Transportation Research Part A 116 (2018) 220–231

Contents lists available at ScienceDirect

Transportation Research Part A

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

Modeling the endogeneity of lane-mean speeds and lane-speed deviations using a Bayesian structural equations approach with spatial correlation

Wen Cheng*, Gurdiljot Singh Gill, Taha Sakrani, Dennis Ralls, Xudong Jia

Department of Civil Engineering, California State Polytechnic University, Pomona, 3801 W. Temple Ave., Pomona, CA 91768, United States

A R T I C L E I N F O

Keywords:
Random effects
Spatial correlation
Multivariate
Multilane
Endogeneity
Speed prediction

A B S T R A C T

This study focused on the development of speed prediction models for a multilane highway which incorporate the potential endogenous relationship between adjacent lane speeds and speed deviations while considering geometric design, traffic flow, and other variables in the model specification and accounting for the correlation structures due to multilevel nature of data. The Full Bayesian framework was employed to build the hierarchical models which accounted for three correlation structures at multiple levels: the correlation between speeds of adjoining lanes due to multivariate nature; spatially structured correlations between the adjacent segments, and spatially unstructured correlations among segments.

The model estimates which influence the lane-mean speed indicated the directional variation of exogenous factors. For the westbound traffic, the afternoon and night hours were observed to be influential while eastbound traffic was more sensitive to the morning period. The segment length revealed a lane-varying correlation where longer segments influenced a speed reduction for the three lanes closer to the median while the speed in outermost lane exhibited a reverse trend as it increased with longer segments. Both models, with mean speed and speed deviations, demonstrated the significant presence of endogeneity due to mean speeds and speed deviations of adjacent lanes, respectively. This study also assessed the accuracy of predicted mean speed and speed deviations by calculating the measures of discrepancy 差异 between the observed and model predicted speeds. The Bayesian residuals, which incorporated the normal, multivariate, and spatial correlation structures, exhibited significant superiority at the prediction accuracy than the Normal ones. This discrepancy in prediction performance reflected the impact of consideration or exclusion of random effects.

1. Introduction

In the field of transportation engineering, the geometric design and level of service of roadways are greatly influenced by the operational speed. Speed plays a vital role in traffic management and control as it is fundamental to the performance measures of traffic. Due to the general tendency of unavailability of speed distribution information of road entities, it has been best estimated in terms of its basic relationship with other fundamental variables of the road entity such as density or flow. The basis of this approach was laid by the seminal research studies from several decades ago (Greenshields, 1935; Lighthill and Whitham, 1955) which established the relationships between flow, density, and space mean speed. To construct a superior function of speed-flow relationship,

* Corresponding author.
E-mail addresses: wcheng@cpp.edu (W. Cheng), gurdiljotg@cpp.edu (G.S. Gill), tsakrani@cpp.edu (T. Sakrani), dbralls@cpp.edu (D. Ralls), xjia@cpp.edu (X. Jia).

https://doi.org/10.1016/j.tra.2018.06.014
Received 15 May 2017; Received in revised form 4 March 2018; Accepted 20 June 2018
0965-8564/

some studies incorporated different explanatory variables which could potentially influence the vehicular speed such as an array of environment conditions and traffic characteristics. The study by Fitzpatrick et al. (2005) on two-lane tangent sections developed the operational speed model and considered roadway and roadside characteristics such as number of access points, roadside development type, lane width, and others. The study by Einbeck and Dwyer (2011) proposed the use of local principal curves to model the speed-flow relationship. Following the variable selection recommended by previous studies focused on flow theories (Shankar et al. 1995; Li et al. 2009; Golob and Recker, 2003), this study included variables pertaining to roadway conditions such as speed limit, ramp and signal control, and speed distribution. Pei et al. (2012) employed Full Bayesian method to establish a prediction model for speed distribution, in terms of average travel speed and speed deviation, by considering the effects of traffic flow, road geometry, and weather conditions. A study by Silvano and Bang (2015) focused on the impact of posted speed limit (PSL) on the free-flow speed on urban roads by incorporating the road characteristics, such as carriageway width, road environments, and the presence of on-street parking and sidewalks. As noted by Bassani et al. (2016), the recent research on speed prediction models has focused on improving the model prediction capability, unlike the previous studies which aimed to investigate the influential factors.

Apart from the aforementioned studies which were centered on the development of speed prediction models with different exogenous variables, several studies related with other transportation modeling context have utilized endogeneity models to address the simultaneity problems common to transportation field (Derrig et al., 2000; Eisenberg, 2003; Cheng et al., 2013; Washington et al., 2010). The exogenous and endogenous variables differ on the basis of dependency between the independent and dependent variables in a model. Majority of models assume that the causal relationship between such variables is unidirectional, which means that only the independent variables could impact the dependent ones and not the other way around. However, the transportation field poses different simultaneity issues where the dependent variable could also impact the independent ones. For example, in case of safety analysis of intersections crashes, different covariates may be considered which may impact the dependent variable of crash occurrence. One such factor is the presence of left-turn lanes at the intersection. However, the introduction of left-turn lanes on an intersection is also influenced by the perceived safety concern. This builds up an endogeneity problem where the left-turn lane could
If you perceive something, you see, notice, or realize it, especially when it is not obvious.
not be assumed to be influenced by crash occurrence (dependent variable) (Kim and Washington, 2006). Different studies have attempted to circumvent the issue of endogeneity bias in different aspects of transportation research (Dane et al., 2014). The study by
If someone circumvents a rule or restriction, they avoid having to obey the rule or restriction, in a clever and perhaps dishonest way
Bhat and Koppelman (1993) used the Simultaneous Equations Model (SEM) approach to develop an integrated model of employment, income, and household car ownership. They regarded employment and income as endogenous variables in disaggregate travel de-
to separate from a group or mass
mand frameworks and estimated a joint model of employment, income, and car ownership. Abu-Eisheh (2001) developed a vehicle demand and driver population models by employing simultaneous equation estimation techniques as a function of socioeconomic and political variables and observed simultaneity among the dependent variables of automobile ownership and driver population. Kim and Washington (2006) noted that inclusion of endogeneity revealed the unbiased effect of left-turn lanes on crash frequency as left-turn lanes were observed to reduce angle crash frequencies as expected by engineering judgment. Medina and Tarko (2007) explored the relationship between roadway characteristics, driver behavior as influenced by the perceived risk, and safety by formulating an SEM that consists of two structural equations involving two endogenous variables, the objective risk and the speed selected by the drivers.

Some studies incorporated the SEM formulation to address the endogeneity issue for speed prediction models. Shankar and Mannering (1998) investigated the role of endogeneity at a macroscopic level by developing models for lane-mean speeds and speed deviations on a three-lane highway. Due to the linear nature of functional form, three-stage least squares (3SLS) estimation was
lane-mean speeds
employed for model estimation. They found the presence of endogeneity in the form of variables pertaining to ~~lean-mean speed~~ and
If one thing pertains to another, it relates, belongs, or applies to it.
speed deviations of the adjacent lanes while exogenous variables related to the environment, traffic flow, and temporal factors were noted to be influential for lane speeds. Boyle and Mannering (2004) utilized the same 3SLS approach to explore the effect of driver behavior for in- and out-of-vehicle travel advisory systems. The results showed the negative endogenous relationship between mean
An advisory is an official announcement or report that warns people about bad weather, diseases, or other dangers or problems.
speed and speed deviation as various driver-related variables and advisory message indicators were found to be associated with speed and speed deviation. The study by Himes and Donnell (2010) focused on the investigation of geometric design and traffic flow factors to model the mean operating speed and speed deviation on four-lane highways by using 3SLS estimator while exploring the endogeneity and incorporating the contemporaneous correlation between the disturbances across the equations. This study recommended the exploration of more geometric variables (shoulder width, lane width) and focus on areas with more substantial traffic flow environments.

Similar to the above studies which controlled the endogeneity to reduce bias and inconsistency of estimated parameters, some studies aimed to develop more precise models by incorporating different types of correlation structures found in transportation data. Some studies employed the multivariate approach for simultaneously modeling different crash outcomes based on accident type or severity (Aguero-Valverde and Jovanis, 2009; Cheng et al., 2017b; Cheng et al., 2018b) while other studies expanded to explore the spatial correlations (Jonathan et al., 2016; Cheng et al., 2017a; Cheng et al., 2018c). Such correlations are typical in studies of crash prediction models but rare in lane speed models. Recently, a study by Bassani et al. (2016) utilized random effects to account for the variability in the time and space of data for the estimation of free-flow speed on two-lane rural highways. The adopted model structure separated the estimate of the central tendency of speeds from the typical deviations of individual speeds. Unlike the aforementioned crash prediction models, the random effects in this study were normally distributed and hence it did not consider any multivariate or spatial correlations.

In summary, few multilane highway speed prediction models have been estimated. Of those, some have included traffic flow, seasonal, and time-of-day variables in the model specification using a simultaneous equations framework. Other multilane speed prediction models considered only inside and outside lane mean speeds, or estimated a model of mean speed and speed deviation without considering interaction of speed across travel lanes. There is a need to estimate a multilane highway speed prediction model that considers the potential endogenous relationship between adjacent lane speeds and speed deviations while considering geometric design, traffic flow, and other variables in the model specification while accounting for the multilevel nature of data. The hierarchical structure of the models is important to explore the different types of heterogeneity for the observations in the data at different levels. This study incorporates the similar nature of speed data for all observations within a segment, correlation between speeds of adjoining lanes due to multivariate nature, and spatial correlations between the adjacent segments as the upstream vehicles carry the traffic characteristics to downstream segments, and hence spatial correlations need to be accounted for the speed prediction models. This study attempts to develop speed prediction models which combine the impacts on speed from various covariates and evaluate the prediction accuracy of the models for different scenarios which examines the benefit of incorporation of such random effects.

2. Methodology

This study aimed to develop a model for the mean speed and speed deviations for an 8-lane roadway section (Interstate-10) in California and gain a better understanding of the influential factors. For model development, several exogenous factors were considered pertaining to roadway geometry, traffic flow, time of the day, seasons, the day of the week, among others. However, the past research (Shankar and Mannering, 1998) has observed the influence of endogenous variables while modeling the mean speed or speed deviations of each lane. Hence, it was imperative 必要的 to simultaneously model the mean speed and speed deviations of each lane, and account for the endogenous nature of mean speed and deviations of adjacent lanes. This study also incorporated three different correlations at multiple levels. First, as adopted by Bassani et al. (2016), random effects were incorporated to remove the potential bias associated with the dependency between errors of estimated parameters for individual observations, which were comprised of multiple data points within a particular segment. Since the adjacent lanes are essentially a group of lanes belong to a particular segment, the speed of vehicles on such lanes may be interdependent. To account for such correlation among the outcome variables (mean speed or speed deviations of lanes in this study), a multivariate correlation matrix was introduced in the model at the individual point level. Second, at the higher hierarchical level, or, the segment one, a spatial correlation was established between the neighboring segments. This inclusion was an effort to account for the unobserved factors which are associated with space. As the roadway section (I-10) of this study was comprised of adjoining segments, hence such entities may share some unobserved characteristics which are not accounted by the explanatory variables (Lord and Mannering, 2010). Such interdependencies were incorporated into the model to mitigate different types of potential bias in the parameter inferences. Third, in addition to the structured spatial random effects, there is also unstructured random effect accounting for the unobserved heterogeneity at the segment level. Fig. 1 illustrates the multilevel correlations in the data structure of the model.

2.1. Model specification

The full Bayesian (FB) framework was employed to model the mean speed and speed deviations of the lanes while incorporating the aforementioned correlation structures. As noted by another study on speed model development (Pei et al., 2012), the FB framework is preferable in such cases as its flexibility and effective approach renders the capability to incorporate complex correlations with a hierarchical structure of data. FB approach provides a posterior distribution of parameters from Markov-chain Monte Carlo (MCMC) simulation which samples the variables as random, unlike the point estimates generated by the traditional approach of maximum likelihood estimation. This approach has been widely used for crash prediction models due to the multilevel and correlated nature of data (Cheng et al., 2017a). The general functional form of the model is given in the subsections that follow:

Model 1: For lane-mean speeds

This model assumes the lane-mean speeds to be the dependent variable with a normal distribution. The normality of the speed data has been documented by previous research (Lindeman and Ranft, 1978; Fitzpatrick and Collins, 2000).

$$\lambda_{ijm} \sim Normal(S_{ijm}) \tag{1}$$

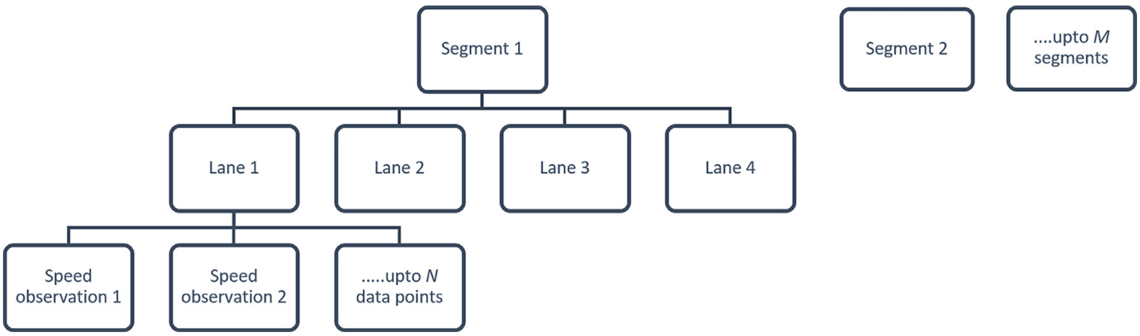

Fig. 1. Multilevel data structure.

Let λ_{ijm}denote the observed lane-mean speed at point i on lane j, at segment m, S_{ijm} is the Bayesian mean expected lane-mean speed, where the lane-mean speed obeys normal distribution.

Lane-mean speed was modeled as a function of covariates and random effects, as shown in Eq. (2):

$$S_{ijm} = \beta_0 + \beta_k X_k + \alpha_k Y_k + \varepsilon_{ijm} + u_m + \phi_m \tag{2}$$

where $\beta_0$is the vector of intercepts, β_k is the vector of independent coefficients, X_k is the vector of independent exogenous covariates, α_k is the vector of independent coefficients, Y_k is the vector of independent endogenous covariates, ε_{ij} is the error term for multivariate correlations among four lanes for each observation i, u_m is the spatially unstructured error term for interdependence of segment m, and ϕ_m represents the spatially structured correlation among segment m. The spatial error term ϕ_m was fit by the conditional autoregressive (CAR) model which accounted for the adjacent segments with binary weights (one and zero). For example, for three consecutive segments named A, B, and C, the weight between segments A and B will be one while for A and C will be zero. The precision of CAR model followed a gamma distribution (refer to Gill et al., 2017). The error term of multivariate correlation was modeled with following distribution:

$$\varepsilon_{ijm} \sim Normal(0, \textstyle\sum)\tag{3}$$

Where

$$\varepsilon_{ijm} = \begin{pmatrix} \varepsilon_{im}^1 \\ \varepsilon_{im}^2 \\ \varepsilon_{im}^3 \\ \varepsilon_{im}^4 \end{pmatrix}, \sum \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{14} \\ \vdots & \ddots & \vdots \\ \sigma_{41} & \cdots & \sigma_{44} \end{pmatrix}\tag{4}$$

$\sum$ is the covariance of the normal distribution for ε_{ijm}. The diagonal element σ_{jj} in the covariance matrix of Eq. (4) represents the variance of ε_{ij}, where the off-diagonal elements represent the covariance of different lane-speeds. The inverse of the covariance matrix represent the precision matrix and has the following distribution:

$$\textstyle\sum^{-1} \sim Wishart(I, J)\tag{5}$$

where I is the J × J identity matrix (Congdon, 2006), and J is the degree of freedom, J = 4.

Model 2: For lane speed deviations

Like some of the previous studies (Shankar and Mannering, 1998; Boyle and Mannering, 2004), in addition to lane-mean speeds, this study also aims to model the lane speed deviations. The speed deviation refers to the difference in upstream and downstream vehicular speeds in the specific lane. The lane-mean speeds influence the speed deviations and hence they both should be modeled as endogenous variables (Shankar and Mannering, 1998). This study incorporates the lane-mean speed as an endogenous variable in the model formulation as follows:

$$D_{ijm} = \beta_0 + \beta_k X_k + \gamma_k Z_k + \varepsilon_{ijm} + u_m + \phi_m \tag{6}$$

where D_{ijm} is the standard deviation of speed in lane j for the observation i, at segment m, γ_k is the vector of independent coefficients, Z_k is the vector of independent covariates which represents the lane-mean speed of lane j, and rest of the symbols are as defined previously.

2.2. Model predictive accuracy

Apart from the identification of influential factors pertaining to lean-mean speed and speed deviations, the speed prediction models are also developed for the objective of accurate estimation of speed. In this study, MAD (Mean Absolute Deviation) and MSPE (mean-squared predictive error) were calculated to assess the prediction capabilities of the models. These criteria are expected to quantify the variability in the prediction performance of models, which will signify the importance of inclusion of different types of random effects. To distinguish the impact of random effects on prediction capability, the models in Eqs. (2) and (6) may be modified as follows:

$$PS_{ijm} = \beta_0 + \beta_k X_k + \alpha_k Y_k \tag{7}$$

$$PD_{ijm} = \beta_0 + \beta_k X_k + \gamma_k Z_k \tag{8}$$

where PS_{ijm} and PD_{ijm} are regarded as the model predicted lane-mean speed and speed deviation, respectively, without accounting for the different types of heterogeneity among the data.

The aforementioned assessment criteria essentially form the basis of residual analysis. In the first case, the normal residuals are calculated as the difference between the model fitted values (PS_{ijm} and PD_{ijm}) and the observed lane-mean speed and speed deviations, respectively. In the second case, the Bayesian residuals are obtained by comparing the observed data with the Bayesian-estimated values (S_{ijm} and D_{ijm}), which account for the random effects (Lawson et al., 2003). The MSPE was calculated as follows:

$$\text{MSPE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - O_i)^2 \tag{9}$$

whereY_i is the model fitted values with and without considering random effects at i by a model, and O_i is the observed one at i by the same model at the same time period. The smaller value of MSPE indicates that the discrepancy between the predicted and observed data values is relatively small and hence that model has better prediction accuracy.

Another similar assessment criterion is MAD. The MAD is defined as:

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - O_i| \tag{10}$$

A relatively low value of MAD and MSPE is preferred which reflects that the model estimated a value which deviates less from the observed reading and hence the model has superiority at prediction accuracy.

3. Data description

The past research endeavors 尽力 have focused on the development of speed prediction models from various perspectives. Many studies investigated the relationship between vehicle speed and various roadway, traffic, or environment features. As the two-lane roadways comprise a significant portion of US roadway structure, most of the studies were dedicated to such roadway classification while limited research was focused on multilane roadways. This study supplements the two-lane studies and contributes to the existing literature of multilane studies by developing speed prediction models for a four-lane roadway. For easier reference in the subsequent text, Lane 1 represents the lane closest to the median in the direction of travel while Lane 4 is the farthest.

The area under the focus of this study was a 44-mile segment on the mainline part of Interstate 10 (I-10) freeway in the Los Angeles County of California (Fig. 2). This segment was selected as it had a continuous division of four lanes in each direction (East and West) and the data were readily available from operational loop detectors for the year 2013. The real-time traffic data from magnetic loop detectors were used to model the lane-mean speed and speed deviations, which are recorded by the Performance Measurement System (PeMS) maintained by the California Department of Transportation (Caltrans). The same data source was used to obtain the certain roadway geometric factors which were expected to impact the speed and deviations on roadway lanes. To ensure that the concerned traffic data were collected for normal conditions (crash-free times), the crash data were obtained from the Transportation Injury Mapping System (TIMS, 2017) database and based on the crash time and milepost, suitable traffic data were extracted from the 5-minute raw detector file. From time's perspective, the data within the time interval of two-to-one hours prior to crash occurrence were extracted, which ensured that traffic data were not biased due to congestion problems from a crash. However, it is worth mentioning that this study intends to develop speed prediction models for normal flow conditions, not necessarily free-flow conditions, and as such the flow data with low speeds were also incorporated for model development. The only flow data excluded from the analysis was concerning the congestion resulting from crash occurrence. As this study also incorporated the spatial

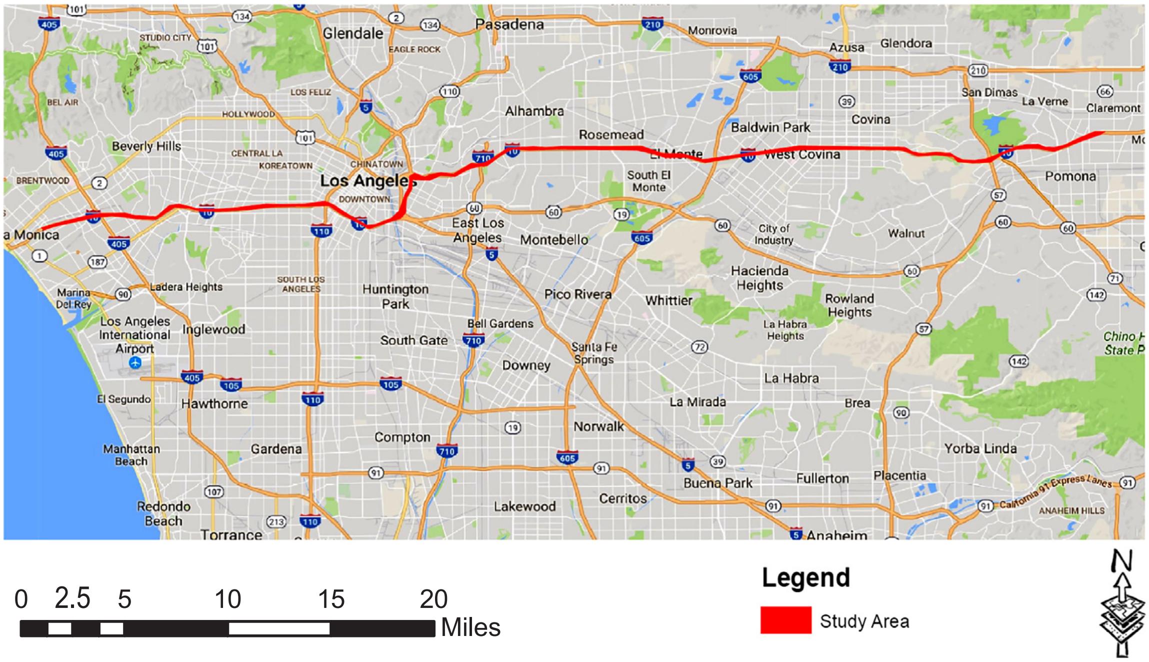

Fig. 2. The study area location.

Table 1
Variables Considered for the Models.

Factor	Variables	Minimum, Mean, Maximum
Traffic data	Speed Lane 1	3, 58.55, 84.8 (mph)
	Speed Lane 2	7.9, 56.31, 81 (mph)
	Speed Lane 3	4.8, 49.53, 77.4 (mph)
	Speed Lane 4	4.1, 47.2, 72.8 (mph)
	Speed Deviation Lane 1	0.07, 6.48, 27.13 (mph)
	Speed Deviation Lane 2	0.1, 6.23, 25.97 (mph)
	Speed Deviation Lane 3	0.14, 5.99, 22.18 (mph)
	Speed Deviation Lane 4	0.39, 6.78, 20.66 (mph)
	Total Flow	13, 383, 806 (veh/h)
Geometry	Inner Shoulder Width	0, 6.9, 10 (Ft)
	Outer Shoulder Width	0, 7.9, 13 (Ft)
	Road Surface Width	36, 50.9, 84 (Ft)
	Segment Length	0.11, 0.4, 0.88 (Miles)
Time of Year (dummy)	Winter	0, 0.23, 1
	Spring	0, 0.3, 1
	Summer	0, 0.26, 1
	Fall (reference)	0, 0.18, 1
Time of Week (dummy)	Monday	0, 0.14, 1
	Tuesday	0, 0.1, 1
	Wednesday	0, 0.12, 1
	Thursday	0, 0.15, 1
	Friday	0, 0.18, 1
	Saturday	0, 0.14, 1
	Sunday (reference)	0, 0.14, 1
Time of Day (dummy)	AM peak	0, 0.24, 1
	PM peak	0, 0.47, 1
	Night	0, 0.17, 1
	Early Morning (reference)	0, 0.1, 1

correlation between segments, from space's perspective, the corresponding station IDs for consecutive upstream and downstream detectors were first obtained, and then the geometric data for the location of loops and their mutual distance were collected from TIMS. Moreover, given that incorrect traffic data might be obtained due to malfunctioning of loop detectors, the data were double checked and invalid or unusable data were excluded if they satisfied infeasible conditions such as the average occupancy was greater than 100%, the vehicle count was greater than 0 veh/30 s while the occupancy was 0%, the occupancy was greater than 0% while the vehicle count was equal to 0 veh/30 s, and so on. Table 1 illustrates the variables incorporated for model development.

As evident from the variables in Table 1, the data sources had limited robustness in terms of variety of variables. The dataset was virtually enhanced by employing dummy variables. The authors acknowledge this limitation, which has also been experienced by existing literature using similar dataset, but hope that the model specification of random effects would compensate the unavailability of data to some degree. Considering the explanatory variables, the geometric factors of shoulder widths, roadway width, and segment length were incorporated in the model. The segment length is essentially the distance between the loop detectors, whose placement and frequency per mile are designated by Caltrans. For a given roadway handling two-way traffic, the number of loop detectors in each direction could be different, even though the roadway length is equivalent. Since the number of loop detectors may have a discrepancy for two directions of a roadway, the number of segments and their lengths may also be different while maintaining the equivalency for overall roadway length for both directions. Given that, the authors strived to keep the roadway length for both directions equal, which resulted in unequal number of segments for the two directions. For the 44-mile roadway stretch, this study had 58 segments for the East direction and 73 segments for West. Considering the difference in the number of loop detectors in each direction, the number of speed data points were also different as the loop detectors generated the flow data. Given that, the East direction had 331 data points for the speed for all four lanes while for the West direction had a larger dataset of 605 data points.

4. Results

Two sets of models were developed for the prediction of lane-mean speed and speed deviations of four lanes. For each set of models, data were available for two directions (East and West) and hence the coefficients for explanatory variables were estimated separately, as shown in Tables 2 and 3. The models were run using a freeware WinBUGS package (Spiegelhalter et al., 2003) employing an MCMC algorithm. For calibration of the model, first 20,000 iterations were discarded as burn-in and further 40,000 iterations were regarded for parameter estimation. To ensure convergence of models, the sample MC errors were recorded to be less than 4% of the associated standard deviation. Visual inspection of the trace plot was conducted as well.

4.1. Lane-mean speed

Table 2 illustrates the estimated coefficients for the exogenous and endogenous variables for both directions of travel. To restate,

Table 2
Parameter estimates for mean-speeds.

Direction East					Direction West				
Variable	mean	SD	2.50%	97.50%	Variable	mean	SD	2.50%	97.50%
Lane-mean speed for lane 1 (dependent)					*Lane-mean speed for lane 1 (dependent)*				
Intercept	−111.8	10.22	−132	−92.24	Intercept	−74.93	7.266	−89.51	−61
Spring	6.296	2.616	0.8272	11.2	Wednesday	2.658	1.348	0.0976	5.351
AM peak	7.301	3.764	0.1544	15.17	PM peak	4.399	1.492	1.435	7.371
Segment Length	−17	5.409	−28.11	−6.727	Night	3.522	1.487	0.62	6.461
Total flow	−0.021	0.008	−0.0383	−0.0043	Total flow	−0.014	0.003	−0.0206	−0.0087
Log of lane 2 speed	−42.72	2.389	−47.67	−38.33	Log of lane 2 speed	−42.72	2.389	−47.67	−38.33
Log of lane 3 speed	69.04	2.641	64.04	74.31	Log of lane 3 speed	69.04	2.641	64.04	74.31
Log of lane 4 speed	21.16	1.484	18.4	24.23	Log of lane 4 speed	21.16	1.484	18.4	24.23
Lane-mean speed for lane 2 (dependent)					*Lane-mean speed for lane 2 (dependent)*				
Intercept	−92.44	8.722	−109.7	−75.65	Intercept	−64.7	6.897	−78.19	−50.85
Wednesday	−5.89	2.836	−11.56	−0.4543	Outer Shoulder width	−0.39	0.164	−0.7194	−0.069
Friday	−5.241	2.653	−10.67	−0.1448	Road Surface Width	−0.232	0.077	−0.3893	−0.0844
AM peak	7.183	3.183	1.175	13.83	PM peak	−3.876	1.487	−6.806	−0.9257
Segment Length	−13.97	4.627	−23.52	−5.11	Segment Length	−13.21	3.274	−19.59	−6.664
Total flow	−0.018	0.007	−0.0324	−0.0035	Log of lane 1 speed	−28.71	2.112	−32.71	−24.54
Log of lane 1 speed	−28.71	2.112	−32.71	−24.54	Log of lane 3 speed	53.57	2.262	49.1	57.97
Log of lane 3 speed	53.57	2.262	49.1	57.97	Log of lane 4 speed	17.28	1.219	14.95	19.7
Log of lane 4 speed	17.28	1.219	14.95	19.7					
Lane-mean speed for lane 3 (dependent)					*Lane-mean speed for lane 3 (dependent)*				
Intercept	−47.07	6.491	−59.64	−34.12	Intercept	−44.37	7.26	−58.45	−29.7
Winter	−5.461	1.463	−8.455	−2.666	Inner Shoulder width	−0.375	0.167	−0.7022	−0.047
Spring	−4.774	1.435	−7.486	−1.873	Outer Shoulder width	−0.586	0.168	−0.9247	−0.256
Summer	−4.145	1.372	−6.73	−1.363	Road Surface Width	−0.232	0.08	−0.3937	−0.0805
Log of lane 1 speed	24.07	2.136	19.96	28.29	Thursday	−3.248	1.358	−5.795	−0.4731
Log of lane 2 speed	11.59	2.011	7.64	15.49	PM peak	−3.301	1.561	−6.377	−0.1929
Log of lane 4 speed	−11.44	0.923	−13.32	−9.708	Segment Length	−16.91	3.362	−23.49	−10.24
					Log of lane 1 speed	24.07	2.136	19.96	28.29
					Log of lane 2 speed	11.59	2.011	7.64	15.49
					Log of lane 4 speed	−11.44	0.923	−13.32	−9.708
Lane-mean speed for lane 4 (dependent)					*Lane-mean speed for lane 4 (dependent)*				
Intercept	−37.23	13.79	−63.8	−9.746	Intercept	−55.94	6.267	−68.34	−43.82
Winter	−7.334	3.84	−15.3	−0.0056	PM peak	3.9	1.208	1.512	6.29
Spring	−9.128	3.728	−15.93	−1.295	Night	3.447	1.192	1.078	5.776
Segment Length	17.77	7.45	3.638	32.74	Total flow	−0.01	0.002	−0.0149	−0.0051
Log of lane 1 speed	58.7	4.277	50.45	67.32	Log of lane 1 speed	58.7	4.277	50.45	67.32
Log of lane 2 speed	56.1	4.292	48	64.72	Log of lane 2 speed	56.1	4.292	48	64.72
Log of lane 3 speed	−98.13	5.958	−109.8	−86.84	Log of lane 3 speed	−98.13	5.958	−109.8	−86.84
SD (ϕ)	0.3334	0.5092	0.00121	1.693	SD (ϕ)	0.4145	0.7113	0.00156	2.139
SD (u)	0.7672	0.8441	5.22E−4	2.142	SD (*PleaseCheck*	1.603	1.291	5.83E−4	4.047
SD (ε)	15.9	0.3356	15.26	16.56	SD (ε)	11.84	6.421	7.347	34.41

Notes: 1. Only the variables with 5% level of significance are shown in the table.
2. SD refers to Standard Deviation
3. ϕ refers to spatial correlation between segments; u refers to random error between segments; ε refers to multivariate error between lanes

Lane 1 represents the lane closest to the median in the direction of travel while Lane 4 is the farthest.

In the West direction, the evening peak hours and night hours are observed to be significant while the morning peak hours are influential for the East direction. For the West direction, the correlation between lane speed and time of day changes within lanes: the outermost lanes show a positive impact while the inner two lanes have a negative relationship. The positive correlation for outer lanes suggests that during the afternoon peak hours and night time, the average speed of vehicles in these lanes tends to increase. These findings are in line with a previous study (Shankar and Mannering, 1998). The discrepancy of impact of daytime on the speed of lanes of different directions is greatly influenced by the direction of work commute. As this study was focused on a busy Interstate in Los Angeles County area, the deviations may be justified due to location of study area.

The lane-mean speeds for neighboring lanes were incorporated as covariates in the model for the lane-mean speed of a particular lane, since previous research observed them to be significant endogenous variables. As shown in Table 2, it should be noted that the lane-mean speeds demonstrate the phenomenon of endogeneity. The lane-mean speeds of four lanes were individually modeled as dependent variables while the other lanes formed the independent part. For all four equations, the correlation between specific lane-pairs exhibits same signs for both directions which signify the presence of endogeneity and also demonstrates the robustness of the model. For example, in the case of lane-mean speed of the first lane as a dependent variable, the speed of lane 2 is negatively correlated while for lanes 3 and 4, the correlation is positive. The trend is observed to be maintained for both directions of travel as

Table 3
Parameter estimates for mean speed deviations.

Direction East					Direction West				
Variable	mean	SD	2.50%	97.50%	Variable	mean	SD	2.50%	97.50%
Speed deviation for lane 1 (dependent)					*Speed deviation for lane 1 (dependent)*				
Segment Length	−4.353	2.226	−8.714	−0.0103	Outer Shoulder width	−0.358	0.084	−0.521	−0.1863
Mean-speed lane 1	0.027	0.009	0.0086	0.0458	Monday	−2.079	0.788	−3.65	−0.5034
Log of speed deviation in lane 2	−4.337	0.619	−5.556	−3.17	Tuesday	−2.321	0.804	−3.885	−0.7645
Log of speed deviation in lane 3	9.448	0.731	8.102	10.93	AM peak	1.886	0.951	0.1118	3.847
Log of speed deviation in lane 4	−4.744	0.5	−5.789	−3.818	Log of speed deviation in lane 2	−3.827	0.313	−4.442	−3.209
					Log of speed deviation in lane 3	4.732	0.487	3.783	5.693
					Log of speed deviation in lane 4	3.817	0.48	2.848	4.761
Speed deviation for lane 2 (dependent)					*Speed deviation for lane 2 (dependent)*				
Mean-speed lane 2	0.023	0.009	0.0047	0.0415	Outer Shoulder width	−0.282	0.073	−0.423	−0.1295
Log of speed deviation in lane 1	−2.083	0.386	−2.885	−1.38	Monday	−1.686	0.626	−2.936	−0.4518
Log of speed deviation in lane 3	6.837	0.474	5.98	7.823	Tuesday	−1.807	0.648	−3.078	−0.5467
Log of speed deviation in lane 4	−3.253	0.436	−4.13	−2.44	AM peak	2.156	0.767	0.7141	3.74
					total flow	−0.003	0.001	−0.006	−2.7E−05
					Mean-speed lane 2	0.024	0.007	0.0103	0.0395
					Log of speed deviation in lane 1	−1.916	0.204	−2.295	−1.492
					Log of speed deviation in lane 3	3.43	0.404	2.645	4.233
					Log of speed deviation in lane 4	3.01	0.386	2.238	3.765
Speed deviation for lane 3 (dependent)					*Speed deviation for lane 3 (dependent)*				
Intercept	−4.209	1.952	−8.054	−0.3558	AM peak	2.009	0.661	0.723	3.308
Winter	−1.51	0.574	−2.639	−0.3927	PM peak	1.768	0.641	0.5142	3.027
Wednesday	−1.482	0.741	−2.937	−0.0186	Segment Length	−3.116	1.31	−5.727	−0.4861
Saturday	−2.409	0.742	−3.852	−0.9355	Total flow	−0.006	0.001	−0.009	−0.0042
Segment Length	3.978	1.213	1.601	6.389	mean speed lane 3	0.035	0.006	0.0224	0.0478
Mean-speed lane 3	0.028	0.009	0.0105	0.0459	Log of speed deviation in lane 1	1.661	0.354	0.9821	2.356
Log of speed deviation in lane 1	1.898	0.392	1.126	2.675	Log of speed deviation in lane 2	3.421	0.367	2.694	4.145
Log of speed deviation in lane 2	2.621	0.378	1.873	3.354	Log of speed deviation in lane 4	−1.448	0.205	−1.851	−1.052
Log of speed deviation in lane 4	2.419	0.287	1.863	2.99					
Speed deviation for lane 4 (dependent)					*Speed deviation for lane 4 (dependent)*				
Saturday	4.335	1.298	1.714	6.806	AM peak	1.849	0.717	0.4504	3.23
Segment Length	−4.598	2.143	−8.756	−0.3197	PM peak	1.875	0.691	0.5058	3.211
Log of speed deviation in lane 1	−3.314	0.408	−4.147	−2.533	Segment Length	−3.045	1.386	−5.787	−0.2345
Log of speed deviation in lane 2	−5.541	0.574	−6.699	−4.451	Total flow	−0.006	0.001	−0.009	−0.004
Log of speed deviation in lane 3	9.187	0.659	7.919	10.5	Mean-speed lane 4	0.033	0.007	0.0186	0.0489
					Log of speed deviation in lane 1	1.846	0.374	1.118	2.587
					Log of speed deviation in lane 2	4.122	0.389	3.368	4.878
					Log of speed deviation in lane 3	−2.653	0.244	−3.132	−2.177
SD (ϕ)	0.01924	0.03029	0.0012	0.1039	SD (ϕ)	0.02162	0.0389	0.0012	0.1489
SD (u)	0.002358	0.00399	4.47E−4	0.0123	SD (u)	0.7249	0.3153	0.0010	1.218
SD (ε)	4.376	0.1008	4.18	4.572	SD (ε)	3.519	0.3936	3.051	3.998

Notes: 1. Only the variables with 5% level of significance are shown in the table.
2. SD refers to Standard Deviation.
3. ϕ refers to spatial correlation between segments; u refers to random error between segments; ε refers to multivariate error between lanes.

well as for all four equations. The negative correlation between lanes 1 and 2 is observed again in the equation when lane 2 becomes a dependent variable while the positive correlation between lane-pairs 1 and 3, and 1 and 4 remains intact when lane 3 and lane 4 are modeled as dependent variables, respectively. Another pair with similar dependency is mean speeds of lanes 3 and 4, which exhibit a negative sign. It seems that only adjacent lane speeds would significantly influence in-lane speeds. Even though such common trends are illustrated in both directions, it is possible that the findings might be subject to the omitted variable issues or lack of consideration of high-order interaction within the variables, which are the common limitations of the typical statistical models.

As previously mentioned, the relationship between mean speeds of lane-pairs remains constant (negative or positive) while the variables interchange between being dependent or independent, but it is observed that the impact of the lane-mean speed of an adjacent lane is negative, i.e. an increase in the mean speed of an adjacent lane is expected to impact a subsequent decrease in the mean speed of adjoining lane. For an exploration of the rationale behind the aforementioned trends, it is worth mentioning that Lane 1 is the closest to the median while Lane 4 is the farthest. This distinction is important to understand the practical difference between the lane-mean speeds of these lanes which also influences the vehicular flow. The lane-pairs of lanes 1 & 2 and 3 & 4 are observed to have a negative interdependence of mean-speeds which means that an increase in the mean speeds of vehicles on one lane will possibly force a decrease in the speed of another. Since the area of focus of this study is a four-lane roadway (in one direction), the middle two lanes (Lanes 2 and 3) are subjected to speed differences from adjoining lanes, where Lane 1 generally experiences relatively high mean-speeds (being closest to median) and Lane 4 has a tendency to have relatively low mean-speed as it serves to be the lane for entry or diversion for on-ramps and off-ramps, respectively. In case of lane-pair of lanes 1 & 2, a negative correlation suggests that when the vehicles in Lane 1 expects to make a transition to Lane 2, that vehicle may slow down and decrease the mean speed of Lane 1 in order to match the relatively low speed of Lane 2 for a safe transition of lanes. Similar justification may be given for the trend between lane-pair of lanes 3 & 4. The outer lane (Lane 4) serves as the medium for vehicles to enter or exit the roadway from on-ramps and off-ramps, respectively, and the vehicle speed tends to be comparatively low in this lane. The lane-change maneuver from Lane 3 to Lane 4 requires slow-down of speed in Lane 3 in case of absence of a gap in Lane 4. This difference in lane-speeds of the lane-pairs 1 & 2 and 3 & 4 seems to be the cause of within-pair negative correlation. The positive interdependency of lane-pair 2 & 3 may be justified by the similarity in characteristics of traffic speed and flow for those lanes as the vehicles in Lane 1 generally tend to drive for long distances while Lane 4 vehicles are trying to drive on-off the road, but the vehicles in Lanes 2 and 3 tend to have relatively constant speed and flow. This consistency allows the direct proportionality of impact of changes within the pair.

As shown in Table 2, the length of segment is also observed to be statistically significant with the lane-mean speed. For the Lanes 1, 2, and 3, a negative correlation is observed while the mean speed of the fourth lane is noted to be positively influenced by the length of segment. The possible justification of the impact of longer segment length on decreased mean speeds of Lanes 1, 2, and 3 may be the greater opportunity of lane-change maneuvers in case of longer segments. The issue of weaving comes into play in such a scenario where the mean speeds of the vehicles decrease. The opposite correlation between segment length and mean speed of Lane 4 may be clarified by the role and position of that lane, as stated in previous paragraph. Segment length reflects the position of on and off ramps which directly influence the traffic activity on Lane 4 due to the interaction of vehicles. A smaller segment tends to decrease the mean speed compared to a longer segment as the vehicles experience free-flow conditions, unlike a smaller segment.

Among other explanatory variables, the consistent negative correlation between vehicular flow and mean speeds for all lanes in both directions suggests that the lane speed reduces with gradual increase in flow, which signifies the conditions of congestion. A similar impact was observed for the geometric factors such as inner and outer shoulder width, and road surface width. Even though the research findings are consistent with some of the previous studies (Bassani et al., 2016; Figueroa Medina and Tarko, 2005), many other studies have also found that wider travel lanes and shoulders are correlated with higher operating speeds. The counterintuitive results might result from the complex interaction among concerned crash risk factors. It is recommended to employ more sophisticated model formulations such as non- or semi-parametric Dirichlet process (Cheng et al., 2018a), or random parameters, to investigate the real-time speed more effectively due to better flexibility associated with the aforementioned formulations. The interaction between the independent significant variables may also be addressed by treating them as a criterion for separating the original dataset in different subsets. For instance, the shoulder and road surface widths may be revealed with potentially less erroneous inferences if different models are developed from subsets of data which are divided on the basis of ranges of such widths. Both results and recommendation sections were modified to illustrate the possible reasons and potential enhancement of the study.

In terms of the random effects, all three terms associated with structured and unstructured correlations are observed to be statistically significant. This demonstrates the multilevel nature of data and the justification for including the specific random effects for model development. As evident from Table 2, the multivariate error term (ε) accounted for the largest variability, followed by the random error between segments (u), and the spatial correlation between segments (ϕ). The statistical significance for all three terms suggests that their non-inclusion may lead to biased estimates and subsequently erroneous inferences, where the role of multivariate error for the dependency between neighboring lanes is more profound.

4.2. Lane-mean speed deviation

The speed deviations on the adjacent lanes are observed to be statistically significant and in some cases of East direction lanes, the deviations explain a large amount of variability in the model as very few other covariates are designated as statistically significant. The speed deviations in adjacent lanes reflect the lane-changing opportunities and their significance in the model for all lanes highlight the aspect of endogeneity. The consistent trends maintained by lane-pairs, similar to the case of lane-mean speed discussed earlier, are observed for the westbound direction. In-lane mean speeds are also incorporated as a covariate and the positive correlation with speed deviation indicates that vehicles tend to increase speed deviations on a lane with corresponding changes of mean speed. The length of segment for the outermost lane (Lane 4) is noted to be negatively correlated with speed dispersion of the vehicles in that lane. The smaller segment length allows less time for the vehicles to adjust speeds for the target lanes or coming onto/ going off the road and hence greater deviations in speeds of vehicles may occur. Similar to the model of lane-mean speed, the geometric factors, as well as the traffic flow, are observed to be negatively correlated with speed deviations. In terms of random effects, similar to the case of model estimates for lane-mean speeds, all three terms are observed to be statistically significant for both directions. However, unlike the lane-mean speed results, the coefficients for the three terms are noticeably smaller, which may suggest that their impact is relatively small while modeling the standard deviation of speed.

4.3. Model predictive accuracy

MAD and MSPE were calculated for assessment of the predictive capability of alternate models, which would demonstrate the impact of consideration of random effects from traditional speed prediction models. As shown in Table 4, the 'Bayesian vs. Observed' represents the models with the random effects which incorporate the different types of correlations at different hierarchical levels of data structure. The two types of 'Predicted vs. Observed' represent the Normal residuals, where 'Pred 1′ represents the predicted value without spatial random effect and 'Pred 2′ represents the predicted value without all random effects. The Normal residuals are

Table 4
Model prediction and consistency of alternate models.

Model	Lane #	Dir.	MAD			MSPE		
			Bayes. vs. Obs.	Pred. 1 vs. Obs.	Pred. 2 vs. Obs.	Bayes. vs. Obs.	Pred. 1 vs. Obs.	Pred. 2 vs. Obs.
Lane-mean speed	1	East	1.237	1.351	11.320	2.990	3.273	219.069
	2	East	1.183	1.493	9.868	3.057	3.960	172.669
	3	East	1.029	1.404	6.407	1.942	3.540	73.663
	4	East	0.906	1.099	17.064	1.284	1.932	488.69
Lane-mean speed	1	West	1.131	1.654	5.881	2.738	6.103	74.684
	2	West	1.354	1.695	5.219	3.660	5.853	61.463
	3	West	1.116	1.408	5.621	2.928	5.143	70.312
	4	West	1.297	1.669	4.983	3.586	6.766	50.283
Lane-mean speed deviation	1	East	0.597	0.603	4.391	0.786	0.823	35.21
	2	East	0.537	0.612	3.324	0.836	0.855	20.343
	3	East	0.694	0.725	2.527	0.823	0.848	10.662
	4	East	0.677	0.689	4.331	0.811	0.820	33.335
Lane-mean speed deviation	1	West	0.327	0.493	3.794	0.321	0.437	23.721
	2	West	0.379	0.514	3.041	0.424	0.491	15.380
	3	West	0.385	0.518	2.448	0.390	0.486	11.074
	4	West	0.365	0.568	2.734	0.340	0.618	13.536

Notes: 1. Dir. represents Direction; 2. Bayes. represents Bayesian; 3. Obs. represents observed; 4. Pred. 1 represents the predicted value without spatial random effect; 5. Pred. 2 represents the predicted value without all random effects.

separated into two types to explore the impact of spatial random effects compared to others. Since MAD and MSPE measure the deviation from the observed data, so a relatively low value is desirable which indicates better prediction capability.

The Bayesian residuals consistently exhibit superiority at accurately predicting the mean speeds and speed deviations across the different lanes for both directions. The greatest deviation between Bayesian and Normal residuals is observed for Lane 4 of eastbound traffic, with a difference of 16.158 points. This discrepancy reflects the immense impact of consideration of random effects on model performance at accurate prediction of speed estimates. More dramatic differences for particular lanes are observed for models with lane-mean speeds and subtle, but still significant, differences are noted for speed deviation models between Bayesian and Normal residuals. In case of MSPE, the differences are more pronounced with the greatest discrepancies noted for lane-mean speed of eastbound. Overall, the consistent and significant difference between the Bayesian and Pred 2 justifies the implementation of random effects. To quantify the impact of spatial correlation on model performance, Pred 1 may be compared to Bayesian. Since Pred 1 is different from Bayesian only due to the exclusion of spatial effects, their comparison depicts the loss of model performance if the spatial correlations between segments are not accommodated. As shown in Table 4, Pred 1 consistently demonstrates inferior performance for both directions and across all four lanes. The difference is subtler in case of lane-mean speed deviations since the spatial correlation term, in this case, was observed to be much smaller compared to the case of lane-mean speed (refer to Tables 2 and 3). Nonetheless, the results demonstrate that inclusion of spatial random effects improves model performance, albeit to a lesser extent compared to the other effects (multivariate error between lanes and random error between segments). Hence, it is justifiable to incorporate the structured correlation to generate the posterior estimates with higher precision. The authors also developed a visual illustration to provide an interactive comparison of model estimated speeds and the observed speed. As shown in Fig. 3, the remarkable significance of random effects is evident from the largest discrepancy in case of Pred 2, which excludes all random effects. The advantage for the predictive accuracy lend by the inclusion of random effects makes their use imperative for modeling of lane-mean speeds. Overall, the visual depiction of consistency of superior prediction performance and the quantified differences observed for MAD and MSPE provide a robust illustration of the advantage of incorporation of correlation structures in speed prediction models. Moreover, existing literature has documented that the models which fail to capture the correlations among data are prone to biased results (Lord and Mannering, 2010).

5. Conclusions and recommendations

This study focused on the investigation of influential factors on lane-mean speed and speed deviations of a multilane freeway. Speed prediction models were developed for eastbound and westbound lanes of a 44-mile freeway segment on I-10 in California. The models incorporated various exogenous variables pertaining to roadway geometry, traffic flow, time of the day, seasons, day of the week, among others. The endogenous relationship of mean speeds of adjacent lanes was also explored while accounting for different correlation structures within the data. The Full Bayesian framework was employed to build the hierarchical models which take into consideration three types of random effects: the multivariate heterogeneity of multiple lanes for all individual observation points; spatially structured and unstructured random effects. Moreover, the predictive accuracy of the models was evaluated to investigate the role of random effects on the prediction capabilities of speed prediction models.

The following conclusions are drawn from the model estimates and evaluation results:

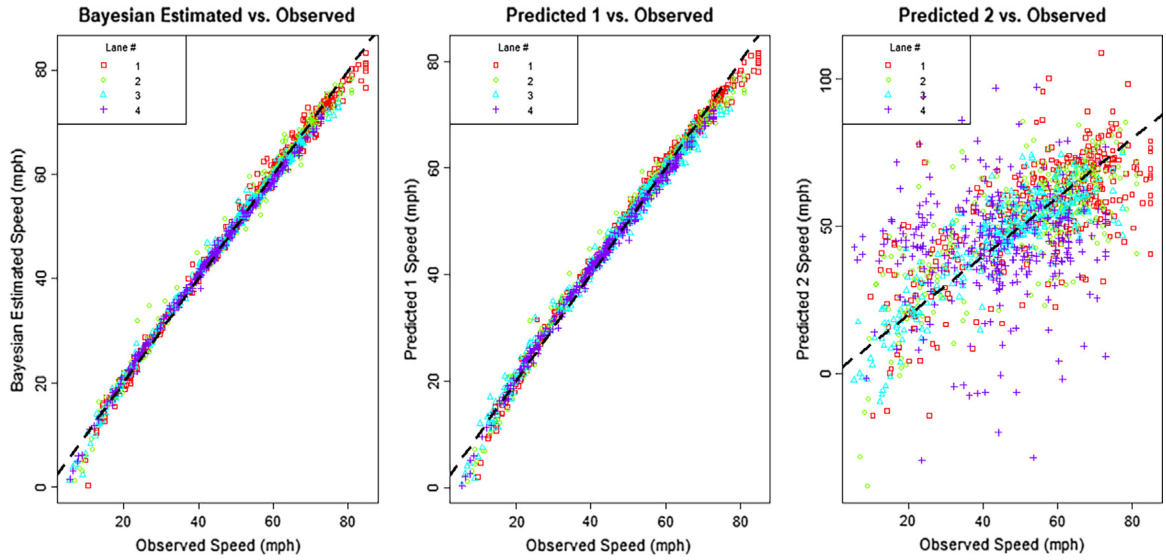

Fig. 3. The plot of various estimated speed against the observed speed (East Direction).

1. The directional variation due to exogenous factors was evident for the lane-mean speed. For the westbound traffic, the afternoon and night hours were observed to be influential while eastbound traffic was more sensitive to the morning period. Further variations were noted within the lanes which warrant the consideration of the directional factor while studying locations which are sensitive to work commute directions. This information may help the planners in devising more informed policies by accommodating the directional impact.
2. The length of segment also revealed a lane-varying correlation. Longer segments influenced a speed reduction for the three lanes closer to the median while the speed in outermost lane exhibited a reverse trend as it increased with longer segments. This phenomenon may be due to the lane-changing behavior (weaving), which is dominant for the first three lanes while the urgency of speed changes for exiting and incoming traffic dominates the outermost lane. Such phenomenon may have remained hidden for model with aggregated speeds for all lanes and hence it is recommended to employ a methodological framework which models the lanes separately while also accounting for their shared characteristics.
3. Significant presence of endogeneity was observed for mean speeds and speed deviations of adjacent lanes. The trends were influenced by the placement of lanes relative to the median which rendered the distinct traffic characteristics. The middle two lanes had a positive interdependency which may be attributed to similar traffic flow between a relatively high-speed lane (Lane 1) and low-speed lane (Lane 4). The negative correlation of lanes with Lane 1 and Lane 4 stems from the speed differences of these lanes with the adjoining lanes. The influence of endogenous variables observed in this study suggests that it is imperative to account for the endogeneity aspect to obtain more precise (less biased) posterior estimates, which directly impact the implementation of engineering solutions or policy design and subsequently their effectiveness.
4. The Bayesian estimates, which incorporated the normal, multivariate, and spatial correlation structures, exhibited a significant superiority at the prediction accuracy than the predicted ones. This discrepancy in prediction performance reflected the impact of consideration or exclusion of random effects. These findings justify the use of complex methodology which incorporates the different types of random effects. The additional computational effort of the complex model (Bayesian) was neutralized by the significantly better performance at prediction. This study recommends the inclusion of multilevel correlation structures to account for the presence of interdependencies so as to obtain relatively less biased results.

Although this study developed multivariate spatial models, more advanced correlation structures may be explored which take into account the higher order neighboring segments (second or third adjacent neighbors) and assigns continuous weights instead of a dichotomous weight utilized in this study (Gill et al., 2017). It is expected that the inferences drawn in this study may exhibit some degree of variation pertaining to the location of study area which influences many directional factors such as the commute for work, grade, sunlight glare, and so on. It is recommended to study more diverse locations and observe if the results of this study are consistent. Also, this study incorporated limited explanatory variables pertaining to roadway geometry due to their unavailability which may have rendered the models prone to omitted variable bias. It is recommended to include a diverse set of variables from factors which potentially influence the speed of vehicles such as weather changes, land-use, and demographics, among others. Moreover, some of the counterintuitive results (such as the impact of roadway surface and shoulder widths on speeds) as illustrated in the study maybe verified by employing more sophisticated model formulations (e.g., non- or semi-parametric Dirichlet process, or random parameters) due to the intrinsic benefits associated with the aforementioned formulations. Finally, the correlation structures employed in this study to model speed prediction may be extended to perform a safety analysis while considering the lane-mean speed and their deviations.

References

Abu-Eisheh, S., 2001. Modeling automobile demand and driver population in palestinian territories: Simultaneous-equation estimation method. Transport. Res. Record: J. Transport. Res. Board 1752, 108–116.

Aguero-Valverde, J., Jovanis, P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. Transport. Res. Record: J. Transport. Res. Board 2136, 82–91.

Bassani, M., Cirillo, C., Molinari, S., Tremblay, J.M., 2016. Random effect models to predict operating speed distribution on rural two-lane highways. J. Transport. Eng. 142 (6), 04016019.

Bhat, C.R., Koppelman, F.S., 1993. An endogenous switching simultaneous equation system of employment, income, and car ownership. Transport. Res. Part A: Policy Pract. 27 (6), 447–459.

Boyle, L.N., Mannering, F., 2004. Impact of traveler advisory systems on driving speed: some new evidence. Transport. Res. Part C: Emerg. Technol. 12 (1), 57–72.

Cheng, W., Gill, G., Vo, T., Zhou, J., Sakrani, T., 2018. Use of multivariate dirichlet process mixture spatial model to estimate active transportation-related crash counts. J. Transport. Res. Record (in press).

Cheng, W., Gill, G.S., Ensch, J.L., Kwong, J., Jia, X., 2018b. Multimodal crash frequency modeling: multivariate space-time models with alternate spatiotemporal interactions. Acc. Anal. Prev. 113, 159–170.

Cheng, W., Gill, G.S., Zhang, Y., Cao, Z., 2018c. Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. Acc. Anal. Prev. 112, 84–93.

Cheng, W., Wang, J.H., Bryden, G., Ye, X., Jia, X., 2013. An examination of the endogeneity of speed limits and accident counts in crash models. J. Transport. Saf. Sec. 5 (4), 314–326.

Cheng, W., Gill, G.S., Dasu, R., Xie, M., Jia, X., Zhou, J., 2017a. Comparison of Multivariate Poisson lognormal spatial and temporal crash models to identify hot spots of intersections based on crash types. Acc. Anal. Prev. 99, 330–341.

Cheng, W., Gill, G.S., Sakrani, T., Dasu, M., Zhou, J., 2017b. Predicting motorcycle crash injury severity using weather data and alternative Bayesian multivariate crash frequency models. Acc. Anal. Prev. 108, 172–180.

Congdon, P., 2006. Bayesian Statistical Modeling, 2nd ed. Wiley, New York.

Dane, G., Arentze, T.A., Timmermans, H.J., Ettema, D., 2014. Simultaneous modeling of individuals' duration and expenditure decisions in out-of-home leisure activities. Transport. Res. Part A: Policy Pract. 70, 93–103.

Derrig, R.A., Segui-Gomez, M., Abtahi, A., 2000. The effect of seat belt usage rates on the number of motor vehicle-related fatalities. In: Proceedings of the 2000 Risk Theory Society Seminar, Minneapolis, Minnesota (pp. 14–16).

Einbeck, J., Dwyer, J., 2011. Using principal curves to analyse traffic patterns on freeways. Transportmetrica 7 (3), 229–246.

Eisenberg, D., 2003. Evaluating the effectiveness of policies related to drunk driving. J. Policy Anal. Manage. 22 (2), 249–274.

Figueroa Medina, A., Tarko, A., 2005. Speed factors on two-lane rural highways in free-flow conditions. Transport. Res. Record: J. Transport. Res. Board 1912, 39–46.

Fitzpatrick, K., Collins, J.M., Speed-profile model for two-lane rural highways. In: Transportation Research Record: Journal of the Transportation Research Board, No. 1737, TRB, National Research Council, Washington, D.C., 2000, pp. 42–49.

Fitzpatrick, K., Miaou, S.P., Brewer, M., Carlson, P., Wooldridge, M.D., 2005. Exploration of the relationships between operating speed and roadway features on tangent sections. J. Transport. Eng. 131 (4), 261–269.

Gill, G., Cheng, W., Xie, M., Vo, T., Jia, X., Zhou, J., 2017. Evaluating the influence of neighboring structures on spatial crash frequency modeling and site ranking performance. Transport. Res. Record: J. Transport. Res. Board, No. 2659 (10.3141/2659-13).

Golob, T.F., Recker, W.W., 2003. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. J. Transport. Eng. 129 (4), 342–353.

Greenshields, B., 1935. A study of traffic capacity. Highway Res. Board Proceed. 14, 448–477.

Himes, S.C., Donnell, E.T., 2010. Speed prediction models for multilane highways: simultaneous equations approach. J. Transport. Eng. 136 (10), 855–862.

Jonathan, A.V., Wu, K.F.K., Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. Acc. Anal. Prev. 87, 8–16.

Kim, D.G., Washington, S., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Acc. Anal. Prev. 38 (6), 1094–1100.

Lawson, A.B., Browne, W.J., Rodeiro, C.L.V., 2003. Disease mapping with WinBUGS and MLwiN (Vol. 11). John Wiley & Sons.

Li, Y., Lu, H., Bian, C. and Sui, Y.G., 2009, April. Traffic speed-flow model for the mix traffic flow on Beijing urban expressway. In: Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on IEEE (vol. 3, pp. 641–644).

Lighthill, M., Whitham, G., 1955. On kinematic waves II. A theory of traffic flow on long crowded roads. Proc. Royal Society of London, Part A 229 (1178), 317–345.

Lindeman, H., Ranft, B., 1978. Speed on Curves. ETHZ Institute for Traffic Planning and Transport, Zurich, Switzerland.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transport. Res. Part A: Policy Pract. 44 (5), 291–305.

Medina, A.M.F., Tarko, A.P., 2007. Modeling endogenous relationship between driver behavior and highway safety. In: 86th TRB Annual Conference. Washington DC.

Pei, X., Wong, S.C., Li, Y.C., Sze, N.N., 2012. Full Bayesian method for the development of speed models: applications of GPS probe data. J. Transport. Eng. 138 (10), 1188–1195.

Shankar, V., Mannering, F., 1998. Modeling the endogeneity of lane-mean speeds and lane-speed deviations: a structural equations approach. Transport. Res. Part A: Policy Pract. 32 (5), 311–322.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Acc. Anal. Prev. 27 (3), 371–389.

Silvano, A.P., Bang, K.L., 2015. Impact of speed limits and road characteristics on free-flow speed in urban areas. J. Transport. Eng. 142 (2), 04015039.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS Version 1.4 User Manual. MRC Biostatistics Unit, Cambridge http://www.mrc-cam.ac.uk/bugs.

Transportation Injury Mapping System (TIMS), 2017. Safe Transportation Research and Education Center, University of California, Berkeley.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and econometric methods for transportation data analysis. CRC Press.