

Assessment of Alternative Bayesian Hierarchical Models for Estimating Gas Emissions

Gurdiljot Singh Gill and Wen Cheng*

Department of Civil Engineering, California State Polytechnic University, Pomona, USA

*Corresponding author: Wen Cheng, Department of Civil Engineering, California State Polytechnic University, Pomona 3801 W. Temple Ave., Pomona, CA 91768, USA, Tel: 9098692957; Fax: (909) 869-4342; E-mail: wcheng@cpp.edu

Received date: August 18, 2017; Accepted date: September 13, 2017; Published date: September 20, 2017

Copyright: © 2017 Gill GS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Gill GS, Cheng W. Assessment of Alternative Bayesian Hierarchical Models for Estimating Gas Emissions. Glob Environ Health Saf. 2017, Vol. 1 No. 2: 8.

Abstract

This study focused on the comparison of different Bayesian models developed for prediction of two primary air pollutants, ozone and PM_{2.5} emissions. Three alternate models were developed to incorporate different correlation structures: 1) univariate model which served as reference for comparison; 2) univariate spatial model which incorporated the spatial random effects to account for the spatial correlation structures among the TAZs; and 3) multivariate model which addressed the potential correlation among the dependent variables and allowed the simultaneous prediction to generate more precise estimates.

Many socioeconomic variables were observed to be influential such as household density, population, education, and poverty. Such phenomenon indicates the disproportionate impact of Ozone and PM_{2.5} emissions on the specific areas which requires the efforts to emphasize social equity and environmental justice. In terms of factors pertaining to traffic conditions, traffic density was observed to be statistically significant which served as an indicator of vehicular emissions. The univariate spatial model revealed the influence of space for ozone prediction as a significant positive correlation was recorded, which reflected the large amount of variability explained by spatial random effects that may have escaped the explanatory variables. This finding highlighted that, relative to PM_{2.5}, ozone emission models benefit with the inclusion of spatial correlations as such dependency may be more profound.

In terms of model performance at goodness-of-fit, the multivariate model significantly outperformed the others by demonstrating lowest posterior deviance without a notable increase in model complexity suggesting the implementation of joint modeling for PM_{2.5} and ozone prediction. However, the spatial model was observed to be superior based on predictive accuracy which indicated the importance of accommodating the spatial correlation

to account for unobserved heterogeneity and obtain more precise posterior estimates with minimum deviation from observed data.

Keywords: Ozone; PM_{2.5}; Spatial correlations; Multivariate; Goodness-of-fit

Introduction

The emissions from vehicular traffic contribute as a major source of air pollution in developing or developed cities due to ever increasing use of vehicles associated with population growth and industrial development. Vehicular emissions are comprised of a variety of pollutants which have significant long-term impact on air quality and may “threaten human health, damage ecosystems and influence climate” [1]. The general constituents of vehicle emissions are carbon monoxide (CO), carbon dioxide (CO₂), oxides of nitrogen (NO_x), hydrocarbons (HC), volatile organic compounds (VOCs), and polycyclic aromatic hydrocarbons (PAHs) [1]. These also include the particulate matter (PM) and precursors of ozone.

Particulate matter (PM) signifies the solid matter particles suspended in air which are classified based on the aerodynamic diameter: coarser particles (with size ranging from 2.5 to 10 μm, PM₁₀) and fine particles (with sizes up to 2.5 μm, PM_{2.5}). Ambient PM contains numerous carcinogenic and toxic substances, heavy metals and stable quinoid radicals. The evidence supported by a plethora of research studies indicates strong epidemiological association between air pollution (e.g., PM_{2.5}) and adverse health impacts such as respiratory inflammation, cardiovascular morbidity, allergy, asthma attacks, beside other illness [2-7]. The studies focused on PM emissions have revealed their association with roadway traffic as vehicular emissions is a significant source of ambient PM in urban areas mainly due to congested conditions that force stop-and-go behavior [8]. Different factors pertaining to transportation have been investigated for correlation with PM emissions, such as the impact of roadway geometric design [9-11], built environment [12-14], driving behavior and fuel-consumption [15], and so on. The results from the

forementioned studies demonstrated the association of PM emissions and the concerned transportation variables. While developing models for prediction of PM emissions to explore the relationship between air quality and transportation related factors, many studies considered the spatial influence since the local vehicle activity has a major impact on the ambient air pollutants, hence they tend to affect areas where they are emitted [16,17]. The spatial dependency was also noted to be influential for the studies focused on filtering the hazardous sites based on emissions from vehicles at roundabout corridors [18], transportation projects [19], or vehicle speed and traffic intensity [20], demonstrating the significance of inclusion of spatial correlation to generate more precise estimates.

Similar to the adverse impact of PM on human health, many epidemiologic studies have indicated substantial association between ambient ozone concentrations and their adverse impact on the respiratory lung diseases and airway inflammation [21,22]. Similar to the emission particles related to vehicular traffic, ozone is a highly reactive oxidant gas that forms major component of air pollution and forms reactive oxygen species (ROS) by oxidizing important biological molecules, which eventually target the cellular compartments of respiratory tract [23-27]. Chronic effects of ozone exposure on lung function development, asthma incidence, and pulmonary inflammation have been suggested [28]. Ozone has been observed to induce lung tumors through free radical mechanisms and especially by the formation of HO• radicals [29], which may interact with the DNA to cause mutagenic damages [30]. Similar to PM emissions, the research has revealed the association of ozone and traffic activity while also observing the spatial dependency. Granier and Brousseau [31] noted that the emissions by road traffic (passenger vehicles and trucks) of ozone precursors (NO_x, CO, hydrocarbons) have a substantial impact on the concentration of tropospheric ozone at the regional and hemispheric scale, which directly relates to adverse health effects. The study by Liu et al. [32] employed a random-effect linear model for the assessment of ozone exposure in a community and observed that the inclusion of traffic and spatial effects accounted for the variability and improved prediction capability for Ozone. Pont and Fontan [33] also observed the variations in the ozone at the macro level of cities dependent on the changes in vehicular traffic for different days of the week. The study by McConnell et al. [34] investigated the dependency of variations in vehicular traffic on the Ozone deficit of a residential community. The ozone prediction model was developed by estimation of nitrogen oxides from the traffic counts near the entities under consideration. This study corroborated the spatial dependency of the NO_x source on the Ozone prediction. Recently, Wang et al. [35] also observed substantial variations of predicted ozone concentration over the spatial entities. Ibarra-Berastegi et al. [36] also utilized the traffic data for short-term forecasting of ozone and NO₂. However, this study addressed the potential correlation among the ozone and NO₂, since both originate from similar sources, by joint modeling of the concerned dependent variables using multiple linear regression model. It was observed that

variables related to traffic accounted for significant variability and the modeling structure demonstrated improved persistence as assessed by different statistical parameters.

The aforementioned studies pertaining to PM and Ozone considered them as separate environmental pollutants based on the assumption of potential dependency among them, hence developing individual models as separate dependent variables. However, recent studies and epidemiological evidence indicate that, there is a positive association between airborne PM and O₃ and hospital admissions for respiratory diseases. The study by Valavanidis et al. [37] investigated the health impact of the interaction of ozone and particulate matter originated from the vehicular traffic and observed that the combination generates synergistically increasing amounts of hydroxyl radicals (HO•), compared to individual action of O₃ or PM, which eventually poses more serious impact to exposed population.

Albeit the literature review illustrates numerous effort to explore the ozone and PM_{2.5} emissions by employing different approaches, there is a lack of comprehensive research addressing the comparison of different correlation structures and the associated benefits. To this end, the present study develops a multivariate modeling structure for simultaneous prediction of PM_{2.5} and Ozone (give the interaction among them) based on the macro-level covariates comprising of factors pertaining to traffic activity (such as traffic density and volume), multimodal indicators (such as access to bike lanes, pedestrian-friendly facilities, transit), and socioeconomic factors (such as poverty, education, population). Two alternate models are also developed: 1) univariate model which ignores the potential correlation among the dependent variables; 2) univariate spatial model which extends the former model by incorporating the random effects to account for the spatial correlations among entities, which are traffic analysis zones (TAZs) of a city in California in this case. The distance-based spatial correlation among the neighboring TAZs is incorporated to account for the unobserved heterogeneity, which results from non-inclusion of spatial characteristics which may impact air emissions [38,39]. These three models were developed to explore the advantages associated with model fit and prediction accuracy while considering the correlation among dependent variables (multivariate) or impact of space (spatial). The univariate model without spatial represents the traditional approach and acts as a reference. These three models are assessed by employing five evaluation criteria, namely: Dbar (posterior deviance), Pd (model complexity), DIC (deviance information criterion), MSPE (mean square predictive error), and Residual Sum of Squares (RSS). These criteria assess the model performance based on goodness-of-fit, explanation of variability, and predictive accuracy. It is anticipated that this study will provide valuable insights to the transportation and environmental agencies of Southern California to promote transportation and environmental justice.

Methodology

The traditional non-Bayesian techniques such as maximum likelihood and/or least squares estimations usually involve

fixing the values of parameters that have an important bearing on the final outcome of the analysis and for which there is considerable uncertainty. On the contrary, Bayesian analyses can take fuller account of the uncertainties related to models and parameter values. In addition, the Bayesian approach has the ability to incorporate prior information which is based on historical data sets or expert knowledge. Finally, the Bayesian analysis provides a convenient setting for a wide range of models, such as hierarchical models and missing data problems. Markov chain Monte Carlo (MCMC), along with other numerical methods, makes computations tractable for virtually all parametric models [40]. Given the various benefits associated with the Bayesian techniques, the study implements the alternative modeling settings through Bayesian framework through the freeware WinBUGS [41]. The following subsections present the detailed analytics of various models in order.

Model 1: Univariate Model without Spatial Correlations.

In general, the gas emissions can be assumed to follow normal distribution which gives rise to the following formulation:

$$\begin{aligned} y_i | X_i &\sim N(\mu, \tau) \\ \mu_i &= \beta_0 + \beta X_i \\ \tau^{-1} &\sim \text{gamma}(0.001, 0.001) \dots \dots (1) \end{aligned}$$

Where y_i represents the concerned gas emissions (Ozone or PM2.5) for zone i ($i=203$); X_i is the vector of independent variables for zone i ; μ is the vector representing posterior mean of the particular gas emissions; τ is the variance whose reciprocal is presumed to follow a gamma distribution; β is the covariate coefficients; β_0 is the intercept representing the base condition, which is assumed to follow a normal distribution as usual.

Model 2: Univariate Model with Spatial Correlations.

This model differs from Model 1 in the way of addressing the spatially structured heterogeneity. Part of the above formulation can be modified as follows:

$$\mu_i = \beta_0 + \beta X_i + \phi_i \dots \dots (2)$$

Where ϕ_i is the spatial random effect which takes on the CAR (conditional autoregressive) correlation:

$$\phi_i | \phi_{k \in N_i} \sim N_j(\sum_{k \in N_i} c_{ik} \phi_k, \sum_i) \dots \dots (3)$$

As shown from the above equation, estimation of the particular emission in any zone is conditional on emissions in neighboring zones. Subscripts i and k refer to a TAZ and its neighbor, respectively, and k belongs to N_i , where N_i represents the set of neighbors of TAZ i . Besides the identification of neighbors, the assigned weights also affect the concerned gas estimations. The present study employed the popularly used distance-based structure where the weight between TAZ i and j is inversely related with the distance between the zone pairs. With this weight structure, it is known that all zones are considered as neighbors to each other, and the TAZs which are relatively closer would have more weights.

Once ϕ_i is estimated, it is also interesting to calculate the percentage of gas emission variability that is due to spatial clustering:

$$\alpha = \text{sd}(\phi_i) / \text{sd}(\mu_i) \dots \dots (4)$$

Where α is defined as fraction, and sd is the marginal standard deviation function. The larger the fraction value, the more variability explained by the spatially structured random effects.

Model 3: Multivariate Model

This model is distinct from previous models in the sense that it jointly estimates the emissions of Ozone and PM2.5, rather than analyze them separately. The resultant formulation is expressed as below:

$$\begin{aligned} y_{ij} | X_{ij} &\sim \text{MVN}(\mu_j, \Sigma) \\ \mu_{ij} &= \beta_{0rj} + \beta_j X_{ij} \\ \Sigma^{-1} &\sim \text{Wishart}(R, n) \dots \dots (5) \end{aligned}$$

Where y_{ij} represents the gas emission for zone i ($i=203$) and type j ($j=2$); X_{ij} is the matrix for predictor variables; μ_j is the vector representing posterior mean of the gas emissions (Ozone and PM2.5); β_j is the corresponding variable coefficients; Σ^{-1} is a symmetric positive definite precision with the scale matrix R and degrees of freedom matrix n ($=2$). The interested readers can refer to the details of Wishart distribution through previous research.

Modeling Evaluation

In order for assessment of modeling performance from different perspectives, alternative goodness-of-fit (GOF) measures were utilized for modeling comparison which are presented in the following subsections.

GOF Measure 1: Deviance Information Criterion (DIC) and its Components

Similar to the Bayesian equivalent of the Akaike Information Criterion (AIC) DIC is also a panelized measure which can be expressed in the equation below:

$$DIC = \bar{D} + P_D \dots \dots (6)$$

Where, \bar{D} represents the posterior mean of the deviance statistic and P_D denotes the effective number of parameters in the model. The smaller the DIC, the better fitness of the model tends to be. In general, the modeling deviance could be reduced with more effective parameters being included. Therefore, the DIC criterion compensates the deviance with the model complexity being taken into consideration. In terms of the guideline suggested by Spiegelhalter et al. a difference of 7 or more points in the DIC is considered significant for modeling performance.

GOF Measure 2: Mean Square Predictive Error (MSPE)

MSPE assumes the form shown as below:

$$MSPE = \frac{1}{n} \sum_{i=1}^n (y_i - o_i)^2 \dots \dots (7)$$

Where y_i is the Bayesian-estimated gas emission of Zone i and O_i is the observed one for the same zone. The smaller the MSPE value, the better fitness to the data.

GOF Measure 3: Chi-squared Residual Sum of Squares (RSS)

RSS is defined as:

$$RSS = \sum_{i=1}^n \frac{(o_i - y_i)^2}{y_i} \dots\dots(8)$$

Where y_i and O_i are as defined previously. Under MSPE, the larger zones are expected to subject to more deviances due to larger areas and gas emissions. RSS tends to remove such bias by calculating the squared residual relative to estimated amount of gas emissions. A particular model is considered more reliable if smaller RSS value is observed.

Data Preparation

This transportation planning-level study analyzed the average annual ozone and PM_{2.5} emissions from the 203 TAZs in the city of Irvine, California. The variables used for model development and the associated descriptive statistics are shown in **Table 1**. As evident, diverse variables were incorporated pertaining to transportation, air-quality, and socioeconomic factors. PM_{2.5} and ozone were considered as the dependent variables and modeled jointly in case of multivariate model. The primary source of air quality data was the Office of Environmental Health Hazard Assessment (OEHHA) while few variables were collected from Air Quality Management District. These data sources served best for the purpose of this study as they focus on the collection of explanatory variables which serve as indicators for more precise reflection of socioeconomic vulnerability to environmental pollutants. Satellite data was incorporated to provide full state coverage for the PM_{2.5} indicator. The main data sources for PM_{2.5} were Air Monitoring Network and California Air Resources Board (CARB). For all measurements in the time period, the mean concentrations were estimated at the geographic center of the census tract using a geostatistical method that incorporates the monitoring data from nearby monitors. Another air pollutant which poses widespread and significant health threat is ozone. The data were mostly obtained from CARB, along with other sources such as local air pollution control districts, tribes and federal land managers. The indicator value for ozone was computed as a mean of daily maximum 8-hour ozone concentration (ppm) recorded over

the summer months (May-October) and averaged over three years (2012 to 2014).

As previously discussed in literature review, the vehicular emissions were regarded as the primary source of ozone and PM_{2.5} emissions and Traffic Density served as the primary indicator of the vehicle exhaust gases. Since the air emissions also indirectly get mitigated by the pedestrian and bicycle activity, possibly in place of vehicular activity, hence the indicator variables which pertain to presence of sidewalks, crosswalks, bicycle lanes and other factors, were considered such as: Bike lane access, Bike lane density, Walk accessibility. In terms of socioeconomic factors, education and poverty for the area residents were incorporated. Indicator for the education shows the percent of the population over age 25 with less than a high school education (5-year estimate, 2011-2015) while the poverty indicator shows the percent of the population living below two times the federal poverty level (5-year estimate, 2011-2015). They both act as a social determinant of awareness and health, where education is often inversely related to the degree of exposure to indoor and outdoor pollution while less impoverished populations are relatively prone to adverse health outcomes when exposed to environmental pollution. Shape file of TAZ boundary and the transportation or socioeconomic TAZ characteristics were provided by SCAG (Southern California Association of Governments). In addition, the distance matrix containing distances among various TAZ centroids were also collected from SCAG for the estimation of distance-based spatial random effect. Since there are 203 TAZs in the city, the matrix includes 203 × 202 distances. Their descriptive statistics can be found in **Table 1** as well.

As evident from **Table 1**, many variables indicated the possibility of being correlated. To obtain a parsimonious model and ensure that non-correlated variables were entered for model development, the correlation analysis was performed on the multiple covariates using the Harrell Miscellaneous package in R software which allowed the calculation of Pearson correlation coefficient. The variables observed to be correlated at a significance level of 0.05 were eliminated in multiple steps using engineering judgment to prevent exclusion of any potential influential variables which would result in loss of precision of estimated parameters. The final variables selected for development of models are shown in **Table 3**.

Table 1 Summary Statistics of Variables for TAZ's of the City of Irvine.

Variables	Description	Mean	SD	Min.	Max.
DVMT	Daily vehicle miles traveled	5,4262.44	56,156.84	112.57	276,079.90
Acre	TAZ Area (acre)	282.90	431.75	0.69	5,062.95
Median	Median house income (\$)	48,440.78	50,635.10	0	183,347
Pop_den	Population density by area (persons/acre)	6.18	7.96	0	32.40
HH_den	Household density (hh/acre)	2.34	3.15	0	13.62

Emp_den	Employment density (jobs/acre)	10.34	17.43	0	121.10
Ret_den	Retail job density	0.79	2.02	0	17.45
% age 5_17	% of population age 5-17	8.64%	8.78%	0	27%
% age 18_24	% of population age 18-24	5.79%	7.42%	0	40%
% age 24_64	% of population age 24-64	38.35%	36.12%	0	95%
% age 65+	% of population age 65 or older	6.25%	10.21%	0	83%
K12	K12 student enrollment (thousand)	0.39	1.00	0	5.52
College	College student enrollment	0.11	1.00	0	12.59
Int34_den	Intersection density (3- and 4- legs)	0.12	0.12	0	0.62
BKInACC	Bike lane access (1=if a TAZ has bike lane)	0.92	0.28	0	1
BL_den	Bike lane density	3.40	1.80	0	7.26
Rail	1=at least one rail station in a TAZ	0.01	0.10	0	1
TTbus_D	Total Bus Stop Density (stops per mile)	0.05	0.09	0	0.53
Exbus_D	Stop density for Express Bus and BRT	0.002	0.007	0	0.06
HFLbus_D	High-Frequency Bus Stop Density (local bus headway <= 20 mins)	0.001	0.004	0	0.03
WalkAcc	Walk Accessibility	3.87	9.46	0	74.53
% Arterial	Percent of main arterial (45-55 mph) of TAZ	10.61%	17.33%	0	80%
RetSer_den	Retail Service Job Density	2.98	6.29	0	0.27
Jobmix13	Job Mix (13 Sectors)	0.543	0.282	0	0.93
Pct_Art	Percent of main arterial	0.106	0.173	0.	0.8
Mlt_pct	% of households living in multiple unit	0.120	0.169	0	0.5
HQTA_pct	% of TAZ area are in non-freeway HQTA	0.197	0.36	0	1
TPA_pct	% of TAZ area are in TPA	0.02	0.10	0	0.84
BLdenIND	Bike Lane Density Indicator	3.40	1.79	0	7.26
Population	Average Population of TAZ (thousand)	18.777	10.447	0.549	50.588
Education	Population over 25 age with less than high school education	-15.531	170.358	-995.8	105.4
Poverty	Population living below two times federal poverty level	51.758	38.011	4.7	227
Pestic	Pesticide Use	2485.9	4347.1	0	20318.8
Toxic Release	Toxicity Weighted Concentration	52713.6	59649.2	4447.8	263112.6
Traffic Density	Number of vehicles in a specific area	3178.1	1770.2	509.8	10493.1
Haz Wa	Hazardous waste	1.110	1.291	0	8.25
Solid	Solid Waste Sites	8.498	6.82	0	27.45
Pollution	Contamination of air	15.92	6.82	4.6	39.71
PM 2.5	Annual mean concentration of Particulate Matter 2.5 (μ g/m ³)	23.86	11.88	8.7	76.74
O ₃	Average daily maximum 8-hour Ozone concentration (ppm)	0.12	0.05	0.046	0.37
Distance	Distance among TAZ centroids (miles)	4.06	2.09	0.16	11.78

Results

This study employed the freeware statistical package WinBUGS for development of three models for the prediction of PM_{2.5} and ozone emissions at the macro-level of TAZ. For assessment of advantages associated with different correlation structures, the univariate model without spatial random effects (Model 1) was developed to represent the traditional approach. The univariate spatial (Model 2) and multivariate (Model 3) models addressed the presence of potential correlation structures among the concerned entities and dependent variables, respectively. A total of 11,000 MCMC iterations were utilized for parameter estimation after

discarding first 1,000 iterations as burn-in. Two chains were set up for each model starting from diverse initial values. The MCMC convergence was ensured by recording MC errors to be lesser than 5% of associated standard deviations and visual inspection of history plots, trace plots, and Gelman-Rubin diagram. As shown in **Table 2**, the three models demonstrated varying computational effort. The inclusion of correlation structures was observed to be directly associated with the running time as the multivariate model took twice the time while the a substantial increase in computational complexity was observed for the spatial model with a 25 second difference compared to the Base univariate.

Table 2 Running time for alternate models. **Note:** All of them run 11000 iterations with the first 1000 as burn in; 2-chains.

Model	Running time (every 1000 iterations)
1: univariate no spatial	2 seconds
2: univariate spatial	27 seconds
3: multivariate no spatial	4 seconds

Model variable estimates

As shown in **Table 3**, the three models demonstrated robustness as implied from the similar set of influential factors at the significance level of 0.05. most of the socioeconomic variables were observed to be significant such as population, education, and poverty. Although the coefficients for PM_{2.5} and ozone were different but the dependency of these factors illustrates that the vehicular emissions tend to have higher exposure for the TAZs with relatively higher population; poorer inhabitants; and less likely to be educated. The disproportionate impact of vehicle emissions on such areas conveys the need to emphasize social equity and environmental justice for the vulnerable section of society. As reflected from the coefficients of model estimates, PM_{2.5} emissions tend to have stronger correlations with such factors

as compared to ozone pollution. In terms of factors pertaining to traffic conditions and built environment, vehicular traffic density and area of TAZ were observed to be significant. Previous studies have also noted the positive dependency of traffic density and PM_{2.5} emissions/Ozone as vehicular traffic is regarded as a primary source of such pollution and traffic density serves as an indicator of vehicular presence in an area. The provision of walkability tends to lower traffic activity which helps mitigate vehicular emissions. Interestingly, the spatial correlation was observed to be statistically significant only for the ozone which indicates the explanatory variables, which are common for Ozone and PM_{2.5} emissions, were able to sufficiently account for the space-related heterogeneity in case of PM_{2.5} but the spatial random effects, captured the unobserved heterogeneity which escaped the covariates.

Table 3 Posterior Inference for Ozone and PM_{2.5} Emissions. Numbers in parentheses represent uncertainty estimates, or, posterior standard deviations. The statistically significant variable coefficients are shown in bold. α represents the emission variability explained by the spatial structurally heterogeneity.

Count Type	Variables	Model 1	Model 2	Model 3
Ozone	Intercept	0.017 (0.011)	0.014 (0.011)	0.019 (0.013)
	Acre	2.3E-5 (5.5E-6)	2.3E-5 (6.0E-6)	2.3E-5 (6.7E-6)
	HH_den	2.9E-4 (8.1E-4)	3.4E-4 (7.9E-4)	3.2E-4 (0.001)
	Emp_den	9.8E-4 (0.009)	5.9E-4 (0.008)	8.8E-4 (0.010)
	BKlnACC	-5.1E-4 (0.009)	-1.2E-4 (0.009)	-0.001 (0.010)
	WalkAcc	-5.4E-4 (2.4E-4)	-4.5E-4 (2.4E-4)	-5.2E-4 (2.9E-4)
	% Arterial	-0.006 (0.014)	-0.005 (0.014)	-0.006 (0.017)
	Exbus_D	0.133 (0.339)	0.095 (0.328)	0.114 (0.292)
	HFLbus_D	0.532 (0.563)	0.519 (0.536)	-0.060 (0.458)

	Population	0.002 (2.7E-4)	0.002 (2.7E-4)	0.002 (3.3E-4)
	Traffic_D	1.1E-5 (1.6E-6)	1.1E-5(1.6E-6)	1.2E-5 (2.0E-6)
	Education	-6.6E-5 (1.3E-5)	-6.5E-5(1.3E-5)	-6.6E-5 (1.6E-5)
	Poverty	4.2E-4 (7.9E-5)	4.4E-4(7.9E-5)	4.1E-4 (1.0E-4)
	α	N/A	0.336 (0.068)	N/A
PM2.5	Intercept	2.797 (1.839)	2.763 (1.845)	2.645 (1.790)
	Acre	0.004 (9.8E-4)	0.004 (9.9E-4)	0.004 (0.001)
	HH_den	0.076 (0.145)	0.076 (0.145)	0.081 (0.143)
	Emp_den	0.061 (1.509)	0.071 (1.501)	0.077 (1.492)
	BKlnACC	-0.511 (1.489)	-0.496 (1.482)	-0.430 (1.461)
	WalkAcc	-0.082 (0.043)	-0.081 (0.044)	-0.080 (0.043)
	% Arterial	-1.155 (2.483)	-1.147 (2.488)	-1.100 (2.454)
	Exbus_D	8.370 (27.570)	8.526 (27.680)	8.650 (27.480)
	HFLbus_D	13.000 (29.930)	12.870 (29.830)	12.860 (29.830)
	Population	0.448 (0.047)	0.448 (0.047)	0.449 (0.047)
	Traffic_D	0.002 (2.9E-4)	0.002 (2.9E-4)	0.002 (3.0E-4)
	Education	-0.013 (0.002)	-0.013 (0.002)	-0.013 (0.002)
	Poverty	0.098 (0.014)	0.098 (0.014)	0.098 (0.015)
	α	N/A	9.4E-5 (1.4E-4)	N/A

Modeling comparison

For comparison of alternate models based on goodness-of-fit, the penalized criterion of DIC was employed. It accounts for the complexity associated with the number of effective parameters used for model development which generally tends to improve model fit. As shown in **Table 4**, Model 3 was observed to have the best overall fit with the lowest DIC of 197, with a remarkable difference of 259 and 274 points from the second-best (Model 2) and worst model (Model 1), respectively. While comparing the models developed from the same sample size, the difference of 10 points in DIC may be regarded as substantial, which reinforces the significant advantage of multivariate approach to fit the air pollution data. Compared to the univariate model, the correlated error terms in case of multivariate model were noted to increase the complexity due to larger number of effective parameters (Pd difference of 6 points) but the raised complexity was compensated with exceptional fit, as demonstrated by difference of 281 points in posterior deviance. As evident, the governing factor for overall fit among univariate and multivariate model was the significant difference of posterior deviance, where the joint estimation of both air pollutants proved highly beneficial to accommodate the data. These results clearly establish that due to the similarities among the ozone and PM2.5 emissions, the potential bias due to correlation should be accounted for by joint modeling. In terms of the impact of spatial random effects, the highest complexity was observed for Model 2 which may be attributed to the inclusion of much larger number of effective parameters

for model development as the distance-based spatial matrix with a dimension of 203×202 was incorporated to account for the spatial correlations among 203 TAZs. The benefit of incorporation of spatial correlations was extended to better overall fit (DIC=456) due to reduced posterior deviance (relative to Model 1) but inferior performance compared to the multivariate model which exhibited significantly better fit with half the complexity. Understandably, Model 1 exhibited the lowest complexity due to the absence of correlation structures but demonstrated highest posterior deviance that eventually resulted in worst overall fit with highest DIC.

The subsequent step of this study was to compare the alternate models based on the predictive accuracy from different perspectives. Similar to the trend observed for the goodness-of-fit, Model 3 was observed to have the least discrepancy between model predicted and observed emissions, as indicated by the lowest MAD, followed by Model 2 and Model 1. However, in case of RSS, the univariate spatial model (Model 2) was observed to be superior. It is worth recalling that unlike MSPE, RSS accounts for the bias induced due to variations in size of the TAZs. The superior performance of Model 2 seems to indicate that although the inclusion of spatial random effects raises model complexity, but they account for the variability in prediction of air pollutants by addressing the space-related unobserved heterogeneity that may have escaped the explanatory variables. This finding reflects the need for evaluation of model estimates for assessment of predictive performance since the goodness-of-

fit may not be correlated with equivalent performance at emission prediction.

Table 4 Evaluation performance for alternate models. **Note:** the best performance under each criterion is shown in bold.

Model	Dbar	Pd	DIC	MSP E	RSS
1: univariate no spatial	445.26 6	26.33 7	471.71 0	15.08 3	268.46 0
2: univariate spatial	393.08 1	63.08 7	456.16 8	15.08 2	267.58 0
3: multivariate no spatial	164.85 5	32.73 5	197.59 0	15.08 1	268.98 6

Conclusions

This study focused on the comparison of different Bayesian models developed for prediction of two primary air pollutants. The ozone and PM_{2.5} emissions were modeled as the dependent variables for the aggregated data at the transportation planning level of Traffic Analysis Zones while the explanatory variables comprised of various transportation and socioeconomic factors. Three alternate models were developed to incorporate different correlation structures: 1) univariate model which served as reference for comparison; 2) univariate spatial model which incorporated the spatial random effects to account for the correlation structures among the TAZs; and 3) multivariate model which addressed the potential correlation among the dependent variables and allowed the simultaneous prediction to generate more precise estimates. The alternate models were assessed based on the performance at goodness-of-fit and predictive accuracy by employing five evaluation criteria.

In terms of model estimates, all three models were able to identify mostly similar influential factors which indicates the robustness of the different model specifications employed. Many socioeconomic variables were observed to be influential such as household density, population, education, and poverty, which portrays the disproportionate impact of Ozone and PM_{2.5} emissions on the specific areas which requires the efforts to emphasize social equity and environmental justice. In terms of factors pertaining to traffic conditions, traffic density was observed to be statistically significant which was expected considering it served as an indicator of vehicular activity and eventually emissions. The univariate spatial model revealed the influence of space for ozone prediction as a significant positive correlation was recorded. A substantially larger coefficient for the spatial component reflects the large amount of variability explained by the spatially structured random effects, which may have escaped the explanatory variables (unobserved heterogeneity). This finding highlighted that, relative to PM_{2.5}, ozone emission models benefit with the inclusion of spatial correlations as such dependency may be more profound.

In terms of model performance at goodness-of-fit, the multivariate model significantly outperformed the others by demonstrating lowest posterior deviance without a notable

increase in model complexity. However, the spatial model was observed to employ a very large number of effective parameters which increased the computational effort due to complexity but was not followed with an equivalent reduction in posterior deviance. Overall, the model fit results revealed the significant advantage of multivariate approach to fit the air pollution data and suggested the implementation of joint modeling for PM_{2.5} and ozone prediction due to potential similarities. The governing factor for overall fit among the models was established to be the significant differences of posterior deviance while the spatial correlation proved inferior due to raised complexity. However, the spatial model was observed to be superior based on predictive accuracy, as assessed by RSS, which indicated the importance of accommodating the spatial correlation to account for unobserved heterogeneity that may have escaped the explanatory variables.

References

- Sharma AR, Kharol SK, Badarinath KVS (2010) Influence of vehicular traffic on urban air quality-A case study of Hyderabad India. *Transportation Research Part D: Transport and Environment* 15: 154-159.
- Health Effects Institute (2010) *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*.
- Atkinson R, Mills I, Walton H, Anderson HR (2014) Fine particle components and health-a 1 systematic review and meta-analysis of epidemiological time series studies of daily mortality and hospital admission. *Journal of Exposure Science and Environmental Epidemiology*, pp: 1-7.
- Li N, Hao MQ, Phalen RF, Hinds WC, Nel AE (2003) Particulate air pollutants and asthmas A paradigm for the role of oxidative stress in PM-induced adverse health effects. *Clin Immunol* 109: 250-265.
- Cormack MCM, Breyse PN, Matsui EC, Hansel NN, Peng RD, et al. (2011) Indoor particulate matter increases asthma 8 morbidity in children with non-atopic and atopic asthma. *Ann Allergy Asthma Immunol* 106: 308-315.
- Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, et al. (1993) An Association between Air Pollution and Mortality in Six US Cities. *New England Journal of Medication* 329: 1753-1759.
- World Health Organization (2013) *Health Effects of Particulate Matter Final Report*, pp:17-20.
- Joo S, Oh C, Lee S, Lee G (2017) Assessing the Impact of Traffic Crashes on Near Freeway Air Quality (No. 17-01567).
- Boriboonsomsin K, Barth M (2009) Impacts of Road Grade on Fuel Consumption and Carbon Dioxide Emissions Evidenced by Use of Advanced Navigation Systems. *Journal of the Transportation Research Board*, pp. 21-30.
- Park S, Rakha H (2006) Energy and Environmental Impacts of Roadway Grades. *Journal of the Transportation Research Board* 41: 148-160.
- Pérez ZAM, Llopis CD, Camacho TFF, Belkacem I, García A (2017) Impact of Horizontal Geometric Design of Two-Lane Rural Roads on Vehicle CO₂ Emissions (No. 17-04284).

12. Frank LD, Stone B, Bachman W (2000) Linking land use with household vehicle emissions in the central puget sound: methodological framework and findings. *Transport Res Part D: Transp Environ* 5: 173-196.
13. Hong J, Shen Q (2013) Residential density and transportation emissions: examining the connection by addressing spatial autocorrelation and self-selection. *Transport Res Part D: Transp Environ* 22: 75-79.
14. Choi K, Zhang M (2017) The net effects of the built environment on household vehicle emissions: A case study of Austin, TX. *Transportation Research Part D: Transport and Environment* 50: 254-268.
15. Shawarby IEI, Ahn K, Rakha H (2005) Comparative field evaluation of vehicle 30 cruise speed and acceleration level impacts on hot stabilized emissions. *Transport Res Part D: Transp Environ* 10: 13-30.
16. Colville RN, Hutchinson EJ, Mindell JS, Warren RF (2001) The Transport sector as a source of air pollution. *Atmospheric Environment* 35: 1537-1565.
17. Parra MA, Elustondo D, Bermejo R, Santamaria JM (2009) Ambient air levels of volatile organic compounds and nitrogen oxide(NO₂) in a medium size city in Northern Spain. *Science of The Total Environment* 407: 999-1009.
18. Fernandes P, Salamati K, Roupail NM, Coelho MC (2015) Identification of emission hotspots in roundabouts corridors. *Transportation Research Part D: Transport and Environment* 37 : 48-64.
19. Qiu Z, Li X (2016) Quantitative Particulate Matter Hot-spot Analysis for a Transportation Project in China. In *Transportation Research Board 95th Annual Meeting (No. 16-2588)*.
20. Prada FP, Monzon A (2017) Identifying Traffic Emissions Hotspots for Urban Air Quality Interventions: The Case of Madrid City (No. 17-05015).
21. Mudway IS, Kelly FJ (2000) Ozone and the lung: a sensitive issue. *Mol Aspects Med* 1: 1-48.
22. Thurston GD, Ito K (2001) Epidemiological studies of acute ozone exposures and mortality. *J Expo Anal Environ Epidemiol* 11: 286-294.
23. Delinger B, Pryor WA, Cueto R, Squadrito GL, Hedge GLV, et al. (2001) Role of free radicals in the toxicity of airborne fine particulate Matter. *Chem Res Toxicol* 14: 1371-1377.
24. Churg A, Brauer M (2000) Ambient atmospheric particles in the airways of human lungs. *Ultrastruct Pathol* 24: 353-361.
25. Kley D, Kleinmann M, Sanderman H, Krupa S (1999) Photochemical oxidants: state of the science. *Environ Pollut* 100: 19-42.
26. Pryor WA (1994) Mechanisms of radical formation from reactions of ozone with target molecules in the lung. *Free Radic Biol Med* 17: 451-465.
27. Bassett D, Elbon CC, Otterbein S, Barraclough MH, Delorme M, et al. (2001) Inflammatory cell availability affects ozone-induced lung damage. *J Toxicol Environ Health Part A* 64: 547-565.
28. Environmental Protection Agency (2013) Integrated science assessment for Ozone and Related Photochemical Oxidants.
29. Victorin K (1992) Review of the genotoxicity of ozone. *Mutat Res* 277: 221-238.
30. Ito K, Inoue S, Hiraku Y, Kawanishi S (2005) Mechanism of site-specific DNA damage induced by ozone. *Mutat Res* 585: 60-70.
31. Granier C, Brasseur GP (2003) The impact of road traffic on global tropospheric ozone. *Geophysical Research Letters* 30.
32. Liu LJ, Delfino R, Koutrakis P (1997) Ozone exposure assessment in a southern California community. *Environmental Health Perspectives* 105: 58.
33. Pont V, Fontan J (2001) Comparison between weekend and weekday ozone concentration in large cities in France. *Atmospheric Environment* 35: 1527-1535.
34. Connell RM, Berhane K, Yao L, Lurmann FW, Avol E, et al. (2006) Predicting residential ozone deficits from nearby traffic. *Science of the Total Environment* 363: 166-174.
35. Wang M, Keller JP, Adar SD, Kim SY, Larson TV, et al. (2015) Development of long-term spatiotemporal models for ambient ozone in six metropolitan regions of the United States: The MESA Air study. *Atmospheric Environment* 123: 79-87.
36. Ibarra BG, Madariaga I, Agirre E, Uria J (2003) Short-term forecasting of ozone and NO₂ levels using traffic data in Bilbao (Spain). *WIT Transactions on The Built Environment* 64.
37. Valavanidis A, Loridas S, Vlahogianni T, Fiotakis K (2009) Influence of ozone on traffic-related particulate matter on the generation of hydroxyl radicals through a heterogeneous synergistic effect. *Journal of hazardous materials* 162: 886-892.
38. Wang H, Dong Y, Zhang K (2017) A spatial-temporal model to improve PM_{2.5} inference. In *Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference IEEE*, pp: 173-177.
39. Fecht D, Hansell AL, Morley D, Dajnak D, Vienneau D et al. (2016) Spatial and temporal associations of road traffic noise and air pollution in London: Implications for epidemiological studies. *Environment International* 88: 235-242.
40. Congdon P (2014) *Applied bayesian modelling*. John Wiley & Sons 595.
41. Spiegelhalter D, Thomas A, Best N, Lunn D (2003) *WinBUGS User Manual*.