International Journal of Transportation Science and Technology 9 (2020) 29–41

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Evaluating the impact of traffic volume on air quality in South Carolina

Gurcan Comert [a], Samuel Darko [a], Nathan Huynh [b,*], Bright Elijah [a], Quentin Eloise [a]

[a] Computer Science, Physics, and Engineering Department, Benedict College, Columbia, SC 29204, USA
[b] Civil and Environmental Engineering Department, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Article history:
Received 13 February 2019
Received in revised form 24 May 2019
Accepted 28 May 2019
Available online 7 June 2019

Keywords:
Air quality index
Annual average daily traffic
Emissions
Grey systems
Multilevel linear models

ABSTRACT

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. 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. The data are 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 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 + 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 models' estimates suggest that air quality in South Carolina will continue to get worse in the coming years due to increasing 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 other factors, other than AADT, that influence the air quality in these counties and schools.

1. Introduction

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 (Gauderman et al., 2005; McConnell et al., 2010). Studies have also shown higher rates of morbidity and

* Corresponding author.
E-mail address: nathan.huynh@sc.edu (N. Huynh).

https://doi.org/10.1016/j.ijtst.2019.05.008
2046-0430/

mortality for drivers, commuters and 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 cardiovascular effects in adults, as well as 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; Janssen et al., 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 NO_2 reacts with VOC in the presence of heat from sunlight. $PM_{2.5}$ is composed of particulate matter with 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 by Hao et al. (2018) who evaluated the environmental impact of traffic congestion. Alternatively, air quality, or emissions can be more accurately determined 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 Abou-Senna et al. (2013) who 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.

To our knowledge, no such study has correlated the impact 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 study, the Grey models based on Grey System theory are utilized and they are compared against regression models. This approach is adopted 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 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 paper is organized as follows. Section 2 provides a description of the data. Sections 3–3.3 discuss the modeling techniques used in the study: multiple linear regression, multilevel linear regression, and Grey Systems. Section 4 presents the model validation results. Lastly, Section 5 provides concluding remarks and future research directions.

2. Data description

The data used in this study are obtained from the South Carolina Department of Transportation (SCDOT) and the United States Environmental Protection Agency (EPA) websites. Fig. 1 shows the 29 locations of EPA monitoring stations located throughout the state of South Carolina. The South Carolina Department of Health and Environmental Control has stations that monitor $CO, NO_2, O_3, PM_{2.5}, PM_{10}$ and SO_2 throughout the state. However, not every county has sensors that monitored all of these pollutants. Only O_3 and $PM_{2.5}$ are available for every county in the state. Therefore, air quality is limited to just O_3 and $PM_{2.5}$ in this study. 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.

Table 1 shows the emissions and AADT data obtained for 19 South Carolina Counties and selected schools in 2006. Note that the AADT shown in Table 1 represents the average AADT, taken from several count stations located in the proximity of EPA monitoring stations. Similar data were obtained up to 2016, for a total of 11 years. The datasets from EPA tend to contain missing data. To deal with this issue, missing data are treated with mixture models and are imputed using R-package (Gelman et al., 2015). Fig. 2 shows the utilized dataset before and after the missing data imputation. The black regions represent missing data that were subsequently imputed. The color in Fig. 2 denotes standardized values via transformation of $((x - \mu_x)/2\sigma_x)$ of the observations.

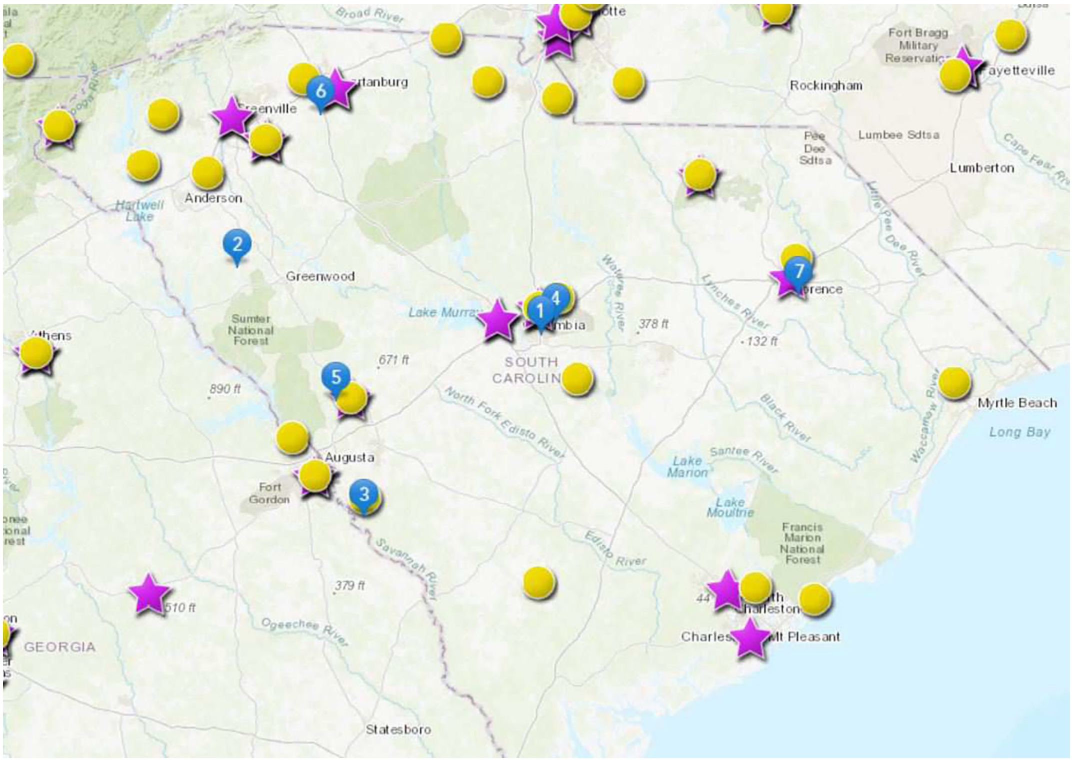


Fig. 1. EPA Stations (star = $PM_{2.5}$, circle = O_3) and nearby schools (numbered marker).

In Fig. 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 the entire county and school.

3. Methods

To determine air pollutant variation with respect to AADT for each of the selected schools in South Carolina, mixed effect multilevel linear regression models as well as multiple linear regression models are utilized. It was observed that AADT is highly correlated to vehicle-miles travelled (VMT), as shown in Fig. 4. For this reason and due to the fact that AADT data are much more readily available, AADT is used in the developed models instead of VMT. They can be simply expressed as additive models $z \sim AADT + Year + County + e$, where the response variables z are O_3 or $PM_{2.5}$ levels, the covariates are AADT and Year, and the factors are counties and schools. For the multiple linear regression model, 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 regression 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 was performed to ascertain that these assumptions are met. From Fig. 5, 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 were observed; however, if there were, the GMs can handle correlated error structure. In addition, regression models are able to handle geographic variations through hierarchical structure. Due to the temporal and spatial nature of the data, this study adopts the combined, LMER + GM, modeling approach as suggested by Clements and Harvey (2010).

3.1. 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 are split into two sets, one for model estimation (2006–2012) and one for model validation (2013–2016).

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			

$$z = b + b_1x_1 + b_2x_t2 + c_i + e \tag{1}$$

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

These models are estimated using the *lm* package in R. Table 2 provides the estimated coefficients and p-values for 3 linear regression models, O_3 for counties and schools, and $PM_{2.5}$ for counties. These models do not have intercepts. Their R^2 values are $0.989, 0.986$ and 0.996%, 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 of this study, it is retained in the model.

3.2. 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, $y_i \in [0, 1]$ are indicator variables, i=$1, \dots, 19$ 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; Gałecki and Burzykowski, 2013) and REML assumes independent observations with homogeneous variance. Table 3 provides the estimated coefficients and p-values for the LMER models. In Table 3, 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).

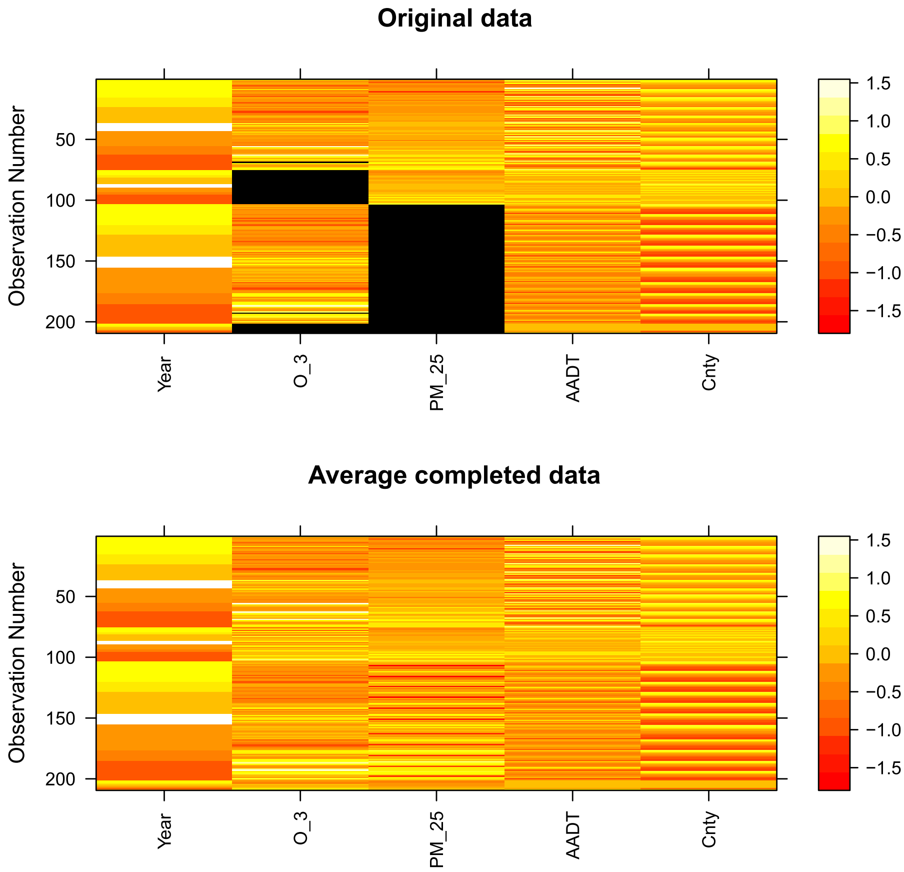


Fig. 2. Imputation of Missing Data Using Gaussian Mixtures.

3.3. Grey systems and its modifications

Grey systems are especially suited for datasets with 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 number of accidents and pavement degradation (Gao et al., 2010; An et al., 2012; Liu et al., 2014).

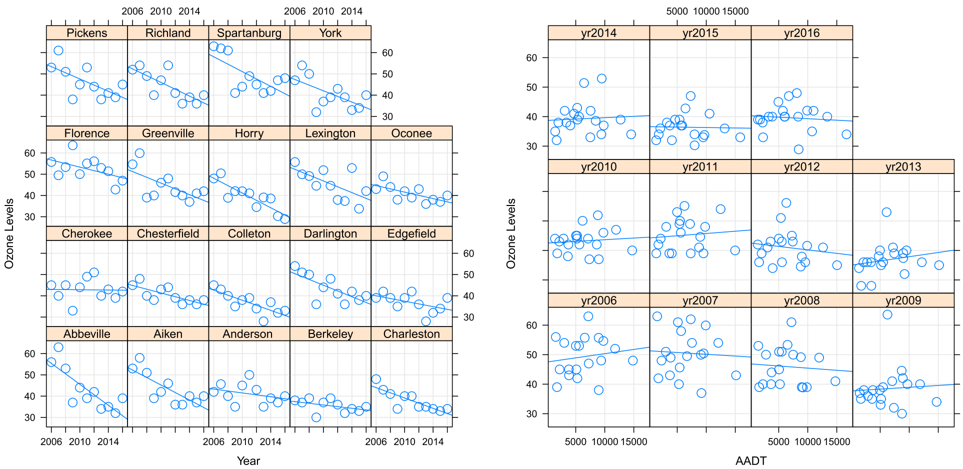
According to the Grey Systems theory, the unknown parameters of the system are represented by discrete or continuous Grey numbers encoded by the symbol $\otimes$. The theory introduces a number of properties and operations on the Grey numbers such as the core of the number $\hat{\otimes}$, its degree of Greyness g°, 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) is briefly discussed here; readers are referred to Deng (1989) for additional information. Suppose that $X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\right)$ denotes a sequence of non-negative observations of a stochastic process and $X^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\right)$ is an accumulation sequence of $X^{(0)}$ computed as in Eq. (3).

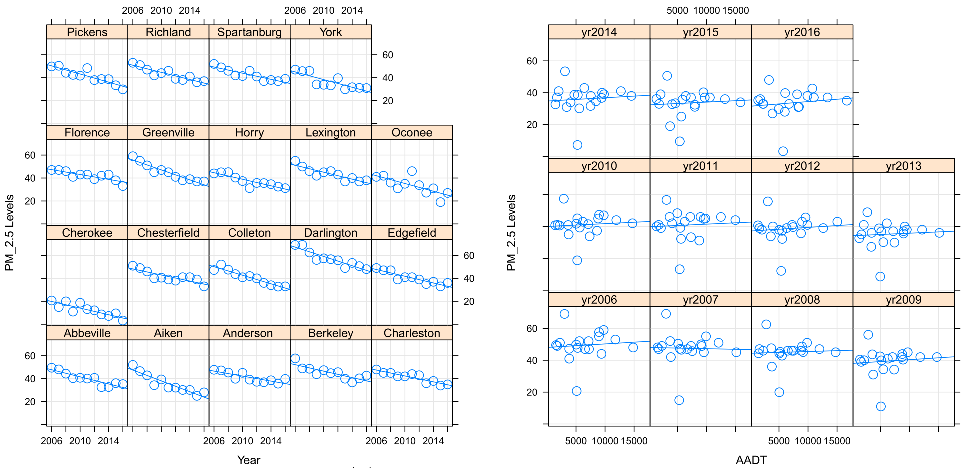
$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \tag{3}$$

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

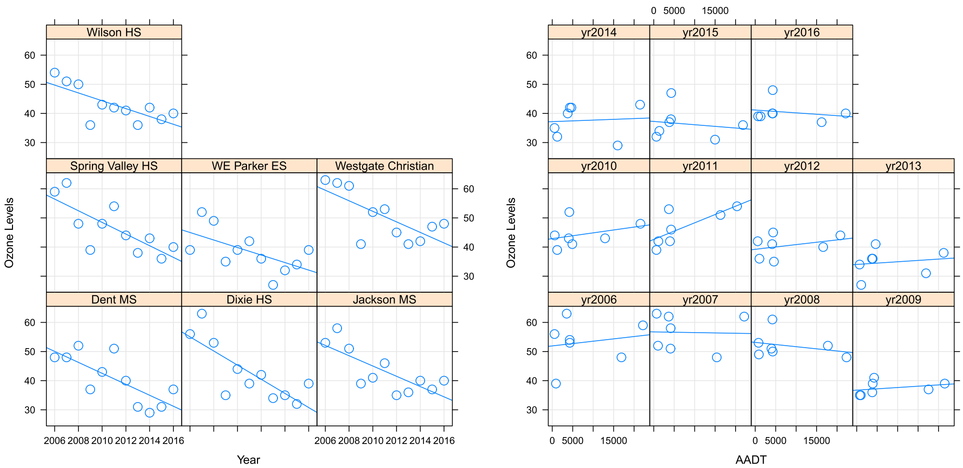
$$x^{(0)}(k) + a x^{(1)}(k) = b \tag{4}$$



(a) O_3 levels for counties



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



(c) O_3 levels for schools

Fig. 3. Correlation matrices of AQIs for different counties, schools, AADT and year.

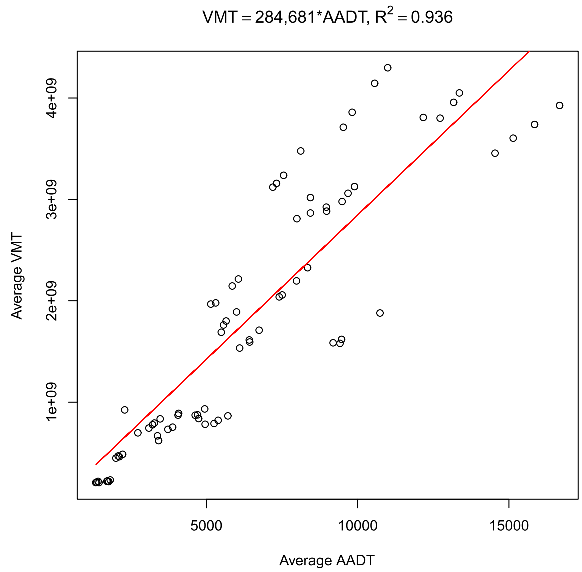


Fig. 4. Correlation between average VMT and average AADT.

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

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

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

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

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

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

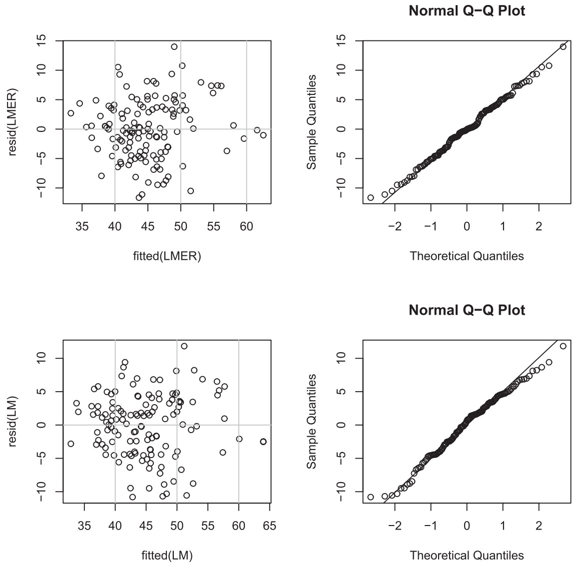
Suppose that $\hat{x}^{(0)}(k)$ and $\hat{x}^{(1)}(k)$ represent the time response sequence (the forecast) and the accumulated time response sequence of GM at time k respectively. Then, the latter can be obtained by solving Eq. (7):

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

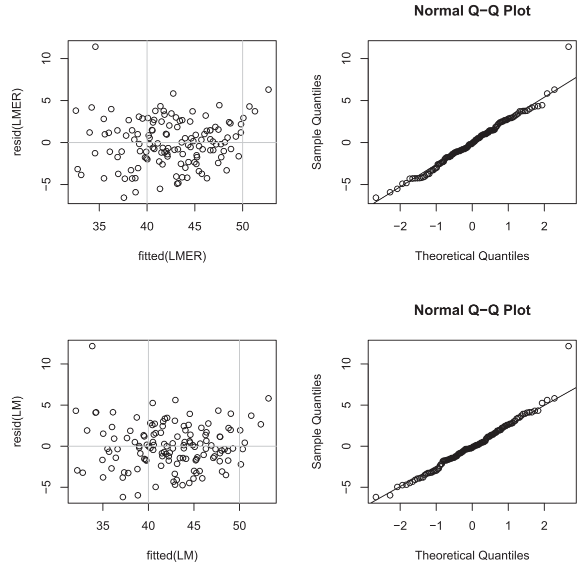
According to the definition in Eq. (3), the restored values of $\hat{x}^{(0)}(k+1)$ are calculated as $\hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$:

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

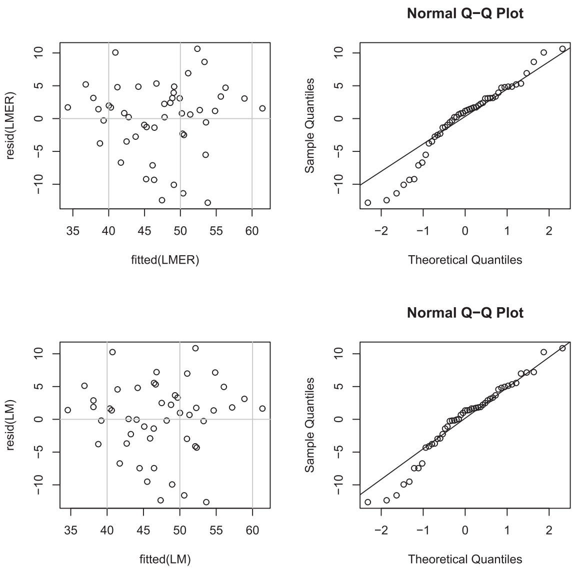
Eq. (9) gives the method to produce forecasts for all k in $2, 3, \ldots, n$. However, for longer time series, a rolling GM is preferred. The rolling model observes a window of a few sequential data points in the series: $x^{(0)}(k+1), x^{(0)}(k+2), \ldots, x^{(0)}(k+w)$, where $w \geqslant 4$ is the 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.



(a) O_3 AQIs counties



(b) $PM_{2.5}$ AQIs counties



(c) O_3 AQIs schools

Fig. 5. Diagnostics for the linear models.

3.4. 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:

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

Table 2
LMs for O_3 and $PM_{2.5}$ AQIs for different counties and schools LM model for O_3 AQIs for different schools.

Variable	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		

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

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

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

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

3.5. Error corrections to Grey models

In order to increase the accuracy of the Grey models, suppose that $\epsilon^{(0)}$=$\epsilon^{(0)}(1), \ldots, \epsilon^{(0)}(n)$ 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 Fourier series (Tan and Chang, 1996) as in Eq. (13).

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

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

The solution is found via the least squares estimate, presuming that $\epsilon^{(0)} \cong PC$ where C is a vector of coefficients: $C = [a_0 a_1 b_1 a_2 \ldots a_n b_n]^T$ and matrix P is:

Table 3
LMER models for O_3 and $PM_{2.5}$ AQIs for different counties LMER model for O_3 AQIs for different schools.

Var.	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			

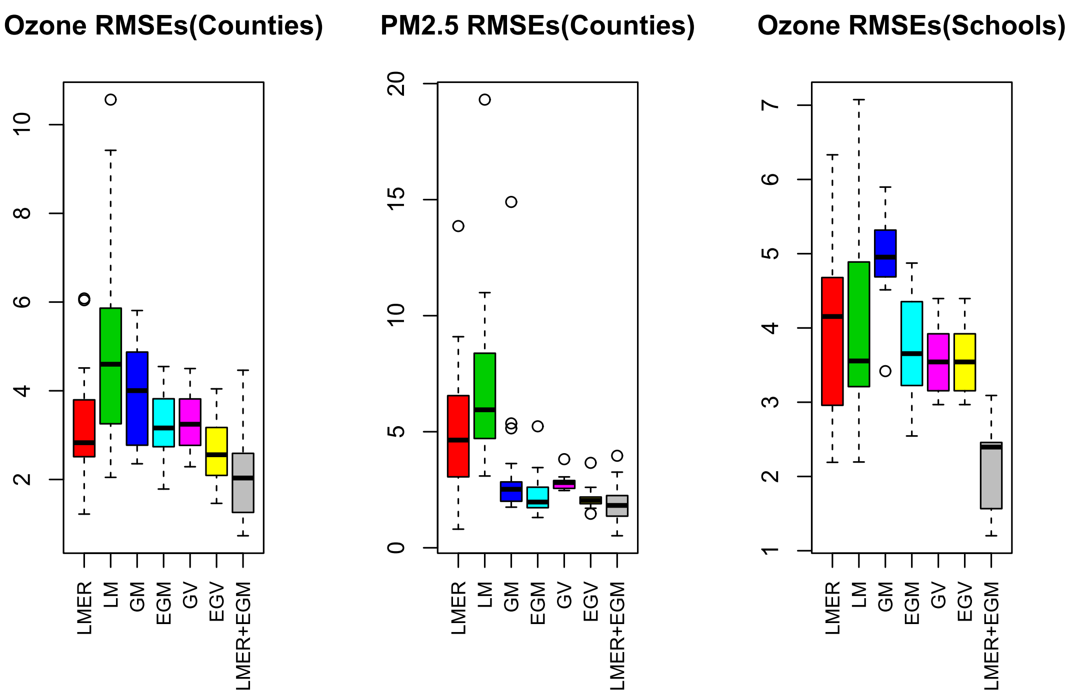
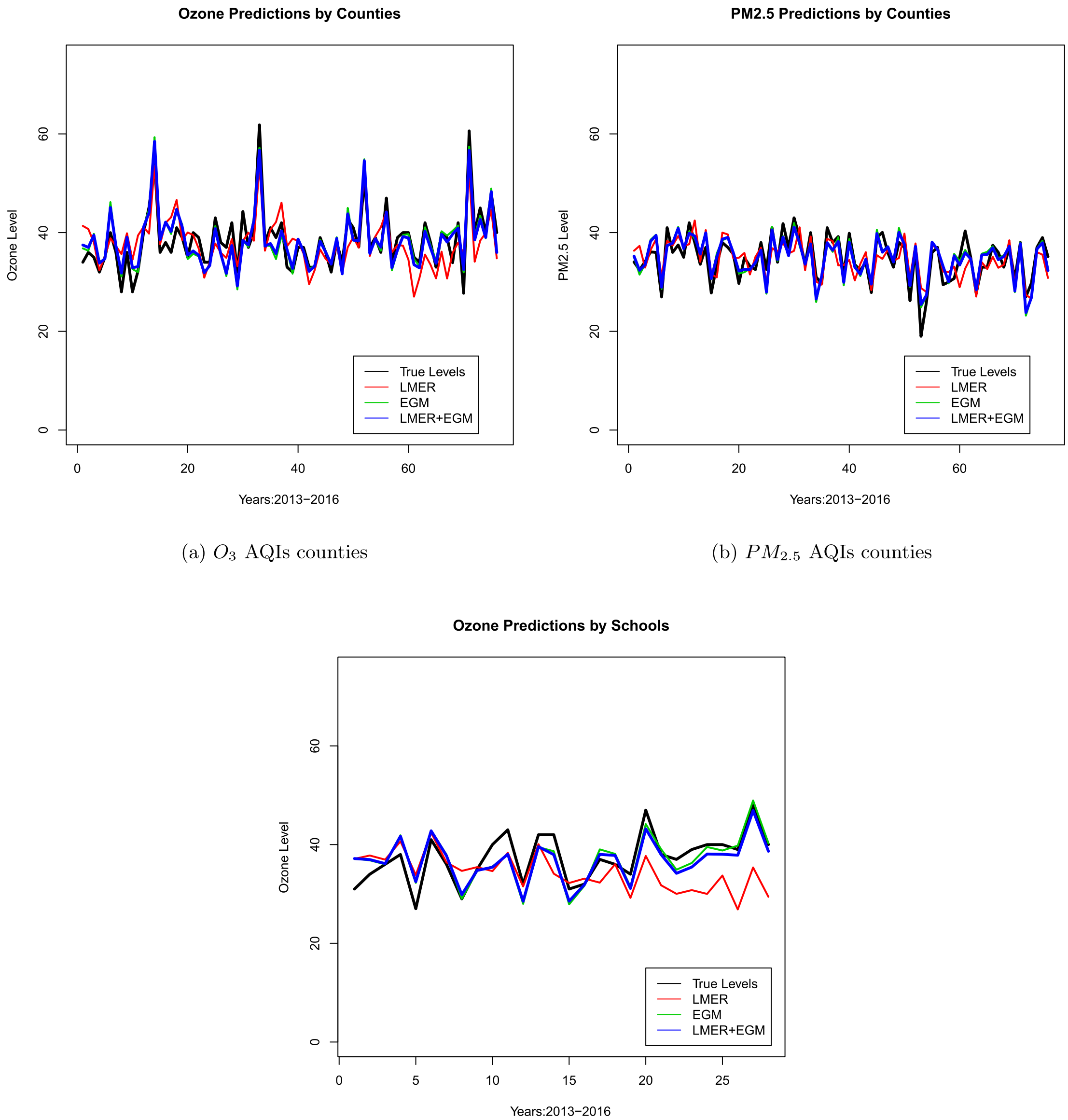


Fig. 6. Prediction errors for 2013–2016 AQIs.

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

4. Modeling results and discussion

This section compares the performance of the linear regression model (LM), LMER, GM, error corrected GM (EGM), GV, error corrected GV (EGV), and LMER + EGM on the validation data set. Average RMSEs for O_3 and $PM_{2.5}$ county-level models are: $[3.2, 5.1, 3.9, 3.3, 3.3, 2.7, 2.1]$ and $[5.2, 7.0, 3.3, 2.3, 2.8, 2.1, 1.9]$. Average RMSEs for O_3 school-level models are $[4.0, 4.1, 4.9, 3.7, 3.6, 3.6, 2.1]$. In each case the highest accuracy was achieved by the combination method. Fig. 6 shows



(a) O_3 AQIs counties (b) $PM_{2.5}$ AQIs counties

(c) O_3 AQIs schools

Fig. 7. Predictions for O_3 and $PM_{2.5}$ levels for different counties.

the RMSEs for the different models in predicting the O_3 and $PM_{2.5}$ levels for counties and 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. This result suggests that perhaps there are correlated residuals in the data. In summary, all models produced estimates within $\pm 10\%$ of true values.

In Fig. 7, 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 produce superior results. In this study, the combined model's weighted forecast is $z_c = \alpha \hat{z}_1 + (1 - alpha)\hat{z}_2$ 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 $\alpha^* = \left(\sum_{t=1}^{T} e_{2t}^2 - \sum_{t=1}^{T} e_{1t}e_{2t}\right) / \left(\sum_{t=1}^{T} e_{1t}^2 + \sum_{t=1}^{T} e_{2t}^2 - 2\sum_{t=1}^{T} e_{1t}e_{2t}\right)$ where $e_{1t}{=}Z_t - \widehat{Z}_{1t}$ and $e_{2t}{=}Z_t - \widehat{Z}_{2t}$ (Newbold and Harvey, 2008). However, in this study, the optimal weight α was empirically derived to be 0.15.

5. Conclusions

This paper developed 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 decrease 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) will have higher levels of O_3 or $PM_{2.5}$ when AADT decreases. This finding suggests that there are other factors, other than AADT, that influence the air quality in these counties and schools. An explanation for this is that 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 EPA's national emissions standards have contributed to air quality improvements since 1990, which enabled many areas of the country to meet standards set to protect public health and the environment. The developed methods can be seen as a step forward in air quality prediction that consider both spatial and temporal factors. These models are important for planning purposes to identify risk areas and to find suitable locations for sensitive facilities such as K-12 schools and hospitals. By knowing which areas are at risk the decision makers can implement countermeasures. There are a number of options to control sources of pollution. From the transportation perspective, the state could seek to implement emission control on vehicles; that is, South Carolina does not require vehicle emission testing. The state could also require the use of cleaner fuel such as California. Future work will focus on developing site-specific models using hourly traffic and air quality measures; high-quality portable air quality sensors will be used.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study is based upon research partially supported by the USDOT Center for Connected Multimodal Mobility (Tier-1 University Transportation Center), U.S. Department of Homeland Security (DHS) Summer Research Team Program Follow-On, and National Science Foundation (NSF, No. 1719501) grants. Any opinions, findings, and conclusions expressed in this paper are solely those of the authors and do not necessarily reflect the views of the USDOT, Center for Connected Multimodal Mobility, DHS, and NSF.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.ijtst.2019.05.008.

References

Abou-Senna, H., Radwan, E., Westerlund, K., Cooper, C.D., 2013. Using a traffic simulation model (vissim) with an emissions model (moves) to predict emissions from vehicles on a limited-access highway. J. Air Waste Manage. Assoc. 63 (7), 819–831.

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. Atmos. Environ. 165, 264–273.

An, Y., Cui, H., Zhao, X., 2012. Exploring gray system in traffic volume prediction based on tall database. Int. J. Digital Content Technol. Appl. 6 (5).

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using LME4. J. Stat. Softw. 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. J. Appl. Econ. 25 (6), 1028–1062.

de Miranda, R.M., Perez-Martinez, P.J., de Fatima Andrade, M., Ribeiro, F.N.D., 2017. Relationship between black carbon (bc) and heavy traffic in são paulo, Brazil. Transp. Res. Part D: Transport Environ.
Deng, J.L., 1989. Introduction to grey system theory. J. Grey Syst. 1 (1), 1–24. URLhttp://dl.acm.org/citation.cfm?id=90757.90758.
Gałecki, 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. 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. Transp. Res. Rec.: J. Transp. Res. Board 1738, 49–55.
Hao, P., Wang, C., et al., 2018. Evaluating environmental impact of traffic congestion in real time based on sparse mobile crowd-sourced data..
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. Environ. Health Perspect. 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. Atmos. Environ. 35 (22), 3875–3884.
Jimenez-Palacios, J.L., 1998. Understanding and quantifying motor vehicle emissions with vehicle specific power and tildas remote sensing. Institute of Technology, Massachusetts.
Ju-Long, D., 1982. Control problems of grey systems. Syst. Control Lett. 1 (5), 288–294.
Kayacan, E., Ulutas, B., Kaynak, O., 2010. Grey system theory-based models in time series prediction. Expert Syst. Appl. 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, p. 68.
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. Environ. Health Perspect. 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 Aff. 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 Air Qual. Res. 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, pp. 268–283. Ch. 12.
Rakowska, A., Wong, K.C., Townsend, T., Chan, K.L., Westerdahl, D., Ng, S., Močnik, G., Drinovec, L., Ning, Z., 2014. Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon. Atmos. Environ. 98, 260–270.
Shaaban, K., Abou-Senna, H., Elnashar, D., Radwan, E., 2019. Assessing the impact of converting roundabouts to traffic signals on vehicle emissions along an urban arterial corridor in qatar. J. Air Waste Manage. Assoc. 69 (2), 178–191.
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.
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.
Xu, X., Liu, H., Anderson, J.M., Xu, Y., Hunter, M.P., Rodgers, M.O., Guensler, R.L., 2016. Estimating project-level vehicle emissions with Vissim and MOVES-matrix. Transp. Res. Rec. 2570 (1), 107–117.
Zhang, K., Batterman, S., 2013. Air pollution and health risks due to vehicle traffic. Sci. Rotal Environ. 450, 307–316.