Modeling Traffic Incident Duration Using Quantile Regression

Asad J. Khattak, Jun Liu, Behram Wali, Xiaobing Li, and ManWo Ng

Traffic incidents occur frequently on urban roadways and cause incident-induced congestion. Predicting incident duration is a key step in managing these events. Ordinary least squares (OLS) regression models can be estimated to relate the mean of incident duration data with its correlates. Because of the presence of larger incidents, duration distributions are often right-skewed; that is, the OLS model underpredicts the durations of larger incidents. Therefore, this study applies a modeling technique known as quantile regression to predict more accurately the skewed distribution of incident durations. Quantile regression estimates the relationships between correlates and a chosen percentile—for example, the 75th or 95th percentile—while the OLS regression is based on the mean of incident duration. With the use of incident data related to more than 85,000 (2013 to 2015) incidents for highways in the Hampton Roads area of Virginia, quantile regression results indicate that the magnitudes of parameters and predictions can be quite different compared with OLS regression. In addition to predicting durations of larger incidents more accurately, quantile regressions can estimate the probability of an incident lasting for a specific duration; for example, incidents involving congestion and delay have an approximately 25% chance of lasting more than 100.8 min, while incidents excluding congestion and delay are estimated to have a 25% chance of lasting more than 43.3 min. Such information is helpful in accurately predicting durations and developing potential applications for using quantile regressions for better traffic incident management.

Traffic incidents occur frequently on roadways, resulting in congestion, commuter anxiety, and harmful vehicular emissions (*1–3*). One traffic incident management strategy is to disseminate accurate incident duration information to travelers (e.g., through variable message signs), who can then make more informed travel decisions (*4, 5*). Another approach would be to actively redirect traffic in a road network to avoid incident-induced congestion. In both cases, accurate predictions of incident durations are required.

Incident duration is defined as the time between the occurrence of an incident and the clearance of the roadway (*6–8*). Traditionally, researchers have applied ordinary least squares (OLS) models (i.e., the linear regression models) to predict incident duration (*9–13*). By definition, OLS models examine the (conditional) mean of incident durations. Therefore, incidents that are much shorter or longer than average cannot be accurately captured with OLS models. To model those incidents, this study proposes to use quantile regression. Quantile regression is a statistical technique that can relate quantiles of the incident duration distribution to explanatory variables (*14*). While traffic operations managers might be more interested in the higher quantiles, that is, longer duration incidents, quantile regression, as shall be shown, is equally suitable for modeling shorter-than-average incidents. This study discusses potential applications of quantile regression in traffic incident management. In general, with quantile regression, transportation professionals (e.g., traffic operators in transportation management centers) can benefit by accurately predicting the incident duration and potentially reducing large-scale incident durations through appropriate solutions.

A. J. Khattak, 322 John D. Tickle Building; J. Liu, 325 John D. Tickle Building; and B. Wali and X. Li, 311D John D. Tickle Building, Department of Civil and Environmental Engineering, College of Engineering, University of Tennessee, 851 Neyland Drive, Knoxville TN 37996-2313. M. Ng, Department of Information Technology and Decision Sciences, Strome College of Business, Old Dominion University, Norfolk, VA 23529. Corresponding author: A. J. Khattak, akhattak@utk.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2554, Transportation Research Board, Washington, D.C., 2016, pp. 139–148.
DOI: 10.3141/2554-15

LITERATURE REVIEW

Various techniques have been reported in the literature for modeling traffic incident duration. The techniques can be grouped into several categories: statistical models, tree modeling, intelligence techniques, and mixed modeling. Brief discussions of each follow.

Statistical models. Linear regression models were estimated to provide real-time incident information to travelers. OLS regression, OLS with logarithmic transformation, and a series of truncated regression models were targeted at skewed data distributions and sequential availability of incident information in real time (*9–13*). Partial least squares regressions were also studied (*15*). Traditional negative binomial and modified negative binomial were also used (*16*). Various studies have developed parametric accelerated failure time survival models for incident durations arising from crashes and hazards and for incidents involving stationary vehicles (*17–21*).

Tree modeling. Ji used decision trees to predict freeway incident durations on the basis of the multimodal fusion algorithm (*22*). Chang and Chang reported good performance of the classification tree method for short-duration incident predictions (*23*).

Intelligence techniques. Neural networks were used in various studies (*24, 25*). However, they have not been used to update duration prediction information dynamically (*26*).

Mixed modeling. Lin et al. combined a discrete choice model and a rule-based model for predicting incident duration (*27*). He et al. used the hybrid tree-based quantile regression (*28*). Xiaoqiang et al. used the classification and regression tree method (*29*). The classification tree, rule-based tree model, and discrete choice model were studied sequentially by Kim et al. to improve prediction accuracy (*30*). Li et al. applied topic modeling, the multinomial logistic model, and the parametric hazard-based model (*31*).

Model comparison has also been of interest. Li and Shang compared prediction models, including the classification and regression tree, chi-squared automatic interaction detector, and exhaustive chi-squared automatic interaction detector, on the basis of performance criteria such as mean absolute percentage error and root mean square error (RMSE) (*17*). They found that RMSE and mean absolute percentage error were relatively low for 15- to 45-min-long durations, while for long durations, prediction accuracy was largely decreased.

Researchers have found that the prediction accuracy for long-duration incidents is generally lower than for short-duration incidents (*23*). The benefit of predicting long durations is not as visible as for shorter durations, as the distribution of incident durations is rather dispersed (*28*). Therefore, quantile regression is chosen as the key method able to account for the dispersed distribution of responses. Quantile regression has been explored by researchers in various fields. Machado and Silva successfully applied quantile regression to health care through a jittering procedure (*32*). Qin et al. (*33*) and Qin (*34*) explored the application of quantile regression on traffic crash data.

To summarize, the gaps in the existing literature on incident prediction are related to (*a*) prediction accuracy of durations and (*b*) the practice of using "black box" models to assist in incident management. In regard to prediction accuracy, while previous studies have demonstrated the application of various modeling techniques to predict incident durations, their prediction accuracy has been a recurring concern, owing to the skewed distribution of incident durations. Theoretically, quantile regression should provide more accurate incident duration predictions since it can account for dispersed and skewed distributions of incident durations. In regard to practice, some researchers have developed models—such as the classification tree model (*23*), classification and regression tree, and chi-squared automatic interaction detector (*35*)—for predicting short versus long durations. Their models may be good in predicting the duration of particular types of incidents. However, these models can be black boxes that do not provide clear intuition to users about correlations between various factors and incident durations. The estimation of correlations is important for incident duration prediction as it can help develop solutions for incident management. Quantile regression is able to estimate variations in correlates of incidents, which means that more focused solutions that address long- and medium-duration incidents can be developed. Such information can be very helpful for incident management.

METHODOLOGY

Data Sources

This study used various data sources, including incident data provided by the Hampton Roads Smart Traffic Center in Virginia Beach, Virginia. These data were collected by the Safety Service Patrol (SSP) of the Hampton Roads area. The records cover the incidents that occurred in the 2013 to 2015 period on freeways; the records include the start and end times, incident duration, incident type, agencies that responded to incidents, and so on. Other data sources used include the road inventory data provided by the Hampton Roads Planning District Commission.

OLS and Quantile Regression

For completeness, in this section the OLS and quantile regression techniques to be used in the next section of this paper are briefly reviewed. This study compares the traditional OLS model with the quantile regression model, which is considered to be more suitable to model the dispersed distribution of incident durations.

OLS Model

The OLS model is given by

$$y_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} + \varepsilon_i \tag{1}$$

where

y_i = dependent variable, that is, duration of ith incident (min), $i = 1, 2, \ldots, m$;
β_0 = intercept;
β_j = coefficient of independent variable j, $j = 1, 2, \ldots, n$;
x_{ij} = value of independent variables j in ith incident; and
ε_i = estimation error or residual for ith incident.

The error ε_i is assumed to be normally distributed with a mean of zero and a finite variance. Coefficients of the independent variables are estimated by minimizing the mean squared error criterion:

$$\sum_{i=1}^{m} \left(y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2 \tag{2}$$

The resulting least squares estimates of β_0 and β_j are then denoted by $\hat{\beta}_0$ and $\hat{\beta}_j$, respectively. OLS models provide intuitive estimations of the relationship between incident duration and associated factors: one unit increase in an independent variable leads to an increase of $\hat{\beta}_j$ in the mean incident duration, with all other variables held constant.

Quantile Regression

OLS models may be a good choice for predictions in which the mean values are of interest. For a more complete picture of the distribution of incident durations, quantile regression becomes more appropriate. Particularly, rather than modeling only the average incident duration (as in OLS regression), quantile regression can model the relationship of any quantile with a set of explanatory variables (*8*).

Contrary to OLS models that minimize the mean squared error, quantile regression minimizes a sum that gives asymmetric penalties $(1-q)|\varepsilon_i|$ for overprediction and $q|\varepsilon_i|$ for underprediction, where q is the quantile point of the outcomes. For example, if one wants to model the median incident duration, one would choose $q = 0.5$. The prediction errors in quantile regression are given by

$$\varepsilon_i^q = y_i - \hat{\beta}_0^q - \sum_{j=1}^{n} \hat{\beta}_j^q x_{ij} \tag{3}$$

where $\hat{\beta}_0^q$ is the estimated intercept at quantile point q, $0 < q < 1$, and $\hat{\beta}_j^q$ is the estimated coefficient of independent variable j at quantile point q. More specifically, the coefficients $\hat{\beta}_0^q$ and $\hat{\beta}_j^q$ are estimated by minimizing the following objective function (*14*):

$$\sum_{i:y_i \geq \beta_0^q + \sum_{j=1}^{n} \beta_j^q x_{ij}}^{n} q \left| y_i - \beta_0^q - \sum_{j=1}^{n} \beta_j^q x_{ij} \right| + \sum_{i:y_i < \beta_0^q + \sum_{j=1}^{n} \beta_j^q x_{ij}}^{n} (1-q) \left| y_i - \beta_0^q - \sum_{j=1}^{n} \beta_j^q x_{ij} \right| \Bigg) \tag{4}$$

where y_i is a dependent variable, that is, the duration of ith incident (min), $i = 1, 2, \ldots, n$; and x_{ij} is the value of independent variables j in the ith incident.

Incident Duration Prediction

From the perspective of modeling outcomes, OLS models provide intuitive results, giving a single number that is the predicted mean, while quantile regression can provide estimates for any quantile q, where q can be any number between 0 and 1. Thus, quantile regression can be seen as providing estimates of the entire (conditional) distribution of incident durations given certain conditions and does not give incident duration prediction directly, that is, it does not provide a single number of how many minutes an incident may last. This study applies a location-based prediction method to predict the incident durations with quantile regression.

Location-Based Prediction

Location-based prediction can be applied if regional historical incident data are available (*36*). It assumes that traffic safety outcomes do not change dramatically in a short period; the durations of incidents in one segment or intersection remain in the same quantile of all incidents in a region. For example, if the historical data show that durations of incidents in one segment are likely to be at the 75th percentile, the predicted durations for this segment are approximately the estimates of quantile regression at the 75th percentile. In this study, the quantile regressions for duration prediction are made at the 5th, 15th, 25th, . . . , 95th percentiles, as shown in Figure 1. Thus, the predicted duration can be obtained at the 5th percentile regression if the observed value is less than the 10th percentile, or at the 15th percentile regression if the observed value is within the 10th to the 20th percentile, and so forth. With the location-based prediction method, the incident duration can be predicted with

$$\hat{y} = \left\{ \hat{y}_m \left| \begin{array}{l} m = 5, \text{ if } q_0 < \overline{y} \le q_{10} \\ m = 15, \text{ if } q_{10} < \overline{y} \le q_{20} \\ \quad\vdots \\ m = 95, \text{ if } q_{90} < \overline{y} \le q_{100} \end{array} \right. \right\} \tag{5}$$

where

$\hat{y}$ = predicted incident duration using location-based prediction method,
$\hat{y}_m$ = predicted incident duration at center of interval m (i.e., percentile location),
$\overline{y}$ = average of historical incident duration at particular location (e.g., bottleneck), and
q_p = pth percentile value of durations of incidents in region.

Model Comparison

This study compares the two modeling techniques—that is, OLS and quantile regression models—by calculating the RMSE for the resulting incident duration predictions. A smaller RMSE indicates a better prediction. The RMSE can be calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n}} \tag{6}$$

where

n = number of observations,
y_i = observed duration for ith incident in data set, and
$\hat{y}_i$ = predicted duration for ith incident in data set.

MODELING RESULTS

Descriptive Statistics

Table 1 presents descriptive statistics of variables selected for analysis and modeling. Figure 2 shows the distributions of incident durations of valid observations, $N = 85{,}624$. Observations with missing information were removed from the data set. The descriptive statistics of selected variables seem to be within reasonable ranges. The distribution of incident duration is widely dispersed. The mean duration was 50.96 min, with a standard deviation of 107.13 min. The maximum incident duration was 1,419 min. Thus, it is clear that the dispersed distribution of incident duration implies that the mean duration does not appropriately represent a full picture of all incidents.

The variable "detection source" refers to how an incident is detected. Seven dummy variables were created: the SSP, closed-circuit television (CCTV), citizen call, contractor call, field device or police, Virginia Department of Transportation field staff, and the Virginia State Police. The majority of incidents, 60.4%, were reported through SSP. In regard to incident type, disabled incidents represented 60.7% of the sampled incidents. Three roadway types were considered in the analysis: Interstates, primary roads, and urban roads; 83% of the incidents occurred on Interstates.

In regard to temporal characteristics, the developed models incorporate the associations between a.m. peak (0600 to 1000 hours), p.m. peak (1600 to 1900 hours), midday (1000 to 1600 hours), and night (1900 to 0600 hours) and incident durations, respectively. Definitions for the aforementioned temporal variables are adopted while taking guidance from several past studies, for example, see the *Urban Mobility Scorecard* (*37*). Of the incidents that occurred, 36% and 30% occurred during the night and at midday, respectively.

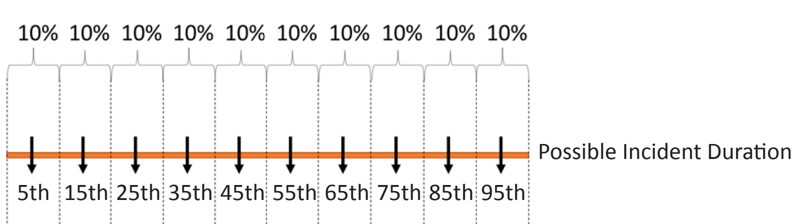

FIGURE 1 Intervals and locations of quantile regression.

TABLE 1 Descriptive Statistics of Incident Data from Hampton Roads, Virginia

Variable	Valid *N*	Mean	Frequency	SD	Min.	Max.	VIF
Incident duration (min)	85,624	50.960	na	107.134	1	1,419	na
Detection source							
SSP	85,624	0.604	51,717	0.488	0	1	na
CCTV	85,624	0.203	17,382	0.402	0	1	1.64
Citizen call	85,624	0.003	257	0.059	0	1	1.01
Contractor call	85,624	0.103	8,819	0.304	0	1	2.11
Field device or police	85,624	0.001	86	0.040	0	1	1.01
Virginia DOT field staff	85,624	0.006	514	0.079	0	1	1.13
VSP	85,624	0.076	6,507	0.266	0	1	1.16
Incident type							
Accident	85,624	0.097	8,306	0.296	0	1	na
Congestion/delay	85,624	0.037	3,168	0.189	0	1	2.57
Disabled vehicle	85,624	0.607	51,974	0.488	0	1	4.66
Other	85,624	0.255	21,834	0.436	0	1	7.76
Vehicle fire	85,624	0.002	172	0.045	0	1	1.03
Roadway type							
Interstate	85,624	0.830	71,068	0.374	0	1	2.69
Primary	85,624	0.040	3,425	0.197	0	1	1.61
Urban	85,624	0.007	599	0.088	0	1	1.12
Time of day							
a.m. peak	85,624	0.176	15,070	0.380	0	1	na
Midday	85,624	0.300	25,687	0.458	0	1	1.92
p.m. peak	85,624	0.161	13,785	0.367	0	1	1.64
Night	85,624	0.362	30,998	0.480	0	1	2.03
Day of week							
Weekday	85,624	0.767	65,674	0.422	0	1	na
Weekend	85,624	0.232	19,865	0.422	0	1	1.02
Injury count	85,624	0.017	na	0.175	0	6	1.37
Number of involved vehicles	85,624	0.814	na	0.627	0	11	3.37
Rescue responded (1–yes, 0–no)	85,624	0.029	2,483	0.168	0	1	1.63
Work zone involved (1–yes, 0–no)	85,624	0.002	171	0.046	0	1	1.02

NOTE: VIF = variance inflation factor; na = not applicable; DOT = department of transportation; VSP = Virginia State Police.

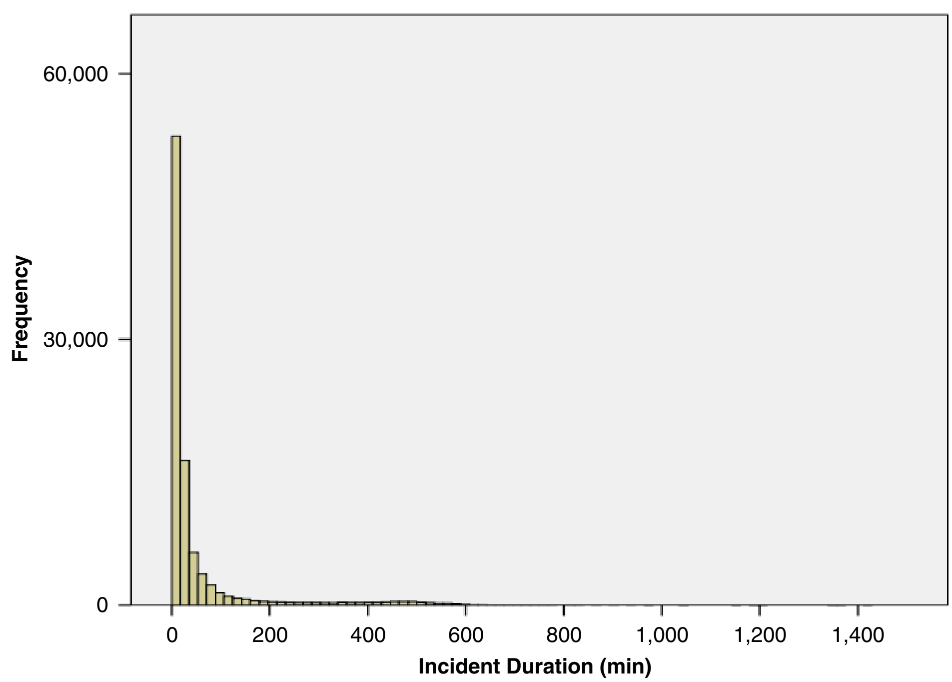

FIGURE 2 Duration distribution of traffic incidents in sample: Hampton Roads (*N* = 85,624).

Moreover, the descriptive statistics reveal that 76.7% of the incidents occurred on weekdays. On average, 0.814 vehicles were involved in sampled incidents, whereas the mean injury count in the data set was found to be 0.017. Last, rescue services responded to only 2.9% of the incidents.

Incident Duration Models

Table 2 presents the outputs of OLS and quantile regression models estimated at the 25th, 50th, 75th, and 95th percentiles. Most of the variables are statistically significant (at the 95% level). The signs of the coefficients are as expected. In general, the coefficients of the OLS model are within the range of the coefficients estimated by the quantile regression models.

The OLS model provides only one set of coefficients, indicating the amount of increase or decrease in the average incident duration with one unit increase in an independent variable, with other variables being held constant. Quantile regression provides one set of coefficients for each quantile considered. For a given quantile, the interpretation of the coefficients is the same as in an OLS model; it is the change in the incident duration in a given quantile category, with one unit increase in the independent variable. Figure 3 presents the coefficients of key factors at continuous quantiles, relative to the coefficients estimated with OLS regression. The coefficients of quantile regression vary across different quantiles, while OLS coefficients are constant.

From the OLS model, it can be seen that compared with SSP detected incidents, those detected by CCTV, contractor call, and Virginia State Police are expected to be 27.92, 24.93, and 6.41 min

TABLE 2 OLS and Quantile Regression Models

	OLS (mean)		25th Percentile		Median (50th percentile)		75th Percentile		95th Percentile	
Variable	β	*t*	β	*t*	β	*t*	β	*t*	β	*t*
Detection source										
SSP	Base		Base		Base		Base		Base	
CCTV	27.92	31.17	5.00	48.93	11.00	38.73	17.00	13.00	23.00	7.44
Citizen call	–1.86	–0.39	6.00	11.06	10.00	6.64	10.00	1.44	3.00	0.18
Contractor call	24.93	18.61	5.00	32.66	9.01	21.55	9.00	4.59	36.00	7.78
Field device or police	86.25	12.43	11.00	13.88	22.00	9.99	108.50	10.70	99.00	4.13
Virginia DOT field staff	27.62	7.33	5.00	11.62	9.00	7.53	11.00	2.00	27.00	2.08
VSP	6.41	5.65	8.00	61.69	11.00	30.52	11.00	6.63	8.00	2.04
Incident type										
Accident	Base		Base		Base		Base		Base	
Congestion/delay	40.05	15.85	12.00	41.57	–27.00	33.66	57.50	15.57	159.00	18.22
Disabled vehicle	–15.22	–11.97	–3.00	–20.64	–13.00	–32.19	–27.00	–14.52	–49.00	–11.15
Other	45.92	24.08	–2.00	–9.18	–9.00	–14.86	35.50	12.73	343.00	52.06
Vehicle fire	7.88	1.29	9.00	12.93	7.00	3.62	8.00	0.90	14.00	0.67
Roadway type										
Interstate	9.93	1.25	2.00	13.99	6.00	15.11	8.00	4.37	21.00	4.86
Primary	32.37	1.81	2.01	9.63	8.00	13.86	15.00	5.64	37.00	5.89
Urban	33.26	3.45	2.03	5.06	10.00	9.11	19.00	3.76	26.00	2.18
Time of day										
a.m. peak	Base		Base		Base		Base		Base	
Midday	–12.86	–15.24	0.00	0.00	–2.00	–7.47	–5.00	–4.05	–51.00	–17.50
p.m. peak	–6.91	–7.16	0.00	0.00	–2.01	–6.53	–4.00	–2.83	–47.00	–14.10
Night	12.14	14.43	1.00	10.40	1.00	3.74	3.00	2.44	–19.00	–6.54
Day of week										
Weekday	Base		Base		Base		Base		Base	
Weekend	4.13	6.21	0.00	0.00	0	0.00	1.00	1.03	12.00	5.21
Injury count	9.86	5.40	10.50	50.31	8.00	13.79	8.00	3.00	7.00	1.11
Number of involved vehicles	5.03	5.97	3.00	31.10	4.00	14.92	5.50	4.46	5.00	1.71
Rescue responded (1–yes, 0–no)	18.48	8.94	21.00	88.88	25.00	38.03	20.00	6.62	46.00	6.44
Work zone involved (1–yes, 0–no)	–10.95	–1.85	5.00	7.41	0.00	0.00	–7.50	–0.87	–1.00	–0.05
Constant	16.23	6.98	1.00	3.76	12.00	16.26	33.50	9.86	116	14.44
Number of observations	85,624		85,624		85,624		85,624		85,624	
Total sum of squared errors	685,567,430		na		na		na		na	
Model sum of squared errors	105,690,879		na		na		na		na	
R^2	.15		.04[a]		.05[a]		.10[a]		.41[a]	
Raw sum of deviations	na		837,475.3		1,549,636		2,001,910		1,483,797	
Minimum sum of deviations	na		807,477.5		1,465,654		1,802,375		872,479.4	

[a]Represents pseudo-R^2 for quantile regression; the median (or any other quantile) regression estimates are based on maximum likelihood for double exponential distribution. The goodness-of-fit measure is calculated as pseudo-R^2 = 1 – minimum sum of deviations/raw sum of deviations.

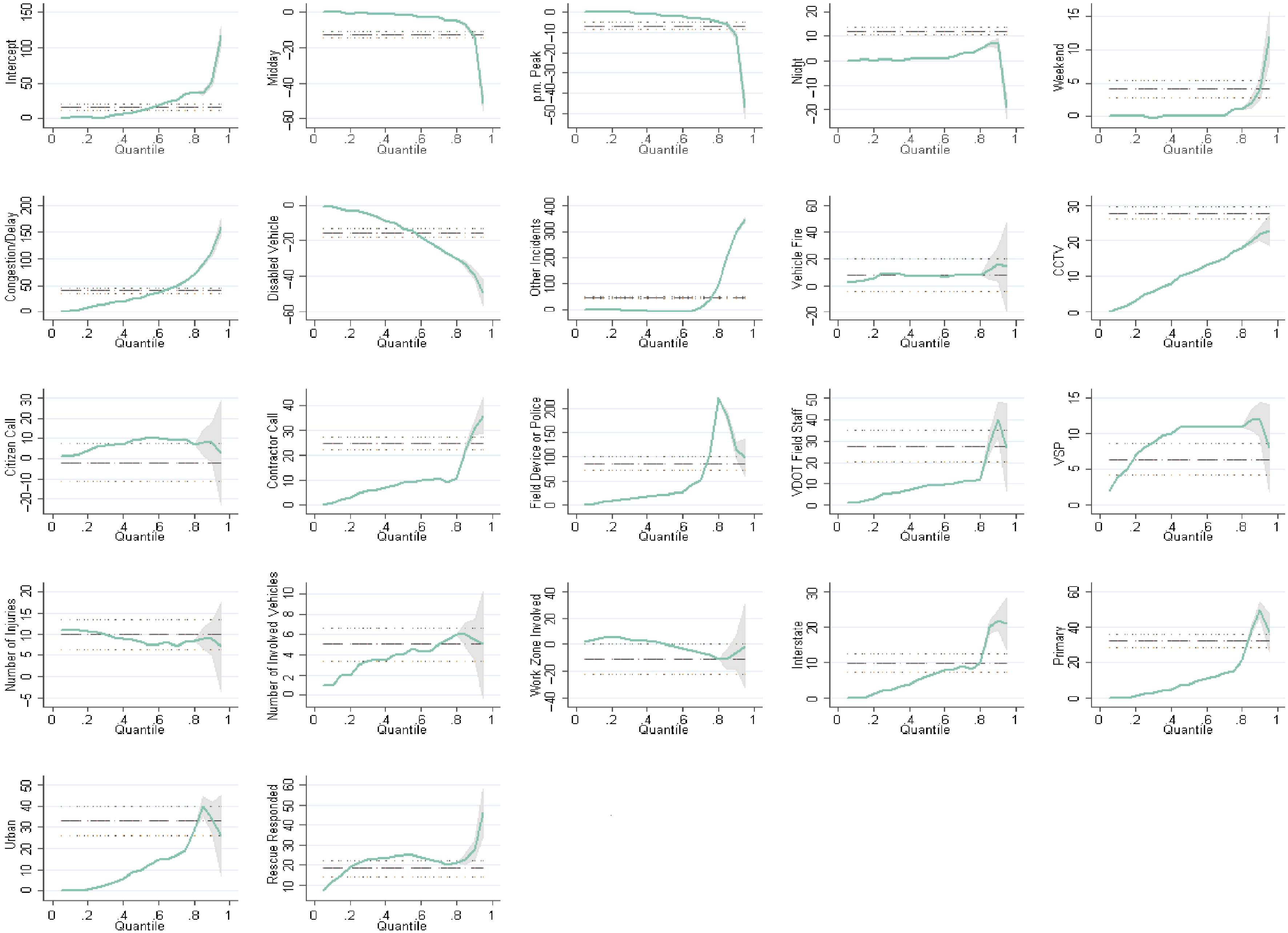

FIGURE 3 Coefficients of OLS and quantile regression models based on Hampton Roads incident data. Black broken line shows estimates from OLS regression; 95% confidence intervals are shown by black dotted lines. Blue line shows estimates from quantile regression; 95% confidence intervals are shown by shaded region (VDOT = Virginia DOT).

longer, respectively. From the quantile regression, the coefficients vary across different percentiles. The differences between SSP and the other detection sources are greater for the upper percentiles (i.e., 75th and 95th percentiles), especially for incidents reported by CCTV, contractor call, and field device or police. For example, for long incidents (in the 95th percentile relative to their duration), when an incident is first reported by CCTV, then the incident duration will be longer by as much as 23 min compared with when the incident is reported by SSP.

On average, the incident duration resulting from congestion or delay is 40.05 min longer than for accidents, while the quantile regression indicates that the associations between incident type being "congestion/delay" and incident durations are significantly higher at the 75th and 95th percentiles. This observation intuitively indicates that once an incident occurs, associations between "congestion/delay" and incident duration become stronger as incident duration increases.

Incidents on freeways are positively correlated with incident durations. On average, an incident on an Interstate is expected to last 9.93 min. However, quantile regression reveals significantly varying positive correlation between Interstate incidents and incident durations, with larger positive correlation at higher quantiles. Likewise, the positive correlation between incidents occurring on urban routes is higher at higher quantiles as compared with lower quantiles. The results from quantile regression thus provide more exhaustive insights about complex interactions, which can help in the development of more-informed incident management strategies.

As compared with a.m. peak incidents, incidents occurring during midday are on average 12.86 min shorter. Nighttime incidents are on average 12.14 min longer than a.m. peak incidents. Contrarily, the results from quantile regression suggest that the association between higher quantile incident duration and midday incident is strongly negative as compared with lower quantile incident duration. There could be several reasons for this finding. For instance, once an incident turns out to be longer, there could be other potential observed or unobserved factors or both that may contribute to an incident's longer duration. In the presence of such unobserved factors that may be associated with longer incident durations, the influence of midday incident on incident duration may be relatively smaller. Quantile regression shows that for incidents that normally last longer than the median, an incident on the weekend may last even longer, according to the larger magnitudes of the coefficients at higher quantiles, as shown in Figure 3. The number of vehicles involved in incident and injury counts has a positive relationship with incident duration. If rescue responds to an incident, the incident is expected to last on average 18.48 min longer compared with an incident that does not receive a response from rescue. The increase would be 46 min at the 95th percentile, indicating a more pronounced positive association. This is, however, merely a correlation since rescue services may, in turn, be needed for larger incidents, and the rescue services likely decrease the duration of the incidents compared with the duration if rescue had not responded.

Using the coefficients from quantile regression, this study proposes another way to interpret the quantile regression results. Table 3 provides the estimation of incident duration by holding all variables at their mean values: the mean incident duration is 44.10 min, 6.68 min at the 25th percentile, 13.86 min at the median, 45.45 min at the 75th percentile, and 186.54 min at the 95th percentile. All these numbers are close to the distributions of the 85,624 incidents sampled in the study. Table 3 allows one to predict the incident duration given a certain value of the independent variable while controlling for other variables at their means. Changes in the probability that an incident with a given duration will occur owing to the change in values of independent variables are quantified.

For example, all other factors are at their means, and only the incident type is allowed to vary. The incident duration at the 75th percentile is estimated to be 45.45 − 2.13 = 43.32 min when the incident is not related to congestion or delay, meaning there is a 25% chance that an incident lasts at least 43.32 min if it is not the result of congestion or delay. When the incident is related to congestion or delay, incident duration at the 75th percentile is calculated to be 45.45 − 2.13 + 57.50 = 100.82 min, indicating a 25% chance that an incident will last 100.82 min or longer. Notably, the 75th percentile incident duration for congestion or delay is 100.82 min, which is close to the 95th percentile estimation for other (unclassified) incidents. The associations of other factors with incidents can be interpreted in the same way. The exact increase or decrease in the chance or probability can be obtained by comparing estimations at other percentiles, such as the 25th or 50th.

Performance Comparison

As mentioned earlier, incident durations can be predicted by the OLS model and by quantile regression models. This study used the location-based method to obtain the predicted values based on the estimation of quantile regression. The quantile regressions for incident duration prediction are made at the 5th, 15th, 25th, . . . , 95th percentiles. To predict incident durations with quantile regression, individual quantile regressions estimated at the 5th, 15th, 25th, . . . , 95th percentiles are used. Next, the incident duration associated with increments of the 10th percentiles are calculated. If a specific observed value for the incident duration value falls within a percentile—for example, if it is less than the 10th percentile (suppose it is equal to 2 min)—then the 5th percentile regression is used to predict incident durations in this bin. Likewise, if the observed incident duration is between the 40th and 50th percentile (i.e., greater than 9 and less than 14 min), then the 45th percentile regression is used to predict the incident duration in this bin, and so on. Thus, the combined predictions (using the 5th, 15th, 25th, . . . , 95th percentile equations) from quantile regression are compared with the single equation (mean) OLS predictions.

The RMSEs are calculated with Equation 6. Their values show the extent of the difference between the predicted and observed incident durations. The RMSE for OLS is 82.29 min, while for the quantile regression with location-based prediction, it is 57.49 min. The quantile regression is observed to be significantly better in predicting incident durations through the location-based method. The location-based method seems the best in regard to accurately predicting the incident duration; however, historical data are required for the use of this method.

POTENTIAL APPLICATIONS

There are potential applications of the quantile regression method in traffic incident management. First, the models can more accurately predict incident durations in real time and, second, analysis of correlates can be used to design strategies for reducing incident durations. Transportation researchers and professionals in different areas may use the method proposed in this study to develop their local quantile regression models for regional incident management.

TABLE 3 Estimation of Incident Duration at Means of Independent Variables

		OLS (mean)		25th Percentile		Median (50th percentile)		75th Percentile		95th Percentile	
Variable	X	β	β * X	β	β * X	β	β * X	β	β * X	β	β * X
Detection source											
SSP	0.604	Base		Base		Base		Base		Base	
CCTV	0.203	27.92	5.67	5.00	1.02	11.00	2.23	17.00	3.45	23.00	4.67
Citizen call	0.003	–1.86	–0.01	6.00	0.02	10.00	0.03	10.00	0.03	3.00	0.01
Contractor call	0.103	24.93	2.57	5.00	0.52	9.01	0.93	9.00	0.93	36.00	3.71
Field device or police	0.001	86.25	0.09	11.00	0.01	22.00	0.02	108.50	0.11	99.00	0.10
Virginia DOT field staff	0.006	27.62	0.17	5.00	0.03	9.00	0.05	11.00	0.07	27.00	0.16
VSP	0.076	6.41	0.49	8.00	0.61	11.00	0.84	11.00	0.84	8.00	0.61
Incident type											
Accident	0.097	Base		Base		Base		Base		Base	
Congestion/delay	0.037	40.05	1.48	12.00	0.44	–27.00	–1.00	57.50	2.13	159.00	5.88
Disabled vehicle	0.607	–15.22	–9.24	–3.00	–1.82	–13.00	–7.89	–27.00	–16.39	–49.00	–29.74
Other	0.255	45.92	11.71	–2.00	–0.51	–9.00	–2.30	35.50	9.05	343.00	87.47
Vehicle fire	0.002	7.88	0.02	9.00	0.02	7.00	0.01	8.00	0.02	14.00	0.03
Roadway type											
Interstate	0.830	9.93	8.24	2.00	1.66	6.00	4.98	8.00	6.64	21.00	17.43
Primary	0.040	32.37	1.29	2.01	0.08	8.00	0.32	15.00	0.60	37.00	1.48
Urban	0.007	33.26	0.23	2.03	0.01	10.00	0.07	19.00	0.13	26.00	0.18
Time of day											
a.m. peak	0.176	Base		Base		Base		Base		Base	
Midday	0.300	–12.86	–3.86	0.00	0.00	–2.00	–0.60	–5.00	–1.50	–51.00	–15.30
p.m. peak	0.161	–6.91	–1.11	0.00	0.00	–2.01	–0.32	–4.00	–0.64	–47.00	–7.57
Night	0.362	12.14	4.39	1.00	0.36	1.00	0.36	3.00	1.09	–19.00	–6.88
Day of week											
Weekday	0.767	Base		Base		Base		Base		Base	
Weekend	0.232	4.13	0.96	0.00	0.00	0	0.00	1.00	0.23	12.00	2.78
Injury count	0.017	9.86	0.17	10.50	0.18	8.00	0.14	8.00	0.14	7.00	0.12
Number of involved vehicles	0.814	5.03	4.09	3.00	2.44	4.00	3.26	5.50	4.48	5.00	4.07
Rescue responded (1–yes, 0–no)	0.029	18.48	0.54	21.00	0.61	25.00	0.73	20.00	0.58	46.00	1.33
Work zone involved (1–yes, 0–no)	0.002	–10.95	–0.02	5.00	0.01	0.00	0.00	–7.50	–0.02	–1.00	0.00
Constant	—	16.23	16.23	1.00	1.00	12.00	12.00	33.50	33.50	116	116
Estimate at means Σ(β * X)			44.10		6.68		13.86		45.45		186.54

Predicting Incident Duration

At some critical locations (such as bottlenecks) in the road network, there may be incidents that normally last longer than the regional average. If an incident occurs at such a location, then higher percentile regressions can be applied to predict the incident duration. For example, incident data in Hampton Roads show that the duration of incidents at entrances of the Hampton Roads Bridge Tunnel are longer and in the 75th percentile compared with incidents in the region. Therefore, the 75th percentile regression model can be used to obtain the initial incident duration prediction for this bottleneck. Other triggers that move the models to higher percentiles include unclassified "other" incidents (as opposed to accidents), injury counts, and number of involved vehicles. The model in Table 2 presents the 75th percentile regression for predicting the durations of future incidents at this bottleneck.

Reducing Incident Duration

In addition to incident duration prediction, quantile regression has the potential to provide transportation practitioners with solutions to reduce the duration of incidents. Specifically, the correlates of higher or lower percentile regressions can highlight factors that can potentially reduce incident durations. Incidents on Interstates for smaller incidents at the 25th percentile are associated with 2-min-longer incident durations, but at the 95th percentile, that is, for large-scale incidents, they are associated with 21-min-longer durations. Similarly, if incidents are captured through CCTV, then the durations of larger incidents may increase substantially as compared with those captured via SSP. Strategies that can reduce the number of people injured and the number of involved vehicles can also reduce the durations of larger incidents.

LIMITATIONS

The results of this study depend heavily on the accuracy of information documented in the database. The data collected were based on incident reporters and investigators. Reporting errors may exist. Further, this study analyzed a limited number of factors. If other variables are included in the model specification, the associations between incident duration and related factors may be different. The data used in this study are based on incidents that occurred in Hampton Roads, Virginia, during the 2013 to 2015 period. The results may vary if data from other areas are used for estimation. More detailed data about road geometry and incident characteristics can potentially enhance the model specification. For example, this study did not account for shoulders and ramp characteristics, if they were affected or otherwise. Such data can be added and the modeling framework enhanced to develop more appropriate incident management solutions.

CONCLUSIONS

This study applied the quantile regression technique to predict incident duration, providing a broader range of information for incident duration predictions. Unlike OLS regression models that provide estimates of average incident durations, quantile regression is able to estimate the entire distribution of incident durations by modeling its quantiles.

In general, estimates of the OLS model are within the ranges of the estimates made by the quantile regression models. This study demonstrated the estimation of quantile regression models at the 25th, 50th, 75th, and 95th percentiles. Differences between congestion- and delay-related incidents compared with accidents are greater at higher percentiles, especially at the 75th percentile, implying that congestion has a substantial influence on large incidents that normally last longer than 75% of all incidents. For factors related to the number of involved vehicles and the number of injuries, the greater coefficients are found at higher percentiles. Further, given the quantile regression estimates, this study presented a way to predict the change of probability that an incident with a given duration will occur owing to changes in values of independent variables. It is estimated that compared with the accidents, congestion- and delay-related incidents are associated with a nearly 25% increase in the probability of having an incident lasting for 100.82 min. Last, the OLS and quantile regression models were compared in relation to the accuracy of the incident duration prediction. The comparison showed that the quantile regressions using the location-based method better predicted the incident duration compared with the OLS model.

The information generated by quantile regression is useful in predicting the incident duration for certain groups of incidents, helping with incident management, especially for some areas and road segments where incidents are normally longer than other incidents. Potential applications have been discussed. They can be applied in real-life contexts, benefiting incident managers in transportation management centers. Decision support tools that can apply these models for predictive analytics in transportation management centers are under development by the research team.

ACKNOWLEDGMENTS

The authors thank Hampton Roads Smart Traffic Center, Virginia Department of Transportation, for sharing valuable data. The statistical software Stata was used for modeling. The authors are thankful for the support received from the Southeastern Transportation Center through a grant, the Center for Transportation Research, and the Transportation Engineering and Science Program in the Department of Civil and Environmental Engineering at the University of Tennessee. The Office of the Secretary of Transportation sponsorship is greatly appreciated.

REFERENCES

1. Zhang, H., M. Cetin, and A.J. Khattak. Joint Analysis of Queuing Delays Associated with Secondary Incidents. *Journal of Intelligent Transportation Systems,* Vol. 19, No. 2, 2015, pp. 192–204.
2. Ng, M., A.J. Khattak, and W.K. Talley. Modeling the Time to the Next Primary and Secondary Incident: A Semi-Markov Stochastic Process Approach. *Transportation Research Part B,* Vol. 58, 2013, pp. 44–57.
3. Hu, J., B.J. Schroeder, and N.M. Rouphail. Rationale for Incorporating Queue Discharge Flow into *Highway Capacity Manual* Procedure for Analysis of Freeway Facilities. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2286,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 76–83.
4. Zhang, H., and A.J. Khattak. Analysis of Cascading Incident Event Durations on Urban Freeways. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2178,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 30–39.
5. Khattak, A.J., X. Wang, and H. Zhang. Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2178,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 128–137.
6. Valenti, G., M. Lelli, and D. Cucina. A Comparative Study of Models for the Incident Duration Prediction. *European Transport Research Review,* Vol. 2, No. 2, 2010, pp. 103–111.
7. Lee, Y., and C.H. Wei. A Computerized Feature Selection Method Using Genetic Algorithms to Forecast Freeway Accident Duration Times. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 25, No. 2, 2010, pp. 132–148.
8. Zhang, H., and A.J. Khattak. What Is the Role of Multiple Secondary Incidents in Traffic Operations? *Journal of Transportation Engineering,* Vol. 136, No. 11, 2010, pp. 986–997.
9. Garib, A., A. Radwan, and H.N. Al-Deek. Estimating Magnitude and Duration of Incident Delays. *Journal of Transportation Engineering,* Vol. 123, No. 6, 1997, pp. 459–466.
10. Golob, T.F., W.W. Recker, and J.D. Leonard. An Analysis of the Severity and Incident Duration of Truck-Involved Freeway Accidents. *Accident Analysis and Prevention,* Vol. 19, No. 5, 1987, pp. 375–395.
11. Giuliano, G., Incident Characteristics, Frequency, and Duration on a High Volume Urban Freeway. *Transportation Research Part A,* Vol. 23, No. 5, 1989, pp. 387–396.
12. Khattak, A.J., H.M. Al-Deek, and R.W. Hall. Concept of an Advanced Traveler Information System Testbed for the Bay Area: Research Issues. *Journal of Intelligent Transportation Systems,* Vol. 2, No. 1, 1994, pp. 45–71.
13. Khattak, A.J., X. Wang, and H. Zhang. Incident Management Integration Tool: Dynamically Predicting Incident Durations, Secondary Incident Occurrence, and Incident Delays. *IET Intelligent Transport Systems,* Vol. 6, No. 2, 2012, pp. 204–214.
14. Koenker, R. *Quantile Regression.* Cambridge University Press, United Kingdom, 2005.
15. Junhua, W., C. Haozhe, and Q. Shi. Estimating Freeway Incident Duration Using Accelerated Failure Time Modeling. *Safety Science,* Vol. 54, 2013, pp. 43–50.
16. El-Basyouny, K., and T.A. Sayed. Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1950,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 9–16.
17. Li, R., and P. Shang. Incident Duration Modeling Using Flexible Parametric Hazard-Based Models. *Computational Intelligence and Neuroscience,* Vol. 2014, 2014, p. 33.
18. Wang, X., S. Chen, and W. Zheng. Traffic Incident Duration Prediction Based on Partial Least Squares Regression. In *Procedia—Social and Behavioral Sciences,* Vol. 96, 2013, pp. 425–432.
19. Chung, Y. Development of an Accident Duration Prediction Model on the Korean Freeway Systems. *Accident Analysis and Prevention,* Vol. 42, No. 1, 2010, pp. 282–289.
20. Zou, Y., K. Henrickson, D. Lord, Y. Wang, and K. Xu. Application of Finite Mixture Models for Analyzing Freeway Incident Clearance Time. *Transportmetrica A: Transport Science,* Vol. 12, No. 2, 2016, pp. 99–115.

21. Qi, Y., and H. Teng, An Information-Based Time Sequential Approach to Online Incident Duration Prediction. *Journal of Intelligent Transportation Systems,* Vol. 12, No. 1, 2008, pp. 1–12.
22. Ji, Y. Prediction of Freeway Incident Duration Based on the Multi-Model Fusion Algorithm. Presented at 2011 International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE), Nanjing, China, 2011.
23. Chang, H.-L., and T.-P. Chang. Prediction of Freeway Incident Duration Based on Classification Tree Analysis. *Journal of the Eastern Asia Society for Transportation Studies,* Vol. 10, 2013, pp. 1964–1977.
24. Park, H., A. Haghani, and X. Zhang. Interpretation of Bayesian Neural Networks for Predicting the Duration of Detected Incidents. *Journal of Intelligent Transportation Systems,* Vol. 19, No. 1, 2015, pp. 1–16.
25. Wei, C.-H., and Y. Lee. Sequential Forecast of Incident Duration Using Artificial Neural Network Models. *Accident Analysis and Prevention,* Vol. 39, No. 5, 2007, pp. 944–954.
26. Vlahogianni, E. I., and M. G. Karlaftis. Fuzzy-Entropy Neural Network Freeway Incident Duration Modeling with Single and Competing Uncertainties. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 28, No. 6, 2013, pp. 420–433.
27. Lin, P.-W., N. Zou, and G.-L. Chang. Integration of a Discrete Choice Model and a Rule-Based System for Estimation of Incident Duration: A Case Study in Maryland. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
28. He, Q., Y. Kamarianakis, K. Jintanakul, and L. Wynter. Incident Duration Prediction with Hybrid Tree-Based Quantile Regression. *Advances in Dynamic Network Modeling in Complex Transportation Systems,* Vol. 2, 2013, pp. 287–305.
29. Xiaoqiang, Z., L. Ruimin, and Y. Xinxin. Incident Duration Model on Urban Freeways Based on Classification and Regression Tree. Presented at Second International Conference on Intelligent Computation Technology and Automation (ICICTA '09), Hunan, China, 2009.
30. Kim, W., S. Natarajan, and G.-L. Chang. Empirical Analysis and Modeling of Freeway Incident Duration. Presented at 11th International IEEE Conference on Intelligent Transportation Systems (ITSC 2008), Beijing, 2008.
31. Li, R., F. C. Pereira, and M. E. Ben-Akiva. Competing Risk Mixture Model and Text Analysis for Sequential Incident Duration Prediction. *Transportation Research Part C,* Vol. 54, 2015, pp. 74–85.
32. Machado, J. A. F., and J. S. Silva. Quantiles for Counts. *Journal of the American Statistical Association,* Vol. 100, No. 472, 2005, pp. 1226–1237.
33. Qin, X., M. Ng, and P. E. Reyes. Identifying Crash-Prone Locations with Quantile Regression. *Accident Analysis and Prevention,* Vol. 42, No. 6, 2010, pp. 1531–1537.
34. Qin, X. Quantile Effects of Causal Factors on Crash Distributions. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2279,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 40–46.
35. Ruimin, L., Z. Xiaoqiang, Y. Xinxin, L. Junwei, C. Nan, and Z. Jie. Incident Duration Model on Urban Freeways Using Three Different Algorithms of Decision Tree. Presented at 2010 International Conference on Intelligent Computation Technology and Automation (ICICTA), Changsha, China, 2010.
36. Zhang, H., and A. Khattak. Spatiotemporal Patterns of Primary and Secondary Incidents on Urban Freeways. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2229,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 19–27.
37. Schrank, D., B. Eisele, and T. Lomax. *2012 Urban Mobility Scorecard.* Texas A&M Transportation Institute, College Station, 2012.

The views presented in this paper are those of the authors, who are responsible for the facts and the accuracy of the information provided.

The Standing Committee on Freeway Operations peer-reviewed this paper.