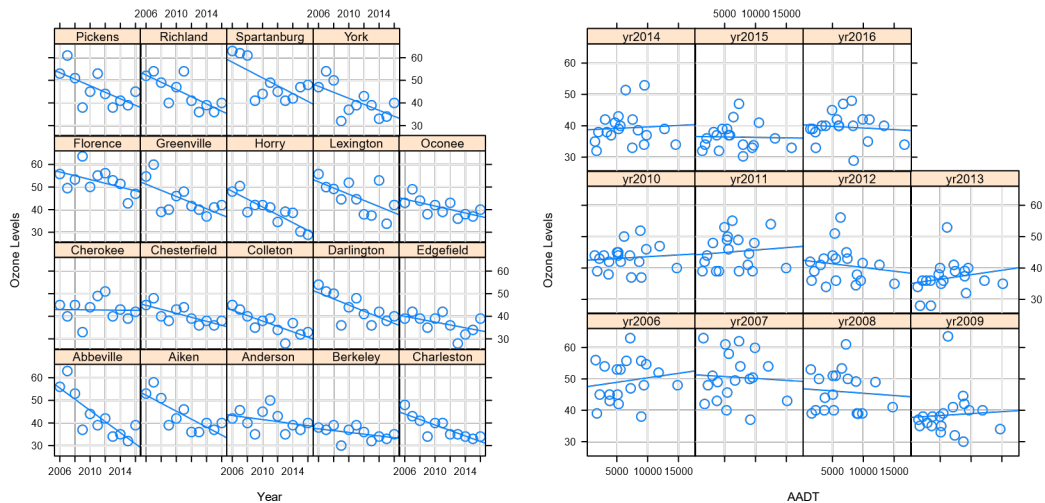
(a) O₃ levels for counties
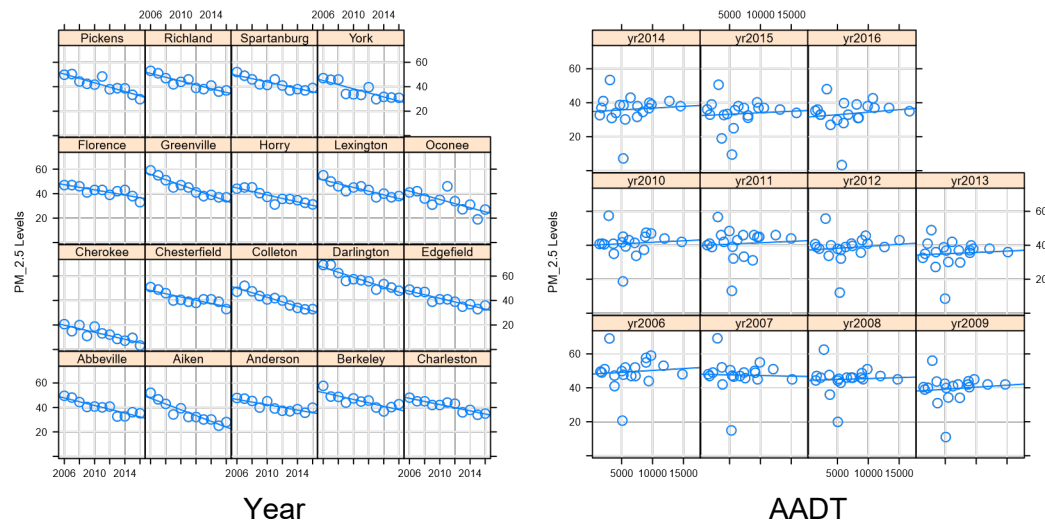
(b) PM₂.₅ levels for counties
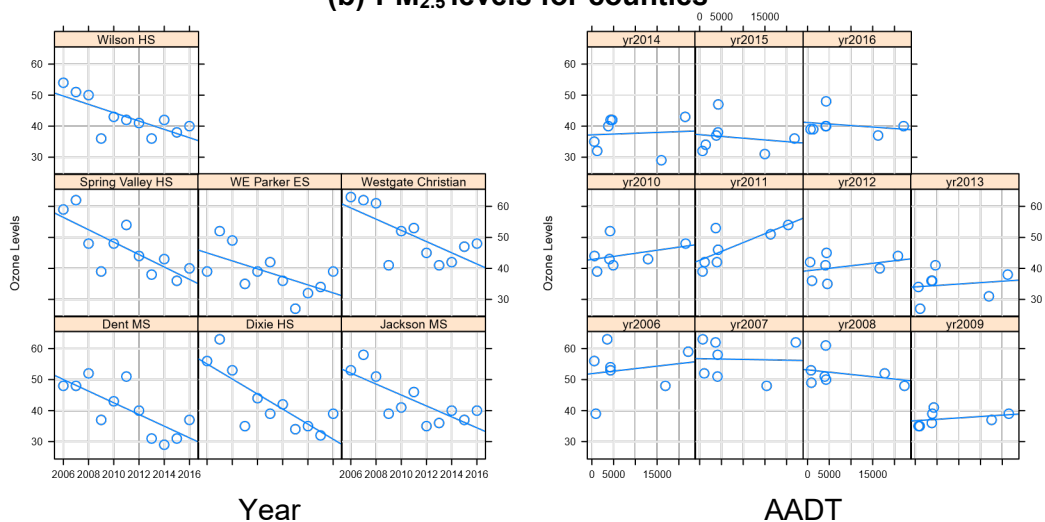
(c) O₃ levels for counties
Figure 3 Correlation matrices of AQIs for different counties, schools, AADT and year

2.2 Microdata, Equipment, and Calibration

Air quality eggs (WickedDevice) has the capability of sampling CO, CO_2, O_3, NO_2, VOC, $PM_{2.5}$, and SO2 in parts per million and/or billion as well as humidity, temperature, and timestamps every 5 seconds. The sensors are able to operate within the range of -20 to 40 ºC, the accuracy of humidity is with 1.8% and 0.2 ºC. It was noted by the producers that 15 seconds exposure of 20% relative humidity would require recalibration. We designed an ad-hoc shelter and did not deploy the sensors on rainy days outside.

Traffic-related air pollution has been associated with adverse cardiorespiratory effects, including increased asthma prevalence. Asthma has affected children, causing them to miss on average four days of school a year. Studies showing the outcome on the impact of traffic volume on air quality around schools have been reported but only a few documentations show a link between traffic volume and air quality in local areas. The research team has been able to develop and explain prediction models for future Air Quality Index (AQI) and compared large scale historical data for traffic volume (average annual daily traffic (AADT)) and AQIs of harmful ozone and $PM_{2.5}$ (USEPA) for specific schools in South Carolina. The team was able to analyze the impact of transportation on air quality around 7 schools in South Carolina from data that covered 2006 to 2016. As the next step, hourly and daily traffic datasets in South Carolina are used to understand correlations within a 3 miles radius.

(a) Calibration   (b) Environmental Eggs at Bridge Creek   (c) Radar set-up
Figure 4 Calibration of Environmental eggs at EPA Parklane Station

Moreover, the Air Quality Sensor (or referred to as the air quality egg) was placed for a week at the Parklane EPA site for calibration (see Figure 4). Then, sensors were deployed at Benedict College around Alumni Hall and the Business Development Center and observed that air pollution readings were slightly higher around Alumni Hall possibly due to higher traffic volume. For measuring the air quality around the College, the sensors and radar were deployed at the intersection between Taylor Street and Harden Street. The results give insights into how these simple sensors would be used for intersection level emissions data and possible utilization of such for eco-friendly signal timing.

Calibration for PM2.5

EPA = 1.3282*Eggs
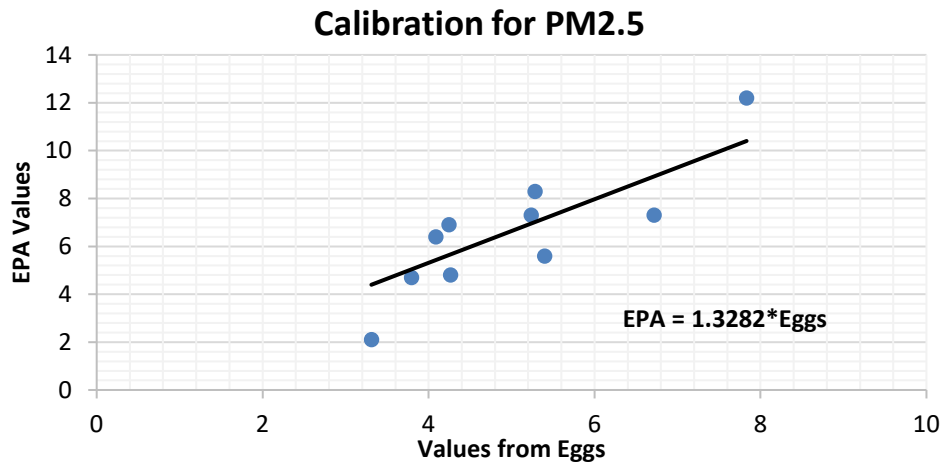
Figure 5 Calibration line for PM$_{2.5}$

Figures 5 and 6 show the calibration lines fitted using EPA and Air Quality Egg (Eggs) values. The lines approximately follow each other. Thus, one can use these equations to be able to transform their observations from environmental egg sensors and utilize them in their models and evaluations.

Calibration for O3

EPA = 0.24*Eggs
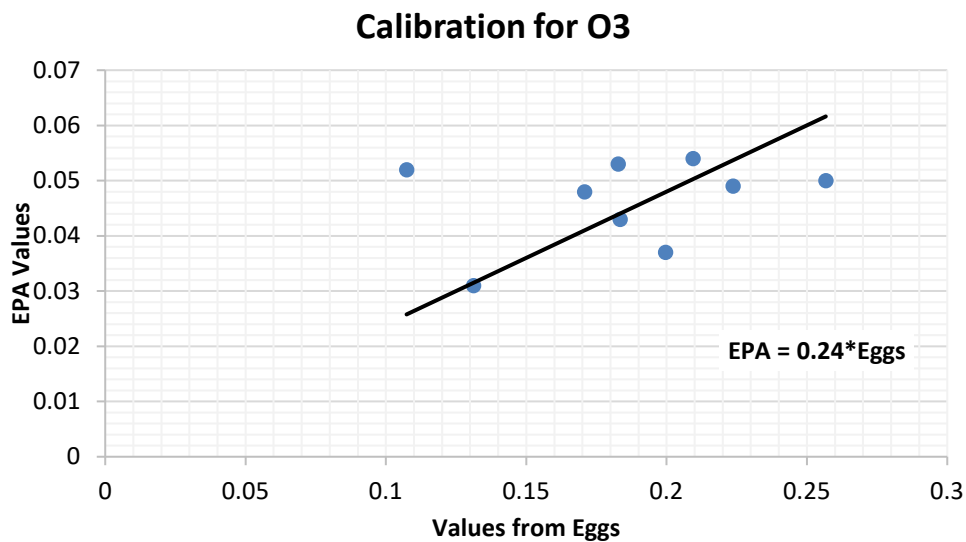
Figure 6 Calibration line for O$_3$

CHAPTER 3

Models for Estimating Emissions

3.1 Bridge Creek Middle School Experiment

The purpose of this section of the project was to test the functionality of the real-time data equipment which includes the traffic count radar and the air monitoring sensors. We assembled a platform that held the instrument. The air quality sensors were powered by portable solar power banks. The air monitoring eggs were placed closer to the road. The traffic radar recorded the traffic count on two different lanes in and out of the school.

With the traffic count radar, we recorded traffic counts for both lanes and we observed the differences in the traffic flow as it varied with time. More traffic count was obtained during the morning hours and during the hours which the school was supposed to dismiss and zero traffic count was recorded at night time. The result didn't show a direct relationship between traffic count and air quality data (i.e., data related to NO_2, $PM_{2.5}$, VOC, and CO_2). This would be due to very low levels that could not be measured by our sensors like temperature, humidity and vehicle speed, etc. which impact the emissions.

$$y = 0.0116x + 0.5014$$
$$R^2 = 0.0694$$
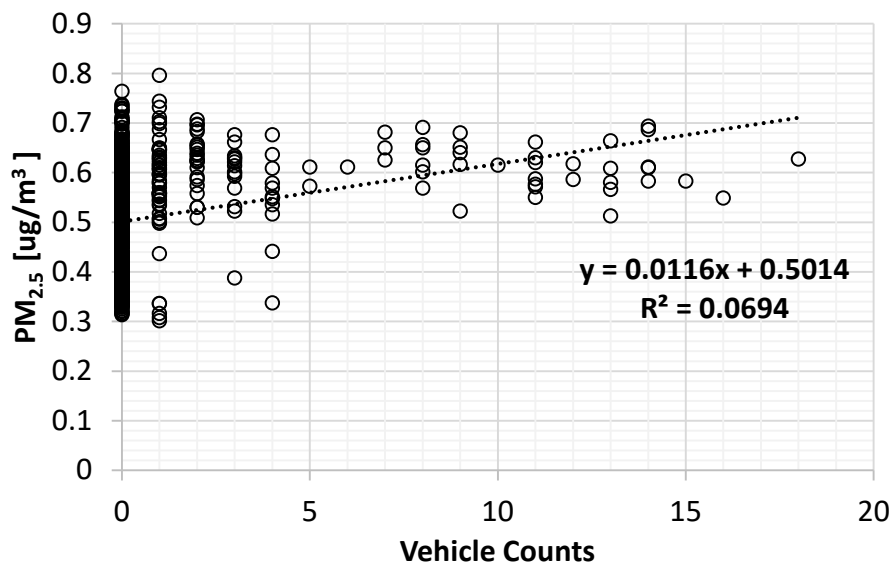
Figure 7 $PM_{2.5}$ vs traffic counts at Bridge Creek

Other air monitoring sensors were placed next to the school building, where we were able to record three pollutants (NO_2, CO_2, and $PM_{2.5}$) and volatile organic compounds. From the result of those sensors, we observe a relationship between NO_2 and the temperature. The CO_2 and VOC plots have also a similar shape but are not related to the temperature and to the traffic count. The $PM_{2.5}$ data is constantly increasing over time but has no strong correlation with the traffic count (see Figure 7). More traffic count was obtained around the school opening and the school closing hours and no traffic was recorded at night time. For the monitoring station located next to the school, we can conclude that the temperature had a significant influence on the NO_2, but not on CO_2 and $PM_{2.5}$. There is no relationship between traffic count and any other pollutant.

Table 2 Regression model for VOC

Variable	Estimate	p-value	Sample Size	R^2
Intercept	32.863	1.04E-07	1503	0.328
Time [min]	0.004	1.37E-01		
Traffic Count	-2.576	1.74E-07		
Temperature[degF]	2.281	4.87E-84		

Table 3 Regression model for CO_2

Variable	Estimate	p-value	Sample Size	R^2
Intercept	113.549	4.118E-07	1503	0.328
Time [min]	0.015	1.371E-01		
Traffic Count	-9.361	1.706E-07		
Temperature[degF]	8.286	4.371E-84		

Next, numerical results for emission values in terms of NO_2, $PM_{2.5}$, VOC, and CO_2 were presented. Multiple regression models are given to discuss possible correlations among variables and factors. According to the results, statistically, only the traffic count in Table 2 is significant for VOC. However, the coefficient of the traffic count is negative. Table 3 shows that traffic count has a similar contribution to CO_2. From Table 4, NO_2 emissions are impacted positively by traffic counts. The overall model shows an explanation of NO_2 values at R^2=0.935. From Table 5, $PM_{2.5}$ the p-value gets closer to being significant, however, the coefficient is very close to zeroEnvironment suggesting that it is not a significant contributor to explain $PM_{2.5}$ variations. In fact, temperature and time of day were found to be important covariates. Given critical emission values, models would be able to give insights where and when they can be considered significant. The coefficients of determination values in the rest of the report would be considered high for experimental data.

Table 4 Regression model for NO_2

Variable	Estimate	p-value	Sample Size	R^2
Intercept	-141.961	0.000	1503	0.935
Time [min]	-0.032	0.000		
Traffic Count	0.108	0.461		
Temperature[degF]	4.541	0.000		

Table 5 Regression model for $PM_{2.5}$

Variable	Estimate	p-value	Sample Size	R^2
Intercept	1.090	0.000	885	0.751
Time [min]	0.000	0.000		
Traffic Count	-0.001	0.209		
Temperature[degF]	-0.012	0.000		

During our experiment, we observed that the NO_2 values were higher from the Air Monitoring Sensors placed away from the school building but were lower than 350 parts per billion (ppb), to be considered clean air the values of NO_2 have to be below 600 ppb, so in this case this level of NO_2 is not a bad impact on the air quality around the school. The values of $PM_{2.5}$ were almost the same for the Air Monitoring Sensors placed close to the road and the one close to the school

building. We found the values of $PM_{2.5}$ were below 5.0 ug/m^3. As the requirement of $PM_{2.5}$ values for clear air is less than 12.0 ug/m^3, there is no impact of $PM_{2.5}$ on the air quality around the school.

From our experiment, we can conclude that Bridge Creek Middle School has clean air and that the traffic count had no influence, based on our collected data, on pollution. After calibration of the Air Monitoring Sensors, more experiments will be done at Bridge Middle School and other schools for a reasonable number of times in order to be able to draw a concrete conclusion on the impact of traffic volume through emission levels on Air Quality Index.

3.2 Benedict College Campus Experiment

Part of an additional experiment, we collected data at Benedict College around Alumni Hall, which is very close to Taylor St. at Harden St. intersection, between June 20 and June 28, 2018. Results are given in Tables 6 and 7 as aggregated average hourly volumes and NO_2 and VOC data. Multiple regression models were fit for both pollution criteria. Traffic volume is not a significant factor for NO2 with 0.84 p-values. However, for VOC, volume was significant with a low p-value of 0.0014. The coefficients are negative suspected in some cases, which would be due to high temperature and humidity levels during summer.

Table 6 Regression model for NO_2

Variable	Estimate	p-value	Sample Size	R^2
Intercept	221.523	5.47E-17	23	0.794
Time [hour]	-3.493	6.19E-08		
Traffic Count	-0.002	8.45E-01		

Table 7 Regression model for VOC

Variable	Estimate	p-value	Sample Size	R^2
Intercept	265.572	4.4862E-12	23	0.622
Time [hour]	1.413	1.3682E-01		
Volume	-0.103	1.4562E-04		

CHAPTER 4
Models for Estimating Air Quality Index

4.1 Introduction

To determine air pollutant variation with respect to AADT for each of the selected schools in South Carolina, mixed effect multilevel linear models, as well as multiple linear regression models, are utilized. They can be simply expressed as an additive model: $z \sim AADT + Year + County + e$, where the response variables z are O_3 or $PM_{2.5}$ observations, the covariates are AADT and Year, and the factors are counties and schools. For the multiple linear models, the coefficients of AADT and Year are fixed regardless of county or school, whereas, in the multilevel model, the coefficients of AADT and Year are variable. Similarly, the error term e is assumed to be fixed for the multiple linear model and variable for the multilevel model. However, this assumption can be relaxed by selecting an appropriate correlation structure and/or using a more sophisticated parameter estimation method.

In the classic regression modeling approach, the following assumptions need to be met: (1) normality of the residuals, (2) constant variance of the errors, (3) correlation of the errors, and (4) nonlinearity of the predictors. In this study, visual diagnostics were performed to ascertain that these assumptions are met. From Figure 8, it can be observed that residuals do not exhibit any pattern and most of the quantile-quantile (Q-Q) plots follow a straight line. Therefore, homogeneous variance and normality can be assumed. No autocorrelation of errors was observed; however, if there were, the Grey Models can handle correlated error structure. In addition, regression models are able to handle geographic variations through a hierarchical structure. Due to the temporal and spatial nature of the data, this study adopts the combined, LMER (Linear Mixed-Effects Regression) and GM, modeling approach as suggested by Clements (Clements and Harvey (2010)).

(a) O_3 AQIs for counties

(b) $PM_{2.5}$ AQIs for counties

(c) O_3 AQIs schools
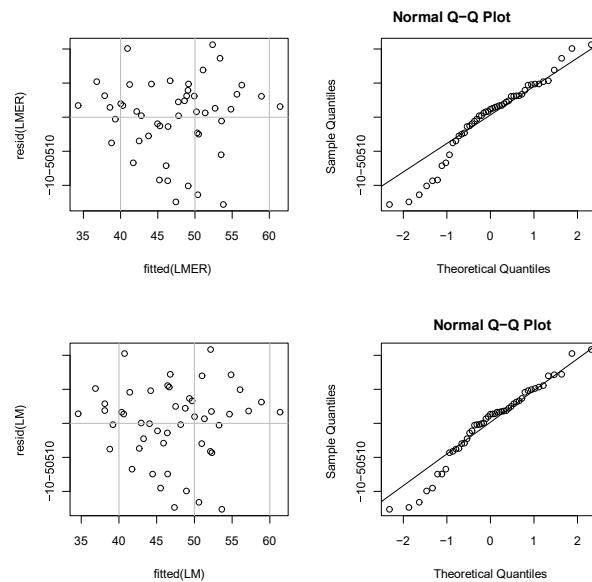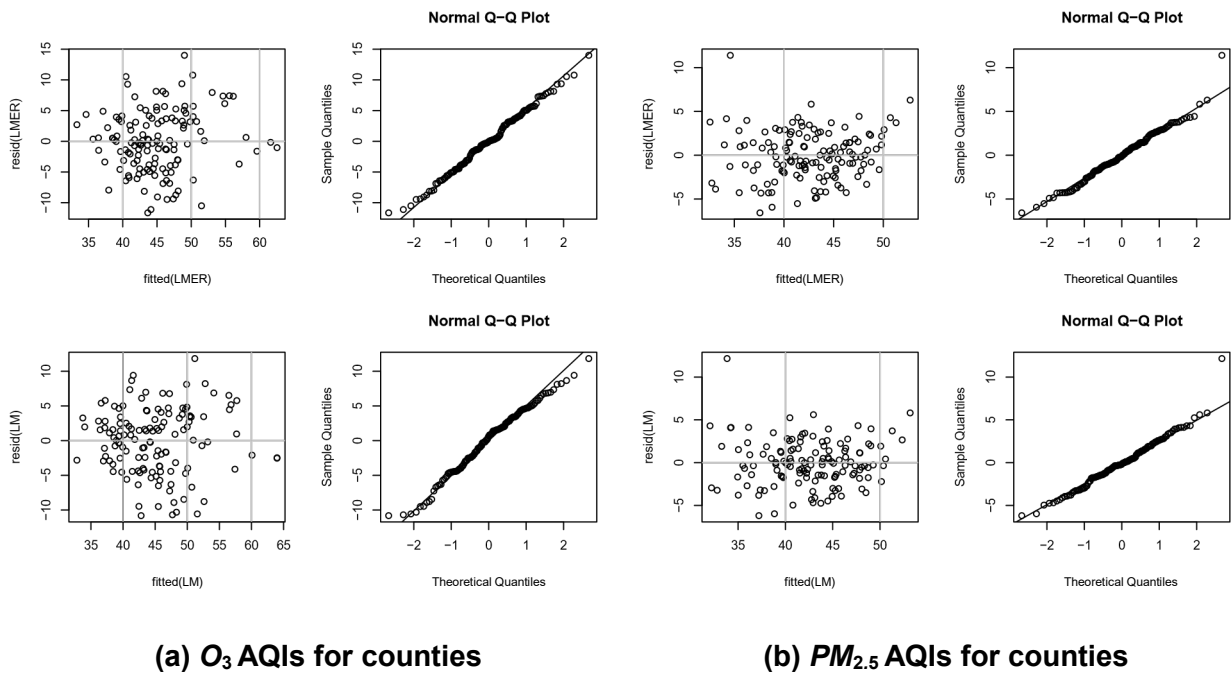
Figure 8 Model fitting diagnostics for the linear models

4.2 Simple Linear Regression Models

For the county-level model, multiple linear regressions as shown in Eq. 1 with ordinary least squares estimators were fitted using data from 2006 to 2012. Note that the data set is split into two sets, one for model estimation (2006-2012) and one for model validation (2013-2016).

$$z = b + b_1 x_1 + b_2 x_2 + c_i + e \tag{1}$$

where z is either O_3 or $PM_{2.5}$ level, x_1 is years, x_2 is AADT and c is county ($i=1,...,19$) and $e \sim N(0, \sigma_z^2)$ is white noise error. The school-level model has a similar specification.

These models are analyzed using R software. Table 8 provides the estimated coefficients and p-values for three linear models. These models do not have intercepts. Their R^2 values are 98.9%, 98.6%, and 99.6%, respectively. Only the AADT coefficient for the school-level model is not statistically significant. However, since AADT has been shown to be a significant covariate in past studies and also in the county-level model in this study, it is retained in the model.

Table 8 Estimated Coefficients and p-values of LMs for O₃ and PM₂.₅ AQIs for different counties, and for O₃ AQIs for different schools*

Variable	O₃		PM₂.₅	
	Estimate	p-value	Estimate	p-value
Year	-1.294	<0.001	-1.462	<0.001
Avg AADT	0.010	<0.001	0.004	0.014
Abbeville	2632	<0.001	2973	<0.001
Aiken	2590	<0.001	2957	<0.001
Anderson	2589	<0.001	2953	<0.001
Berkeley	2545	<0.001	2941	<0.001
Charleston	2491	<0.001	2920	<0.001
Cherokee	2591	<0.001	2952	<0.001
Chesterfield	2620	<0.001	2972	<0.001
Colleton	2601	<0.001	2964	<0.001
Darlington	2616	<0.001	2970	<0.001
Edgefield	2621	<0.001	2973	<0.001
Florence	2580	<0.001	2954	<0.001
Greenville	2548	<0.001	2945	<0.001
Horry	2554	<0.001	2939	<0.001
Lexington	2570	<0.001	2947	<0.001
Oconee	2605	<0.001	2958	<0.001
Pickens	2599	<0.001	2952	<0.001
Richland	2527	<0.001	2933	<0.001
Spartanburg	2581	<0.001	2953	<0.001
York	2569	<0.001	2946	<0.001
Variables	Estimate	p-value		
Year	-2.375	<0.001		
Avg AADT	-0.0008	0.539		
Dent Middle School	4830	<0.001		
Dixie High School	4819	<0.001		
Jackson Middle School	4820	<0.001		
Spring Valley High School	4839	<0.001		
WE Parker Elementary School	4813	<0.001		
Westgate Christian School	4828	<0.001		
Wilson High School	4819	<0.001		

*.**Estimate** column contains estimated co-efficient values, and **p-value** column represents statistical significance for hypothesis testing (evidence for null hypothesis rejection - H₀: fitted parameter=0)

4.3 Multilevel Linear Regression Models

Hierarchical, multilevel, or linear mixed-effect regression models (LMER) can address the changes of covariates (AADT and Year) with respect to different factors (i.e., counties and schools). The LMER specification for counties is shown in Eq. 2.

$$z = b_0 + b_1x_1 + b_2x_2 + y_i[b_{0i} + b_{1i}x_1 + b_{2i}x_2 + e_i] \tag{2}$$

where z is either O_3 or $PM_{2.5}$ level, x_1 is years, x_2 is AADT, $yi \in [0,1]$ are indicator variables, $i=1,...,19$ corresponds to counties, and $e_i \sim N(0,\sigma_i^2)$ is white noise error. The LMER specification for schools is similar.

These models were fitted using the *lme*4 package in R which uses the maximum likelihood (ML) and restricted maximum likelihood estimation (REML), where ML assumes normality and independence (Bates et al. (2015); Ga lecki and Burzykowski (2013)) and REML assumes independent observations with homogeneous variance. Table 9 provides the estimated coefficients and p-values for the LMER models. In Table 9, the "Fixed" estimate corresponds to the first three terms of Eq. 2. The county or school estimate corresponds to the additive effect (fourth term) of Eq. 2.

Next, models for both data types are generated and root means square errors are reported. Based on the close results, we may conclude that AADT or VMT performs very similarly in terms of explaining the air quality index.

Table 9 Estimated LMER Model Coefficient, Parameter, and p-values for O_3 and $PM_{2.5}$ AQIs for different counties, and for O_3 AQIs for different schools*

Var.	O_3			$PM_{2.5}$		
	Int. (p-val)	Year(x1)	AADT(x2)	Int	Year(x1)	AADT(x2)
Fixed	2630.00	-1.288	0.0007	2630.00	-1.288	0.0007
p-values	(<0.001)	(<0.001)	(0.122)	(<0.001)	(<0.001)	(0.085)
Abbeville	2634.99	-1.288	-0.0004	2973.57	-1.459	0.0003
Aiken	2629.44	-1.289	0.0010	2974.10	-1.460	0.0004
Anderson	2628.95	-1.288	0.0004	2979.30	-1.465	0.0005
Berkeley	2632.87	-1.287	-0.0010	2977.04	-1.463	0.0004
Charleston	2632.14	-1.288	-0.0003	2977.50	-1.463	0.0004
Cherokee	2628.90	-1.288	0.0007	2981.43	-1.467	0.0005
Chesterfield	2629.55	-1.288	0.0007	2971.60	-1.457	0.0003
Colleton	2625.80	-1.288	0.0009	2973.18	-1.459	0.0003
Darlington	2632.91	-1.288	0.0003	2970.71	-1.457	0.0003
Edgefield	2626.29	-1.289	0.0014	2972.70	-1.459	0.0003
Florence	2629.72	-1.288	0.0005	2973.95	-1.460	0.0003
Greenville	2630.08	-1.288	0.0005	2968.98	-1.455	0.0002
Horry	2630.63	-1.288	0.0003	2980.12	-1.466	0.0005
Lexington	2625.29	-1.290	0.0027	2971.43	-1.457	0.0003
Oconee	2628.88	-1.288	0.0007	2979.93	-1.465	0.0005
Pickens	2630.36	-1.289	0.0014	2981.44	-1.467	0.0005
Richland	2629.86	-1.288	0.0006	2973.61	-1.459	0.0003
Spartanburg	2627.23	-1.289	0.0021	2971.90	-1.458	0.0003
York	2630.39	-1.288	0.0002	2978.24	-1.464	0.0004

Var.	Int.	Year (x1)	AADT (x2)
Fixed	5008.00	-2.465	-0.00036
p-values	(<0.001)	(<0.001)	(0.396)
Dent MS	5008.50	-2.473	0.00036
Dixie HS	5008.40	-2.469	0.00003
Jackson MS	5008.40	-2.470	0.00010
Spring Valley HS	5006.80	-2.433	-0.00314
WE Parker ES	5008.50	-2.472	0.00030
Westgate Christian	5008.20	-2.465	-0.00032
Wilson HS	5008.40	-2.471	0.00015

*.**Int**. the column contains LMER model intercept values; and **p-value** row represents statistical significance for hypothesis testing (evidence for null hypothesis rejection - H_0: fitted parameter=0)

4.4 AADT versus VMT

We found available 4 years of vehicle miles traveled (VMT) data from 2013 to 2016, which matched the available emissions data, for SC counties from public safety reports. Only regression models were compared. Models were fit with AADT and VMT values and results were provided. First, AADT and average VMT were plotted as in Figure 9 and model comparisons were provided in Table 10.

VMT=284681*AADT, R^2=0.936
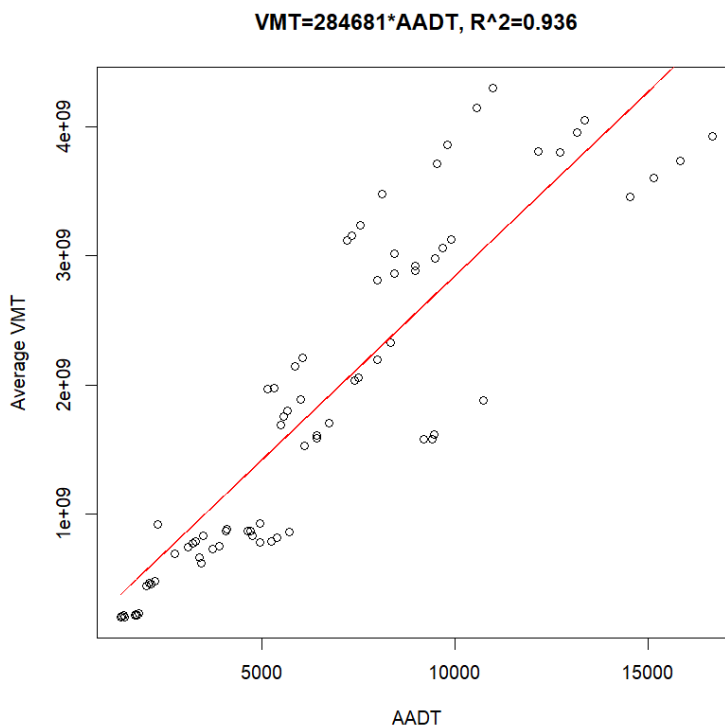
Figure 9 Average VMT versus AADT values for different counties

Table 10 AADT vs VMT model fitting

Criteria	Model	RMSE	AIC
O_3	LMER_AADT	1.918	414.789
	LMER_VMT	1.905	579.560
	LM_AADT	1.908	357.860
	LM_VMT	1.902	357.409
PM_2.5	LMER_AADT	2.680	475.057
	LMER_VMT	2.463	604.262
	LM_AADT	2.662	408.490
	LM_VMT	2.631	406.705

4.5 Grey Systems and its modifications

Grey systems are especially suited for datasets with a low number of observations, as is the case in this study. The Grey Systems theory was developed by Deng in 1982 (Ju-Long (1982)) and since then it has become the preferred method to study and model systems in which the structure or operation mechanism is not completely known (Deng (1989)). Grey System theory applications have been applied mainly in the area of finance (Kayacan et al. (2010)). Its application in

transportation is limited; examples include prediction of accident number and pavement degradation (Gao et al. (2010); An et al. (2012); Liu et al. (2014)).

According to the Grey Systems theory, the unknown parameters of the system are represented by discrete or continuous Grey numbers. The theory introduces a number of properties and operations on the Grey numbers, its degree of Greyness, and whitenization of the Grey number. The latter operation generally describes the preference of the number towards the range of its possible values (Liu et al. (2010)).

In order to model time series, the theory suggests a family of Grey models, where the basic one is the first order Grey model with one variable, will be referred to as GM(1,1). The principles and estimation of GM(1,1) are briefly discussed here; readers are referred to Deng (1989) (Deng (1989)) for additional information. Suppose that $X^{(0)} = (x^{(0)}(1), x^{(0)}(2),...,x^{(0)}(n))$ denotes a sequence of nonnegative observations of a stochastic process and $X^{(1)} = (x^{(1)}(1), x^{(1)}(2),...,x^{(1)}(n))$ is an accumulation sequence of $X^{(0)}$ computed as in Eq. (3).

$$x^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(j) \tag{3}$$

Then, (4) defines the original form of the GM(1,1).

$$x^{(0)}(k) + ax^{(1)}(k) = b \tag{4}$$

Let $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3),...,z^{(1)}(n))$ be a mean sequence of $X^{(1)}$ calculated by formula Eq. (5) and defined for $k = 2,3,\cdots,n$

$$z^{(1)}(k) = [z^{(1)}(k-1) + z^{(1)}(k)]/2 \tag{5}$$

Eq. (6) gives the basic form of GM(1,1).

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{6}$$

If $\hat{a} = (a, b)^T$ and

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} z^{(1)}(2) & 1 \\ z^{(1)}(3) & 1 \\ \vdots & \vdots \\ z^{(1)}(n) & 1 \end{bmatrix}$$

Then, as in Liu and Lin (2006), the least-squares estimate of the GM(1,1) model is $\hat{a} = (B^T B)^{-1} B^T Y$ and Eq. (7) is the whitenization equation of the GM(1,1) model (GM).

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{7}$$

Suppose that $\hat{x}^{(0)}(k)$ and $\hat{x}^{(1)}(k)$ represent the time response sequence (the forecast) and the accumulated time response sequence of Grey model at time k respectively. Then, the latter can be obtained by solving Eq. (7):

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}, k = 1,2,\dots,n \tag{8}$$

According to the definition in Eq. (3), the restored values of $\hat{x}^{(0)}(k+1)$ are calculated as $\hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$:

$$\hat{x}^{(0)}(k+1) = (1 - e^a)\left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak}, k = 1,2,\dots,n \tag{9}$$

Eq. (9) gives the method to produce forecasts for all k in 2, 3,...,n. However, for longer time series, a rolling GM is preferred. The rolling model observes a window of a few sequential data points in the series: $x^{(0)}(k+1), x^{(0)}(k+2),\dots,x^{(0)}(k+w)$, where $w \geq 4$ is the window size. Then, the model forecasts one or more future data points: $\hat{x}^{(0)}(k+w+1), \hat{x}^{(0)}(k+w+2)$. The process repeats for the next k.

4.6 The Grey Verhulst Model (GV)

The response sequence Eq. (9) implies that the basic GM works best when the time series exhibits a steady growth or decline and may not perform well when the data has oscillations or saturated sigmoid sequences. For the latter case, the Grey Verhulst model (GV) is generally used (Liu et al. (2010)). The basic form of the GV is shown in Eq. (10).

$$x^{(0)}(k) + az^{(1)}(k) = b\left(z^{(1)}(k)\right)^2 \tag{10}$$

The whitenization equation of GVM is:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b\left(x^{(1)}\right)^2 \tag{11}$$

Similar to the GM(1,1), the least-squares estimate is applied to find $\hat{a} = (B^T B)^{-1} B^T Y$, where

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & z^{(1)}(2)^2 \\ -z^{(1)}(3) & z^{(1)}(3)^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & z^{(1)}(n)^2 \end{bmatrix}$$

The forecasts $\hat{x}^{(0)}(k+1)$ are calculated using Eq. (12).

$$\hat{x}^{(0)}(k+1) = \frac{ax^{(0)}(1)(a - bx^{(0)}(1))}{bx^{(0)}(1) + (a - bx^{(0)}(1))e^{a(k-1)}} * \frac{e^{a(k-2)}(1 - e^a)}{bx^{(0)}(1) + (a - bx^{(0)}(1))e^{a(k-2)}} \tag{12}$$

4.7 Error Corrections to Grey Models

In order to increase the accuracy of the Grey models, suppose that $\epsilon^{(0)} = \epsilon^{(0)}(1),\dots, \epsilon^{(0)}(n)$ is the error sequence of $X^{(0)}$, where $\epsilon^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$. If all errors are positive, then a remnant GM(1,1) model can be built (Liu et al. (2010)). Whether the errors are positive or negative, $\epsilon^{(0)}$ can be expressed using the Fourier series (Tan and Chang (1996)) as in Eq. (13).

$$\epsilon^{(0)}(k) \cong \frac{1}{2}a_0 + \sum_{i=1}^{z}\left(a_i\cos(\frac{2\pi i}{T}k) + b_i\sin(\frac{2\pi i}{T}k)\right) \tag{13}$$

where $k = 2,3,...,n$, $T = n - 1$, and $z = \left(\frac{n-1}{2}\right) - 1$.

The solution is found via the least squares estimate, presuming that $\epsilon^{(0)} \approx PC$ where C is a vector of coefficients: $C = [a_0a_1b_1a_2...a_nb_n]^T$ and matrix P is:

$$P = \begin{bmatrix} \frac{1}{2} & \cos(2\frac{2\pi}{T}) & \sin(2\frac{2\pi}{T}) & ... & \cos(2\frac{2\pi z}{T}) & \sin(2\frac{2\pi z}{T}) \\ \frac{1}{2} & \cos(3\frac{2\pi i}{T}) & \sin(3\frac{2\pi}{T}) & ... & \cos(3\frac{2\pi z}{T}) & \sin(3\frac{2\pi z}{T}) \\ \vdots & & & & & \\ \frac{1}{2} & \cos(n\frac{2\pi i}{T}) & \sin(n\frac{2\pi}{T}) & ... & \cos(n\frac{2\pi z}{T}) & \sin(n\frac{2\pi z}{T}) \end{bmatrix}$$

CHAPTER 5

Modeling Results and Discussion

Prediction Results

This section compares the performance of the linear model (LM), LMER, GM, error-corrected GM (EGM), GV, error-corrected GV (EGV), and LMER combined with EGM on the validation data set. Average RMSEs for O_3 and $PM_{2.5}$ county-level models, i.e., [LMER, LM, GM, EGM, GV, EGV, LMER+EGM], are [3.2, 5.1, 3.9, 3.3, 3.3, 2.7, 2.1] and [5.2, 7.0, 3.3, 2.3, 2.8, 2.1, 1.9], respectively. Average RMSEs for O_3 school-level models, i.e., [LMER, LM, GM, EGM, GV, EGV, LMER+EGM], are [4.0, 4.1, 4.9, 3.7, 3.6, 3.6, 2.1], respectively. In each case, the highest accuracy was achieved by the combination method. Figure 10 shows the RMSEs for the different models in predicting the O_3 and $PM_{2.5}$ levels for counties and schools. There are no results for $PM_{2.5}$ for schools. It can be observed that the LMER+EGM model has the lowest RMSE as well as the lowest variance of RMSE. For the county-level models, all have RMSE less than 10.0. For school-level models, all predicted levels have RMSE less than 6.0. It can also be observed that the GM models outperformed the LMER and LM models. In summary, all models produced estimates within ±10% of true values.

Ozone (O₃) for Counties PM₂.₅ for Counties Ozone (O₃) for Schools
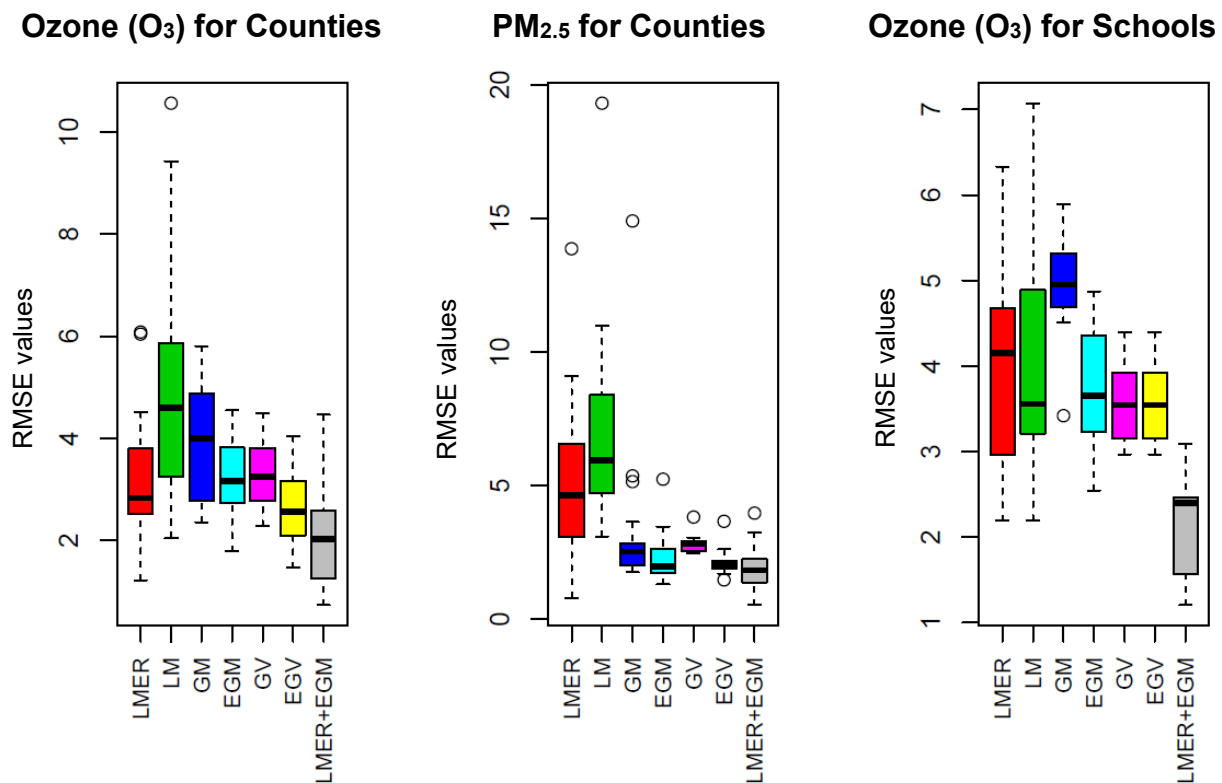
Figure 10 RMSE values related to O₃ and PM₂.₅ levels prediction for different models for 2013-2016 AQIs

In Figure 11, better-performing methods are presented, i.e., LMER, EGM, and LMER+EGM. It can be observed that the LMER model is the least accurate and the LMER+EGM is the most accurate. The results corroborate previous research findings (e.g., Clements and Harvey (2010)) that a combined model with competing methods produces superior results. In this study, the combined model's weighted forecast is $z_c = \alpha z_1 + (1 - \alpha) z_2$ where z_1 and z_2 are predictions from

different models, specifically LMER and EGM. The optimal α_* from training or partial testing data can be determined as $\alpha_* = (\sum_{t=1}^{T} e_{2t}^2 - \sum_{t=1}^{T} e_{1t}e_{2t})/(\sum_{t=1}^{T} e_{1t}^2 + \sum_{t=1}^{T} e_{2t}^2 - 2\sum_{t=1}^{T} e_{1t}e_{2t})$ where $e_{1t} = z_t - \hat{z}_{1t}$ and $e_{2t} = z_t - \hat{z}_{2t}$ (Newbold and Harvey (2008)). However, in this study, the optimal weight α was empirically derived to be 0.15.
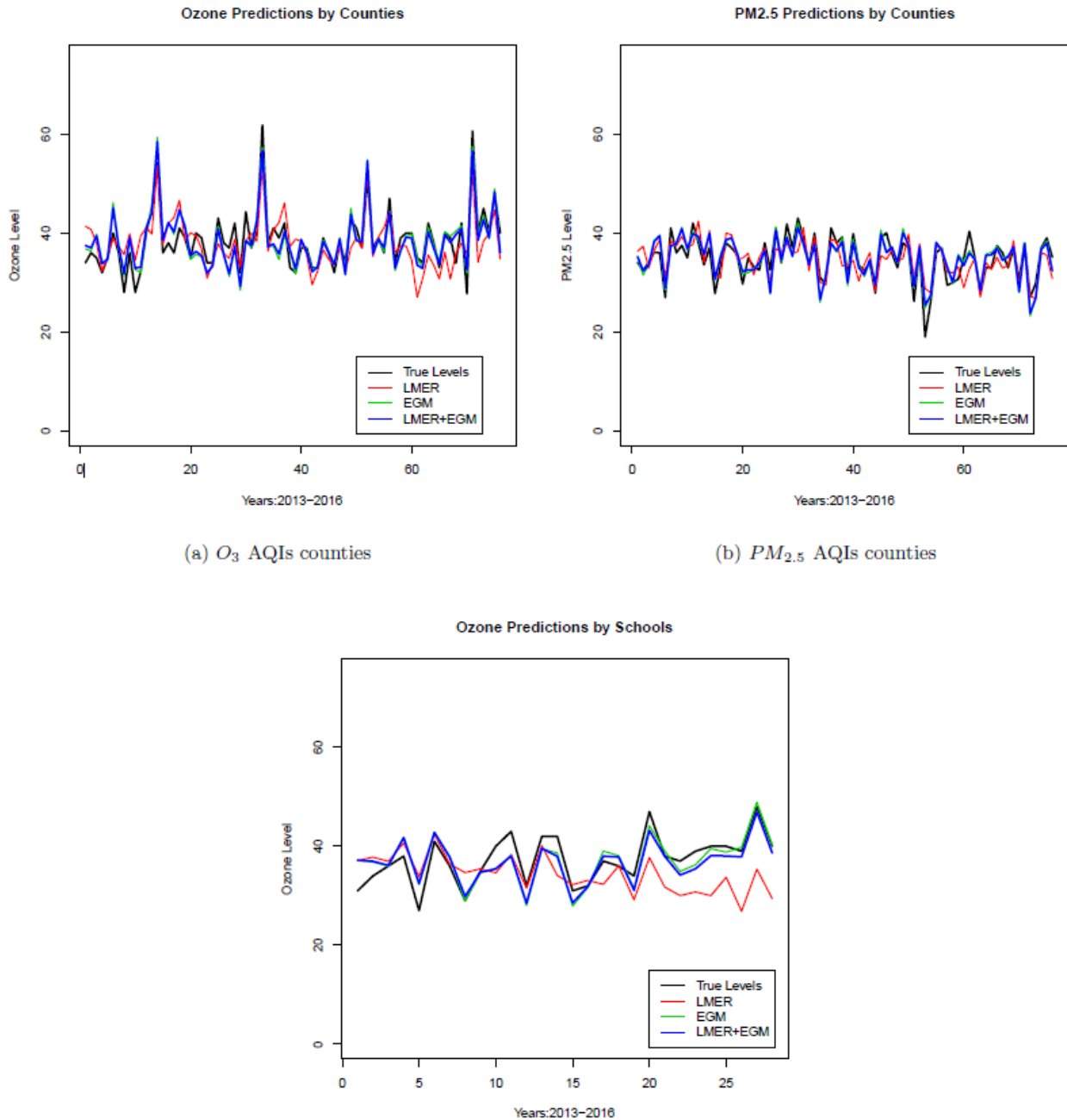

(a) O_3 AQIs counties

(b) $PM_{2.5}$ AQIs counties

Figure 11 Predictions for O$_3$ and PM$_{2.5}$ levels for different counties

CHAPTER 6

Conclusions

This project investigates the effectiveness of low-cost air quality sensors in traffic applications at a school and a college campus. The study included set up, calibration, and data collection at an elementary school and a college campus. Based on the experiments, the following observations are made.

1. Equipment's operation range outside should be carefully selected for reliable data collection.
2. Missing data are possible.
3. Units for different sensors may vary.
4. Calibration should be done for a longer period considering data may not be reported from EPA monitoring sites.
5. Equipment can be used with portable power sources and simply installed on a pole.
6. Weather elements should be considered.
7. Sensors are inexpensive and are able to provide various pollutant data every 5 seconds, thus, they can be used within various intelligent transportation systems real-time applications.

The project also develops prediction models for O_3 and $PM_{2.5}$ levels for different schools and counties in South Carolina. Several types of models were investigated. They include LM, LMER, GM, EGM, GV, EGV, and LMER+EGM. The LM model produced the least accurate estimate while the LMER+EGM model produced the most accurate estimate (average RMSE is less than 5%). The model estimates suggest that air quality in South Carolina will continue to degrade in the coming years. An interesting finding is that some counties (namely, Abbeville, Berkeley, and Charleston) and schools (namely, Spring Valley HS and Westgate Christian HS) have higher levels of O_3 or $PM_{2.5}$ when AADT decreases. Our finding suggests that there are additional factors, other than AADT, which would influence the air quality in these counties and schools. An explanation for this could be these counties or schools are in close proximity to an industrial park. For example, Berkeley County is home to the Boeing plant that assembles the 787 Dreamliner and Charleston County is home to the Port of Charleston.

The methods presented in this study can be seen as a step forward in air quality prediction that considers both spatial and temporal factors. These models are important for planning purposes to identify risk areas and suitable locations for sensitive facilities, such as K-12 schools and hospitals. Our future work will focus on developing site-specific models using hourly traffic and air quality measures and using high-quality portable air quality sensors.

REFERENCES

Adams, M. D., Requia, W. J., 2017. How private vehicle use increases ambient air pollution concentrations at schools during the morning drop-off of children. Atmospheric Environment 165, 264–273.

An, Y., Cui, H., Zhao, X., 2012. Exploring gray system in traffic volume prediction based on tall database.

International Journal of Digital Content Technology & its Applications 6 (5).

Bates, D., M¨achler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67 (1), 1–48.

Brunekreef, B., Janssen, N. A., de Hartog, J., Harssema, H., Knape, M., van Vliet, P., 1997. Air pollution from truck traffic and lung function in children living near motorways. Epidemiology, 298–303.

Clements, M. P., Harvey, D. I., 2010. Forecast encompassing tests and probability forecasts. Journal of Applied Econometrics 25 (6), 1028–1062.

Deng, J. L., Nov. 1989. Introduction to grey system theory. J. Grey Syst. 1 (1), 1–24.

URL http://dl.acm.org/citation.cfm?id=90757.90758

Ga lecki, A., Burzykowski, T., 2013. Linear mixed-effects models using R: A step-by-step approach. Springer Science & Business Media.

Gao, S., Zhang, Z., Cao, C., 2010. Road traffic freight volume forecasting using support vector machine. In: The Third International Symposium Computer Science and Computational Technology (ISCSCT 2010). Citeseer, p. 329.

Gauderman, W. J., Avol, E., Lurmann, F., Kuenzli, N., Gilliland, F., Peters, J., McConnell, R., 2005. Childhood asthma and exposure to traffic and nitrogen dioxide. Epidemiology 16 (6), 737–743.

Gelman, A., Hill, J., Su, Y.-S., Yajima, M., Pittau, M., Goodrich, B., Si, Y., Kropko, J., Goodrich, M. B., 2015. Package mi.

Guarnieri, M., Balmes, J. R., 2014. Outdoor air pollution and asthma. The Lancet 383 (9928), 1581–1592.

Hallmark, S., Fomunung, I., Guensler, R., Bachman, W., 2000. Assessing impacts of improved signal timing as a transportation control measure using an activity-specific modeling approach. Transportation Research Record: Journal of the Transportation Research Board (1738), 49–55.

Howitt, A. M., Moore, E. M., 1999. Implementing the transportation conformity regulations. TR News (202).

Janssen, N. A., Brunekreef, B., van Vliet, P., Aarts, F., Meliefste, K., Harssema, H., Fischer, P., 2003. The relationship between air pollution from heavy traffic and allergic sensitization, bronchial hyperresponsiveness, and respiratory symptoms in dutch schoolchildren. Environmental health perspectives 111 (12), 1512.

Janssen, N. A., van Vliet, P. H., Aarts, F., Harssema, H., Brunekreef, B., 2001. Assessment of exposure to traffic related air pollution of children attending schools near motorways. Atmospheric environment 35 (22), 3875–3884.

Ju-Long, D., 1982. Control problems of grey systems. Systems & Control Letters 1 (5), 288–294.

Kayacan, E., Ulutas, B., Kaynak, O., 2010. Grey system theory-based models in time series prediction.

Expert systems with applications 37 (2), 1784–1789.

Liu, S., Lin, Y., 2006. Grey information: theory and practical applications. Springer Science & Business Media.

Liu, S., Lin, Y., Forrest, J. Y. L., 2010. Grey systems: theory and applications. Vol. 68. Springer Science & Business Media.

Liu, W., Qin, Y., Dong, H., Yang, Y., Tian, Z., 2014. Highway passenger traffic volume prediction of cubic exponential smoothing model based on grey system theory. In: 2nd International Conference on Soft Computing in Information Communication Technology. Atlantis Press.

McConnell, R., Islam, T., Shankardass, K., Jerrett, M., Lurmann, F., Gilliland, F., Gauderman, J., et al., 2010. Childhood incident asthma and traffic-related air pollution at home and school. Environmental health perspectives 118 (7), 1021.

Mohai, P., Kweon, B.-S., Lee, S., Ard, K., 2011. Air pollution around schools is linked to poorer student health and academic performance. Health Affairs 30 (5), 852–862.

Mohammadyan, M., Alizadeh-Larimi, A., Etemadinejad, S., Latif, M. T., Heibati, B., Yetilmezsoy, K., Abdul-Wahab, S. A., Dadvand, P., 2017. Particulate air pollution at schools: indoor-outdoor relationship and determinants of indoor concentrations. Aerosol and Air Quality Research 17 (3), 857–864.

Newbold, P., Harvey, D. I., 2008. Forecast combination and encompassing. In: Clements, M. P., Hendry, D. F. (Eds.), A Companion to Economic Forecasting. Blackwell Publishers, Ch. 12, pp. 268–283.

Rakowska, A., Wong, K. C., Townsend, T., Chan, K. L., Westerdahl, D., Ng, S., Moˇcnik, G., Drinovec, L., Ning, Z., 2014. Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon. Atmospheric Environment 98, 260–270.

Tan, C., Chang, S., 1996. Residual correction method of fourier series to gm (1, 1) model. In: Proceedings of the first national conference on grey theory and applications, Kauhsiung, Taiwan. pp. 93–101.

WickedDevice, Environmental Air Quality Egg Sensors, www.wickeddevice.com (Last visited July 05, 2019)

Wjst, M., Reitmeir, P., Dold, S., Wulff, A., Nicolai, T., von Loeffelholz-Colberg, E. F., Von Mutius, E., 1993. Road traffic and adverse effects on respiratory health in children. Bmj 307 (6904), 596–600.

Xie, Y., Chowdhury, M., Bhavsar, P. and Zhou, Y., 2012. An integrated modeling approach for facilitating emission estimations of alternative fueled vehicles. Transportation Research Part D: Transport and Environment, 17(1), pp.15-20.

Zhang, K., Batterman, S., 2013. Air pollution and health risks due to vehicle traffic. Science of the total Environment 450, 307–316.