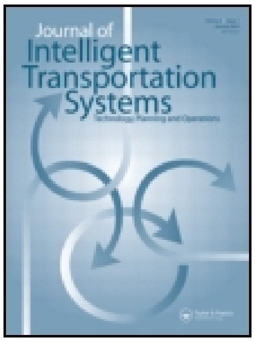

Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: 1547-2450 (Print) 1547-2442 (Online) Journal homepage: http://www.tandfonline.com/loi/gits20

Short-Term Traffic Flow Forecasting for Freeway Incident-Induced Delay Estimation

Runze Yu, Yunteng Lao, Xiaolei Ma & Yinhai Wang

Journal of Intelligent Transportation Systems, 18:254–263, 2014

ISSN: 1547-2450 print / 1547-2442 online
DOI: 10.1080/15472450.2013.824757

Short-Term Traffic Flow Forecasting for Freeway Incident-Induced Delay Estimation

RUNZE YU, YUNTENG LAO, XIAOLEI MA, and YINHAI WANG
Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington, USA

Freeway incidents not only threaten travelers' safety but also cause severe congestion. Incident-induced delay (IID) refers to the extra travel delay resulting from incidents on top of the recurrent congestion. Quantifying IID would help people better understand the real cost of incidents, maximize the benefit-cost ratio of investment on incident remedy actions, and develop active traffic management and integrated corridor management strategies. By combining a modified queuing diagram and short-term traffic flow forecasting techniques, this study proposes an approach to estimate the temporal IID for a roadway section, given that the incidents occurs between two traffic flow detectors. The approach separates IID from the total travel delay, estimates IID for each individual incident, and only takes volume as input for IID quantification, avoiding using speed data that are widely involved in previous algorithms yet are less available or prone to poor data quality. Therefore, this approach can be easily deployed to broader ranges where only volume data are available. To verify its estimation accuracy, this study captures two incident videos and extracts ground-truth IID data, which is rarely done by previous studies. The verification shows that the IID estimation errors of the proposed approach are within 6% for both cases. The approach has been implemented in a Web-based system, which enables quick, convenient, and reliable freeway IID estimation in the Puget Sound region in the state of Washington.

Keywords Congestion; Deterministic Queuing Theory; Incident Induced Delay; Ridge Regression; Short-Term Traffic Flow Forecast

INTRODUCTION

Aside from their adverse impacts on traffic safety, freeway incidents have been identified as one of the major reasons for congestion. A Federal Highway Administration report (Cambridge Systematics, 2005) indicated that approximately 50% of the total congestions on freeways are non-recurrent. The situation is even more severe in urban areas. Approximately one-half to two-thirds of the total travel delay in large metropolitan areas is incident relevant (Center for Urban Transportation Research, 2010). Since congestion mitigation and safety enhancement are among the main goals of most transportation agencies, a number of state (or other local) departments of transportation have invested in incident response programs in a variety of forms.

Making smart decisions for such investments necessitates proper quantification of incident cost. This cost can be reasonably inferred from incident-induced delay (IID), which refers to the extra temporal delays on a roadway segment affected by the incident. IID may further result in secondary incidents, which adds more incident cost to the primary incidents. Two challenges emerge regarding IID quantification for individual incidents. First, it is not an easy task to separate IID from total delay, which comprises both non-recurrent delay (e.g., IID) and recurrent delay. Second, applying the widely used deterministic queuing theory (DQT) methods to IID quantification requires both volume and speed data. Due to the scarcity or poor quality of speed data, the range of applications for previous algorithms is limited.

In this study, a new approach for quantifying IID is introduced. This approach overcomes both aforementioned challenges, by combining modified queuing diagrams and short-term traffic flow forecasting techniques. Both ridge regression and lagged regression are evaluated and compared for their forecasting capabilities. Video cameras are deployed to verify results of the proposed approach, which was rarely done in previous research. The verification demonstrated that the new approach yielded accurate IID estimates with proper

Address correspondence to Runze Yu, Department of Civil and Environmental Engineering, University of Washington, Box 352700, Seattle, WA 98195-2700, USA. E-mail: runze@uw.edu

calibrations. Finally, the approach is implemented in a Web-based system, for large-scale regional applications.

BACKGROUND

Quantifying travel delay is essentially a travel time estimation problem under different traffic conditions, since total travel time is equal to the sum of free-flow travel time, extra travel time caused by recurrent congestion (recurrent delay), and extra travel time cause by non-recurrent congestion. Specifically, IID refers to the extra travel time that all the incident-impacted drivers take to traverse a certain roadway section in comparison with the incident-free scenario.

Historically, how to quantify the travel delay during incidents has attracted numerous research efforts. In terms of research methodology, both DQT-based algorithms (Morales, 1987; Fu & Rilett, 1997; Mak & Fan, 2006) and shockwave-based algorithms (Wirasinghe, 1978; Mongeot & Lesort, 2000) are widely applied. In terms of study data, researchers utilize either traffic simulation outputs (Kamga et al., 2011) or traffic detector measurements (Skabardonis et al., 1996, 2003). In terms of research objective, some researchers study the total delay when incidents occurred (Lindley, 1987; Sullivan, 1997), while others (Skabardonis et al., 2003; Cheevarunothai, 2010; Güner et al., 2011; Chuang, 2011; Cheevarunothai et al., 2012) specifically focused on IID.

Shockwave-based algorithms apply the concepts and modeling techniques of traffic flow theories. Common approaches usually followed the procedure characterized by roadway segmentation, delay estimation for each segment using traffic flow models, and accumulation of segmental delays along the roadway (Mongeot & Lesort, 2000). While providing well-defined theoretical derivations, those models are usually too mathematically complicated to be supported by current traffic data collection infrastructure, creating difficulties for algorithm implementation.

Another widely used approach is DQT. DQT methods calculate delay as the area enclosed by the arrival and departure curves. Therefore, key parameters for delay calculation with DQT methods include arrival rate, departure rate, incident duration, and capacity. Some researchers estimate those parameters in a deterministic fashion (Morales, 1987; Lindley, 1987; Sullivan, 1997), where fixed values are assigned to each parameter based on engineering judgment and historical data. Some other researchers consider those parameters from a stochastic perspective (Qi et al., 2007; Fu & Rilett, 1997; Fu & Hellinga, 2002; Li et al., 2006), where the key parameters are random variables rather than fixed values. Most previous studies using either approach focus on estimating the delay statistics (e.g., mean, standard deviation) for a certain type of incident (e.g., abandoned vehicles, injury collisions, etc.), and thus the estimation accuracy will be impaired when applied to individual incidents.

To estimate delays of individual incidents, some researchers directly apply DQT to traffic detector outputs, which, however, will generate biased estimations (Bertini & Lovell, 2009; Fries et al., 2007). The reason lies in the physical distance between upstream and downstream (U/D) traffic detectors. Theoretically, DQT calculates delay given that arrival and departure flows are captured at the same location (e.g., where incident occurs), but in the reality arrival and departure flows (the demand and capacity curves) are observed by two traffic detectors placed upstream and downstream of the incident location. The travel time from upstream (and downstream) traffic detector locations to incident location must be excluded from the area between arrival and departure curves, which can be done by virtually moving the arrival and departure curves along the timeline (horizontal axis in queuing diagram) to the incident occurrence time, as shown in Figure 1. This process can be realized through calculating the average travel time (Rakha & Zhang, 2005) from the upstream (and downstream) locations to the incident location, where multiple speed data are needed.

Assuming the travel times are accurately estimated and vehicles will not change their routes after an incident, the area between arrival and departure curves still accounts for the total delay with the incident, which consists of both recurrent delay and non-recurrent delay. Since IID is of particular interest in many studies, recurrent delay needs to be estimated and subtracted from the total delay with incident. Quantifying recurrent delay also requires speed data, in order to estimate the travel time when no incident occurs. For example, Skabardonis et al. (1996, 2003) used loop detector data to retrieve IID from the total delay, whose algorithms require high-quality speed information.

In summary, to quantify IID using traffic detector measurements, existing algorithms or models need accurate speed data, either for moving arrival/departure cumulative volume curves or for calculating recurrent delay. However, accurate speed data are not as widely available as traffic volume data, or suffer from lower data quality. Take the loop detector, for example; due to the absence of dual-loop stations or biased

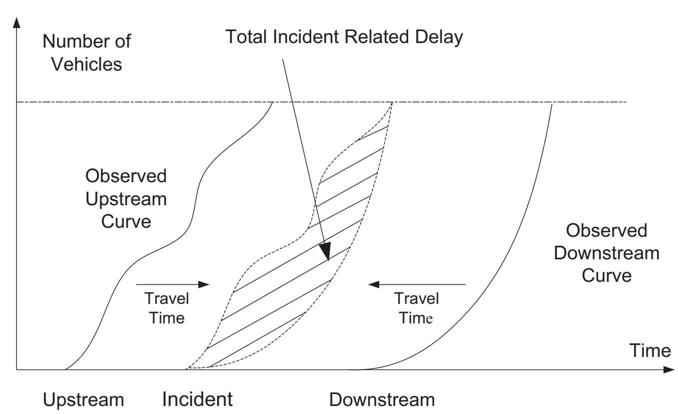

Figure 1 DQT with travel time shifts. The shaded area quantifies the total incident related delay.

occupancy measurements resulting from single loops with incorrect sensitivities (Cheevarunothai et al., 2006; Wang et al., 2009), speed data are highly unreliable. Even if quality measurements of point speeds are available, the accuracy of travel time estimates may still vary depending on how vehicle speeds vary between point measurements, which might become a critical issue during unstable traffic flow conditions. Therefore, a method that only requires the volume input from traffic detector is highly desirable for IID quantification.

METHODOLOGY

As reviewed previously, there are two challenges for IID quantification using traffic detector data: (1) to properly address the separation of the arrival/departure cumulative volume curves by upstream/downstream (U/D) detectors, because the separation of the arrival/departure cumulative volume curves runs counter to the DQT assumptions that vehicle arrivals and departures should be measured at the same point; and (2) recurrent delay must be estimated and excluded from the total delay to obtain the IID. The solutions to both challenges are described in the following sections.

Modified Queuing Diagram

This study proposes a modified deterministic queuing diagram to address the first challenge. Figure 2 shows an example of IID quantification in such a diagram, where arrival and departure flow curves are separated along the timeline, and the horizontal distance between two curves represents the travel time from the upstream traffic detector location to the downstream one. Thus the whole area enclosed by the upstream arrival and downstream departure curves is the total travel time for all the incident-affected vehicles on the freeway segment bounded by the two traffic detectors.

Some previous studies also utilize similarly modified queuing diagram for link travel time estimations (van Woensel et al., 2008; Nam and Drew, 1996). However, IID cannot be directly retrieved from the diagram, because it reflects the total travel time between two traffic detectors. To estimate IID, both recurrent delay and free flow travel time need to be excluded from the observed total travel time. As mentioned earlier in the second challenge, however, this is not easy because given a specific arrival curve, the departure curve under the incident-free scenario could not be observed directly. If this unobservable departure curve can be virtually formed up, then it is straightforward to estimate IID using the modified DQT. Therefore, the key is to construct the departure curve under the incident-free scenario.

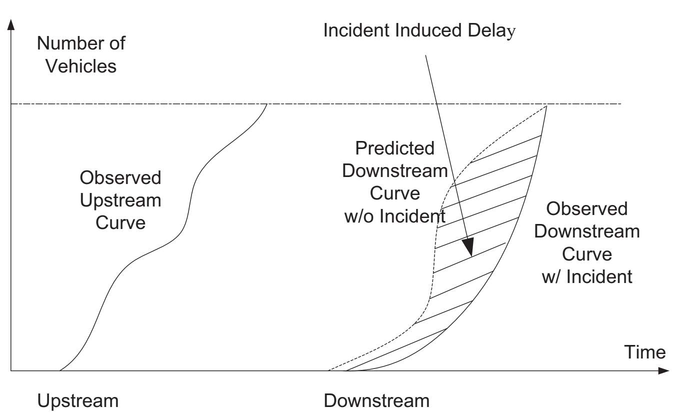

Figure 2 Illustration of the modified queuing diagram.

When there is no incident, the relationship between the U/D flows is assumed to be predictable under certain traffic conditions. This is a valid assumption since if the traffic is under free flow condition (with zero delay) or recurrently congested condition (with predictable spatial and temporal information of congestion), the relationship between the U/D flows stays relatively consistent. Even if the traffic flow is in a transition period, from free flow to congestion or from congestion to free flow, the U/D flows still follow a certain relationship. Therefore, authors use the observed upstream flow curve to predict the downstream flow curve under the incident-free scenario. The feasibility of this approach has been demonstrated in previous short-term traffic flow forecasting studies (e.g., Hobeika and Kim, 1994; Abdulhai et al., 1999).

As shown in Figure 2, the modified DQT approach needs three curves for IID estimation: the observed upstream arrival curve (left solid curve in Figure 2), the observed downstream departure curve for the incident scenario (right solid curve), and the predicted downstream departure curve for the incident-free scenario (right dashed-line curve). IID corresponds to the area enclosed by the observed downstream flow curve with incident and the predicted downstream flow curve without incident. A mathematical expression of the total vehicle travel time with incident VD_{WI} and total vehicle travel time without incident VD_{WOI} can be defined in Eqs. 1 and 2, respectively:

$$VD_{WI} = \int_{t_I}^{t_c} \left(\int_{t_I - t_{UI}}^{t_c} n_{U,i}dt - \int_{t_I + t_{ID}}^{t_c} n_{D,i}dt \right) dt \qquad (1)$$

$$VD_{WOI} = \int_{t_I}^{t_c} \left(\int_{t_I - t_{UI}}^{t_c} n_{U,i}dt - \int_{t_I + t_{ID}}^{t_c} n_{D,i}^{WOI}dt \right) dt \qquad (2)$$

where $n_{U,i}$ and $n_{D,i}$ are upstream and downstream vehicle counts, respectively, in time interval i; $n_{D,i}^{WOI}$ is the predicted downstream vehicle count in time interval i assuming no incident; t_I is the time when the incident occurred; t_c is the time when the incident is cleared; and t_{UI} and t_{ID} are travel times from the upstream loop station to the incident location and from the incident location to the downstream loop station, respectively. Note that upstream vehicles spend t_{UI} to arrive at the incident location; therefore, starting from time $t_I - t_{UI}$, the arriving vehicles recorded by upstream detector have been affected. Similarly, after incident occurs, it takes a vehicle t_{ID} to travel to the downstream detector location to be recorded; therefore, only those vehicles captured by the downstream detector after time $t_I + t_{ID}$ are affected by the incident.

Then IID can be calculated as the difference between VD_{WI} and VD_{WOI}:

$$IID = VD_{WI}\text{-}VD_{WOI} = \int_{i=t_I} \left(\int_{i=t_I\ +\ t_{ID}} n_{D,i}^{WOI} dt\text{-} \int_{i=t_I\ +\ t_{ID}} n_{D,i} dt \right) dt \quad (3)$$

In Eq. 3, $\int_{i=t_I+t_{ID}} n_{D,i}^{WOI} dt$corresponds to the predicted downstream curve (the dashed line in Figure 1), and $\int_{i=t_I+t_{ID}} n_{D,i} dt$corresponds to the observed downstream curve (the right solid line in Figure 1). During the time period $(t_I, t_I + t_{ID})$ when vehicles travel from the incident location to the downstream loop station, $n_{D,i}^{WOI} \approx n_{D,i}$ because the incident only affects the downstream loop station after $t_I + t_{ID}$. Then Eq. 3 can be derived as:

$$IID = \int_{i=t_I} \left(\int_{i=t_I\ +\ t_{ID}} n_{D,i}^{WOI} dt\text{-} \int_{i=t_I\ +\ t_{ID}} n_{D,i} dt + \int_{i=t_I}^{t_I\ +\ t_{ID}} n_{D,i}^{WOI} dt - \int_{i=t_I}^{t_I\ +\ t_{ID}} n_{D,i} dt \right) dt = \int_{i=t_I} \left(\int_{i=t_I} n_{D,i}^{WOI} dt\text{-} \int_{i=t_I} n_{D,i} dt \right) dt \quad (4)$$

Equation 4 indicates that proposed approach avoids speed or travel time (t_{UI} and t_{ID}) estimation and thus significantly eases the calculation process.

Regression Techniques for Prediction

Finding the relationship between U/D flows in an in an incident-free scenario during an incident can be considered as a regression problem. Since most traffic detectors report traffic counts periodically, on a freeway segment a fleet of vehicles first detected by the upstream detector will be counted by the downstream detector after several time intervals. When conducting a multiple regression of the downstream volume to the upstream volume, the predictor variables are highly correlated. Given the multicollinearity in regression, two prediction methods, lagged regression and ridge regression, are investigated in this study.

Lagged regression is a time-series analysis technique (Shumway & Stoffer, 2006). It investigates the correlations between downstream flow count series from time step i and the upstream flow count series from time steps i–r, which actually account for both spatial and temporal correlations. Here r is the time lag between two series and takes values of $r = 1$, 2, 3 . . ., and the lagged regression evaluates which r value corresponds to a correlation stronger than a user-specified level. The lagged regression model used for this study is shown in Eq. 5:

$$y_t = \beta_0 + \sum_{r=l_1}^{l_k} a_r x_{t\text{-}r} + \omega_t \quad (5)$$

where y_t represents the downstream flow count series; x_{t-r} denotes the upstream flow count series with a time shift of r intervals from the downstream series; β_0 is the constant; l_1 to l_k are the lags at which correlations between inputs and outputs are above the correlation threshold; a_r is the lag regression coefficients; and ω_t is white noise.

Ridge regression is one of the nonparametric regression techniques (Draper & Smith, 1998). Unlike multiple linear regression, it uses goodness of prediction as parameter estimation criterion rather than goodness of fit. As shown in Eq. 6, ridge regression shares the same mathematical expression as multiple linear regression:

$$y_i = \beta_0 + \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon \quad (6)$$

An important distinction between ridge regression and multiple linear regression is that ridge regression applies a different procedure for model parameter estimation. Equations 7 through 11 show the parameter estimation procedure with ridge regression. Note that a penalty term λ, called the smoothing parameter, is added to constrain the parameters $\beta_1, \beta_2, \ldots \beta_k$ in Eq. 7:

$$min\left(\sum_{i=1}^{n} \left(y_i\text{-}\beta_0\text{-}\sum_{j=1}^{k} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{k} \beta_j^2 \right) \quad (7)$$

For a specific λ, minimizing Eq. 7 is equivalent to minimizing

$$min\left(\sum_{i=1}^{n} \left(y_i\text{-}\beta_0\text{-}\sum_{j=1}^{k} x_{ij}\beta_j \right)^2 \right) subject\ to \sum_{j=1}^{k} \beta_j^2 s \quad (8)$$

where s is the constraint to $\beta_1, \beta_2, \ldots \beta_k$, which plays a role similar to λ. Solving Eq. 8 yields

$$\beta = (x^T x + \lambda D)^{-1} x^T Y \quad (9)$$

where D is an identity matrix. For smoothing parameter selection, a cross-validation technique is applied: A sample of data is partitioned into two complementary data sets, one for model training and the other for validation; then values of ordinary cross validation (OCV) and generalized cross validation

(GCV) are calculated using Eqs. 10 and 11, respectively:

$$OCV^{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{(y_i - S_{ii})} \tag{10}$$

$$GCV^{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{(1\text{-}tr(S)/n)^2} \tag{11}$$

where $S^{\lambda} = (x^T x + \lambda D)^{-1} x^T$, and $tr(S)$ is the trace of S. The λ value with the lowest OCV or GCV score will be chosen. In this way, cross validation essentially compares different models and selects the one with the best predictions based on the training data set.$\hat{y}$

In summary, the approach proposed for IID quantification involves developing regression models based on historical U/D volume counts free of incidents. When an incident occurs, the real-time upstream volume counts are input to the model to generate the downstream volume counts as if there were no incidents. Finally, model outputs and real-time downstream volume counts are applied to Eq. 4 to calculate IID.

DATA

Incident Data

Incidents recorded in the 2009 Washington Incident Tracking System (WITS) database are used in this study. This database is maintained by the Washington State Department of Transportation (WSDOT) Incident Response (IR) team. *State Route ID*, *direction*, and *mile post* jointly define an incident's location. *Notification time* records the time when an incident was reported to the IR program. Since most incidents were reported through mobile phones, notification time is very close to the start time of the incident. *Arrival time* stores the time when an IR truck arrived at the incident location. *All lanes open time* is the time when all lanes became open to traffic. *Clear time* is the time when the incident had been fully cleared and the IR teams left the incident scene. Incident duration is calculated as the time interval from *Notification time* to *Clear time*.

Loop Detector Data

Loop detector data associated with recorded incidents are acquired from WSDOT and archived in the data server of the Smart Transportation Application and Research Laboratory (STAR Lab) at the University of Washington (Wang et al., 2009). Since most of the loops are configured as single-loop detectors, only volume and occupancy data assure their availability. Data are reported periodically every 20 seconds.

IMPLEMENTATION

Prediction Accuracy Evaluation

Since predicting the downstream flows under the incident-free scenario is a key component in the proposed approach, prediction accuracies of the two regression techniques are compared. The measures for the comparison are mean squared error (MSE) and index of agreement (IA).

Mean squared error (MSE) is widely used as a measure for quantifying the difference between the estimated value and the observed value. It is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{12}$$

where n is the number of observations, $\hat{y}_i$ is the estimated value, and y_i is the observed value. Despite its widespread use, MSE is criticized in that its value is highly related to the magnitude of the observed values. One remedy to this problem would be taking the mean of observed values into consideration. Index of agreement (IA) is proposed by Willmott (1981) as a quantitative indication for the quality of model prediction. It is defined as

$$IA = 1\text{-}\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(|y_i - \bar{y}| + |y_i - \bar{y}|)^2} \tag{13}$$

where n is the number of observations, $\hat{y}_i$ and y_i are the same as in Eq. 12, and $\bar{y}$ is the mean of observed values. IA is dimensionless and scaled between 0 and 1. When IA = 1, perfect agreement is achieved between the estimated and observed values.

Prediction accuracy test is conducted at two locations, one on I-5 northbound between mileposts 157.92 (upstream) and 158.92 (downstream), and the other one on SR-520 westbound between mileposts 1.58 (downstream) and 4.10 (upstream). For each location, a time period free of incident is randomly chosen and corresponding U/D 20-second loop volume data are retrieved. The gray cells in Table 1 are the prediction accuracy for both locations. The ridge regression achieves satisfactory prediction accuracy: IA at the I-5 location (or SR-520) is 0.92 (or 0.87), meaning that the predicted values are 92% (or 87%) agreed with the observed values.

The calibrated model for each location is then applied to three other scenarios to test the model transferability: (1) the same location but on another date; (2) the same date but at another location; and (3) at another location and on a different date. All the locations and date are chosen randomly, to test how well the calibrated model can be applied to different dates and/or locations. Detailed information about training data and test data and test results are also summarized in Table 1. The

Table 1 Prediction accuracy and transferability test results.

Training data			Testing scenario			Prediction method			
	Location			Location		Ridge regression		Lagged regression	
Time	Upstream milepost	Downstream milepost	Time	Upstream milepost	Downstream milepost	IA	RMSE	IA	RMSE
10/15/2009	I-5 157.92	I-5 158.92	10/15/2009	157.92	158.92	0.92	3.48	0.78	5.06
			11/11/2009	157.92	158.92	0.64	6.18	0.38	8.09
			10/15/2009	139.69	140.64	0.74	6.23	0.38	8.90
			11/11/2009	189.98	190.9	0.77	6.30	0.6	8.11
11/18/2009	SR-520 4.1	SR-520 1.58	11/18/2009	4.1	1.58	0.87	2.95	0.64	3.45
			10/14/2009	4.1	1.58	0.52	3.92	0.27	4.36
			11/18/2009	7.98	6.01	0.47	6.62	0.47	6.61
			10/14/2009	7.98	6.01	0.44	6.13	0.44	6.14

test results from Table 1 indicate that the relationship between upstream and downstream volume is related to the location and time period. If we have a high requirement on the accuracy, it is better to estimate the model parameters from location to location.

For the I-5 location, ridge regression performed better than lagged regression in terms of prediction transferability. IA values for ridge regression are generally higher than those of lagged regression. MSE values for ridge regression at I-5 location are half lower than those for lagged regression. This is because the lagged regression coefficients are more relevant to the travel time between the loop stations. For the SR-520 location, ridge regression outperformed lagged regression when applied to a different date, with a higher IA and a lower MSE. In the last two scenarios, ridge regression and lagged regression both had lower prediction accuracies at approximately the same level. IA values for these two scenarios are less than 50%. One reason might be that training data set for the SR-520 case was from both ends of a floating bridge, where the traffic characteristics were different from that at the location for training scenarios 2 and 3, which was a regular freeway section.

This comparison indicates that ridge regression had better or equally good prediction accuracy than lagged regression at these two test sites, which generally holds for other locations. Because the parameter estimation for ridge regression is conducted for better prediction, it uses cross validation for smoothing parameter selection—a technique to evaluate how well the predictions from the trained model can be generalized to other independent datasets.

Additionally, ridge regression considers more variations in traffic than lagged regression. Lagged regression calculates cross-correlation function to find the most correlated previous time steps. Once the model is formulated, only time steps with higher correlation will be kept in the model. In this study ridge regression keeps all the upstream volumes from the previous 15 time steps, and therefore it captures more traffic dynamics that occurred.

Last but not least, compared to the lagged regression, ridge regression is more flexible in terms of choosing U/D loop stations. The most correlated time steps picked up by lagged regression are strongly influenced by the distance between the U/D loop stations, because when traffic flow stays stable, it will take a certain time for vehicles to travel from upstream to downstream locations. This travel time is mainly determined by the distance between the U/D loop stations. However, when applying the trained model to another U/D pair along a specific corridor, with a longer or shorter distance than that in the training data set, the trained model will probably fail to perform well. Ridge regression is a nonparametric regression technique, which does not rely on specific parameters but shrinks the input data to best fit the relationship between inputs and outputs. Therefore, the authors chose ridge regression over lagged regression for implementation.

Model Calibration

It is desirable to develop an individual prediction model for each location and flow condition to enhance the downstream flow prediction accuracy, since the U/D relationship changes correspondingly when traffic flow conditions change and transferring one model to other locations will impair the prediction accuracy, as shown in Table 1. Even for one specific location, different models should be calibrated to accommodate varying traffic flow levels and trends. Based on the statistical analysis, four phases are considered in this research: (1) free flow phase: low volume and high speed, usually from midnight to early morning; (2) increasing phase: the transition phase from free flow to morning peak, with consistently increasing volume until morning peak congestion forms up; (3) daytime phase: even though it might experience morning and/or afternoon peak, traffic demand maintains at a high level over a long time period during the day; and (4) decreasing phase: the transition phase from daytime flow to free flow phase, with consistently decreasing volume. For illustration,

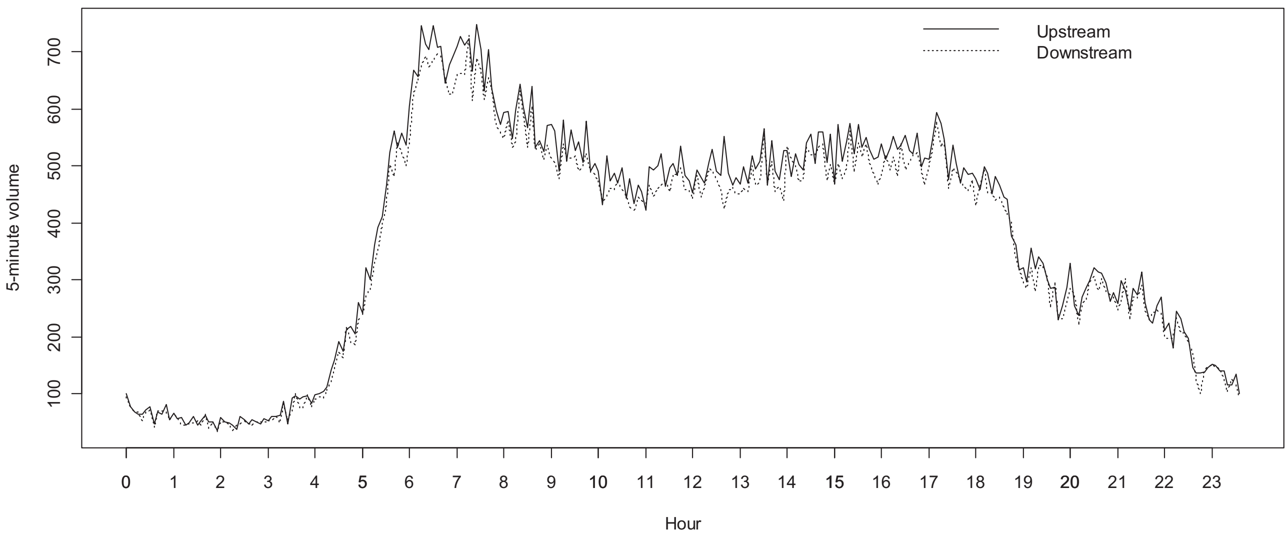

Figure 3 Typical traffic variations over a day.

traffic flow variation is shown in Figure 3, in one day free of incidents on I-5 northbound between milepost 158 and 160. In Figure 3, Phases 1, 2, 3, and 4 roughly correspond to 12 a.m. to 5 a.m., 5 a.m. to 7 a.m., 7 a.m. to 6 p.m., and 6 p.m. to 12 a.m. of the next day, respectively. In this study a specific prediction model is calibrated for each phase.

As traffic phases reflect different temporal flow characteristics, there are four other factors representing various spatial flow characteristics: route, milepost, direction, and lane number. Once a new combination of the factors is determined, a new set of model parameters should be estimated. In this study, 22 types of roadway segments are identified and calibrated, and 88 different models are estimated to represent a variety of scenarios.

Note that the upstream detector will be impacted by the incident once vehicles queue up to its location, which will lower the proposed approach's accuracy. Therefore, a good strategy is to choose an upstream detector sufficiently far away from the incident location without being affected by the incident and meanwhile close enough to the downstream detector so that the U/D relationship is well maintained.

Algorithm Verification

To the authors' knowledge, there are few similar algorithms developed for only calculating the IIDs because most of the previous studies calculate total delay under incident scenarios, which includes both IID and recurrent delay. In addition, algorithm verification using video camera data is rarely conducted in previous studies, since incident occurrence is too unpredictable to be captured by cameras beforehand. In this study, the authors monitored several incident-inclined freeway corridors for several months and two video clips of incidents were recorded for verification purpose. One is on January 26, 2010, on northbound I-5, milepost 158.57 near Boeing Field in Seattle, from 7:53 a.m. to 8:36 a.m.; the other is on April 9, 2010, on westbound SR520 bridge milepost 4.06, from 4:43 p.m. to 5:08 p.m., both in the state of Washington. Both incidents are rear-end accidents and have only property damage. For each incident, the authors captured videos from both an upstream camera and a downstream camera near the corresponding loop detectors. To obtain ground-truth IID from the video, the authors use the average travel time of eight vehicles within a 5-minute interval just before incident to determine travel time without incident. This is reasonable if we assume that the traffic state will not change significantly in a short time. If the incident lasts very long, up to 2 hours or more, the historical data in the same period should be considered. The whole incident duration is broken into 30-second intervals. For each interval, a signature vehicle is identified and its travel time is measured and used as representative travel time for all the vehicles within the interval. Number of vehicles during each interval is also counted at the downstream camera. In the end, time difference in vehicle-minutes (veh-min) between the incident and no-incident scenarios results in the ground-truth IID. The evaluation results are summarized in Table 2.

As shown in Table 2, the proposed approach using ridge regression techniques generates accurate results compared to the ground truth. Percentage errors for both incidents are below 6%. For comparison purpose, lagged regression is also applied for downstream traffic flow prediction. In both verification cases, ridge regression gives more accurate IID estimations than lagged regression. The authors attempt to apply other DQT-based methodologies using loop detectors proposed in previous literature; however, for both cases speed data are not available, leaving this comparison for future exploration.

Table 2 Verification results.

	I-5 Incident 43 min 3363		SR520 Incident 25 min 1228	
Duration Downstream volume	IID (veh-min)	Percentage error	IID (veh-min)	Percentage error
Video	9332	—	8973	—
Ridge	9464	1.40%	8494	–5.60%
Lagged	8961	–4.00%	10,369	15.90%

Results

The proposed approach is coded in Java and applied to the 2009 WITS data set to automate the calculation. The calculation procedure is summarized in Figure 4, and results are summarized by incident types in Table 3. The calculation procedure includes three components: data cleansing component, model training component, and IID calculation component. Estimation results from Table 3 reveal that collision-related incidents, especially incidents with injury, are associated with longer delays. Note that the mean values are significantly larger than the corresponding median values. The reason lies in that of all the 2,523 incidents associated with a nonzero IID, 58.7% of them (1,481 incidents) have an IID of less than 10 vehicle-hours, implying an asymmetrical distribution of IIDs.

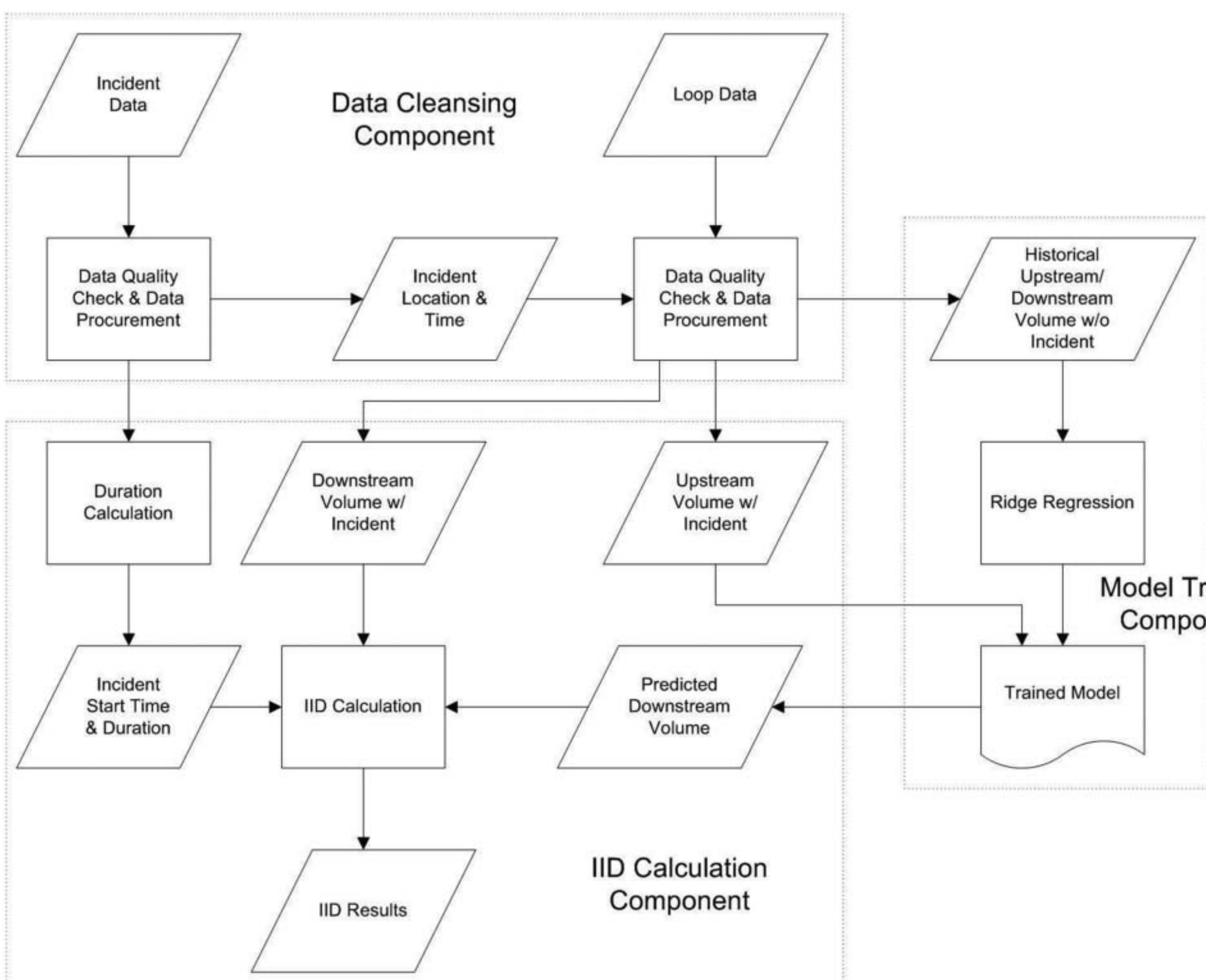

Figure 4 Implementation process.

Table 3 Statistics of incident-induced delays (in vehicle-hours).

Incident type	Frequency	Mean	SD	Median	Maximum
Abandoned vehicle	489	6	33	1	454
Debris blocking traffic	243	17	51	2	591
Disabled vehicle	1618	15	79	1	1993
Injury collision	38	340	628	156	3624
Non-injury collision	135	117	311	7	2460

Additionally, the authors observe that applying the proposed approach to some incidents yielded zero delay estimation. A closer look at those incidents showed that most of them are with incident durations shorter than 10 minutes and occur in light-traffic periods. Very likely, the reduced capacities by these incidents are still above the demands and hence these incidents do not result in significant differences between the downstream flow and predicted flow under incident-free scenarios. Also, applying one calibrated model over different locations and dates may degrade the accuracy to some extent, but the results are still fairly acceptable, as shown in Table 1 earlier. This may be due to the fact that for each location daily traffic follows similar patterns over time and thus can be reflected by a consistent U/D relationship. Location-specific models can be developed to further improve the results when necessary. Considering the ease of computing using the online implementation of the proposed approach, parameter estimation for each location is feasible in practice.

CONCLUSIONS

Incidents are major contributors to urban freeway congestion. Incident-induced delay (IID) quantification helps traffic practitioners and policymakers better evaluate their investment in congestion mitigation measures and develop more effective incident response strategies. In this study, a new approach, based on a modified queuing diagram and short-term traffic flow forecasting, is developed for quantifying IID on freeways. The first advantage of this new approach is that it only requires traffic volume data for IID estimates, making it appealing to most transportation agencies for large-scale implementation. Moreover, the new approach in this article is quite flexible in that it can utilize traffic volume data aggregated in various time periods (e.g., 20 seconds or per minute), and it can be used to calculate IID over either the full duration of the incident or any portion of it. Last but not least, given U/D flow information under different conditions, this approach can be configured for estimation of recurrent delay or non-recurrent delay as needed.

The key technology to enable such an approach is to accurately predict the unobservable downstream volumes under the incident-free scenario. Two regression techniques, lagged regression and ridge regression, are investigated in this study for downstream volume prediction. Their prediction accuracies are evaluated and compared against each other. Ridge regression is chosen for algorithm implementation because it produces better prediction results. The predicted downstream volumes by the ridge regression model are then combined with observed downstream volumes in the modified queuing diagram for IID estimation. To facilitate the calculations, the proposed algorithm is implemented using the Java programming language for automated calculation.

Results from the new approach are verified using video-captured ground-truth data at two locations. The percentage errors using the proposed approach are 1.4% and –5.6%, indicating that the new approach is able to produce fairly accurate IID estimates. In an online system application that computerizes the new approach, IIDs are estimated for 2523 incidents that occurred in the Puget Sound region freeways in the year 2009. Results demonstrate that the proposed approach is easy to use and the computational process is efficient. The proposed algorithm is also flexible in terms of input data aggregation level and time period for the IID analysis. The online system provides a platform for a variety of possible future applications.

When applying this approach to a large-scale network, it is preferable to calibrate the U/D relationships both spatially and temporally since the U/D relationships are location specific and time-period specific. Fortunately, model calibration using ridge regression requires easy efforts and can be automated computationally. Although there are no ramps between the U/D loop station locations for the two recorded incidents, the proposed algorithm can be also extended to the scenarios where ramps are presented. The essence of the short-term flow forecasting in our approach is to capture the relationship between U/D flows. Therefore, once the incident location and on/off ramp locations are given, the ramp volumes count for either upstream volume or downstream volume, and thus can be incorporated into the proposed approach.

FUNDING

The authors acknowledge the Washington State Department of Transportation and Transportation Northwest, University Center for Federal Region 10, for their funding support on this research. The authors also appreciate Timothy Thomson, a graduate research assistant at the Smart Transportation Applications and Research Laboratory (STAR Lab) of the University of Washington, for his help in video data collection and ground-truth travel delay data retrieval.

REFERENCES

Abdulhai, B., Porwal, H., & Recker, W. (1999). *Short term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks*. California PATH Working Paper UCB-ITS-PWP-99-1. Irvine, CA: Institute of Transportation Studies, Department of Civil and Environmental Engineering, University of California Irvine.

Bertini, R. L., & Lovell, D. J. (2009). Impacts of sensor spacing on accurate freeway travel time estimation for traveler information. *Journal of Intelligent Transportation Systems*, **13**(2), 97–110.

Cambridge Systematics, Inc. (2005). *Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation*. Research report for Federal Highway Administration. FHWA, U.S. Department of Transportation, Washington, DC.

Center for Urban Transportation Research. (2010). *Best practices for traffic incident management in Florida*, 2005. Retrieved from http://www.iacptechnology.org/IncidentManagement/BestPracticesFLDOT.pdf

Cheevarunothai, P., Wang, Y., & Nihan, N. L. (2006). Identification and correction of dual-loop sensitivity problems. *Transportation Research Record: Journal of the Transportation Research Board*, **1945**, 73–81.

Cheevarunothai, P., Zhang, G., & Wang, Y. (2010). *Quantifying freeway incident-induced delays using loop detector measurements*. Proceedings CD-ROM for the 89th Annual Meeting of Transportation Research Board, January 2010, Washington, DC.

Cheevarunothai, P., Zhang, G., & Wang, Y. (2012). Using precise time offset to improve freeway vehicle delay estimates. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, **16**(2), 82–93

Chung, Y. (2011). Quantification of nonrecurrent congestion delay caused by freeway accidents and analysis of causal factors. *Transportation Research Record: Journal of the Transportation Research Board*, **2229**, 8–18.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed., chap. 17). New York, NY: John Wiley & Sons.

Fries, R., Chowdhury, C., & Ma, Y. (2007). Accelerated incident detection and verification: A benefit to cost analysis of traffic cameras. *Journal of Intelligent Transportation Systems*, **11**(4), 191–203.

Fu, L., & Hellinga, B. (2002, May 12–15). *Real-time adaptive prediction of incident delay for advanced traffic management systems*. Proceedings of the 2002 Annual Conference of Canadian Institute of Transportation Engineers, Ottawa, Canada.

Fu, L., & Rilett, L. R. (1997). Real-time estimation of incident delay in dynamic and stochastic networks. *Transportation Research Record: Journal of the Transportation Research Board*, **1603**, 99–105.

Güner, A. R., Murat, A., & Chinnam, R. B. (2011). Dynamic routing under recurrent and non-recurrent congestion using ral-time ITS information. *Computers & Operations Research* **39**, 358–373.

Hobeika, A. G., & Kim, C. K. (1999). Traffic-flow-prediction systems based on upstream traffic. In *The 1994 Vehicle Navigation and Information Systems Conference*, Yokohama, Japan, 1994, pp. 345–350.

Kamga, C. N., Mouskos, K. C., & Paaswell, R. E. (2011). A methodology to estimate travel time using dynamic traffic assignment (DTA) under incident conditions. *Transportation Research Part C: Emerging Technologies*, **19**(6), 1215–1224.

Li, J., Lan, C. J., & Gu, X. (2006). Estimation on incident delay and its uncertainty on freeway networks. *Transportation Research Record: Journal of the Transportation Research Board*, **1959**, 37–45.

Lindley, J. A. (1987). A methodology for quantifying urban freeway congestion. *Transportation Research Record: Journal of the Transportation Research Board*, **1132**, 1–7.

Mak, C. L., & Fan, H. S. L. 2006. Heavy flow-based incident detection algorithm using information from two adjacent detector stations. *Journal of Intelligent Transportation Systems*, **10**(1), 23–31.

Mongeot, H., & Lesort, J. B. (2000). Analytical expressions of incident-induced flow dynamic perturbations using the macroscopic theory and an extension of the Lighthill and Whitham theory. *Transportation Research Record: Journal of the Transportation Research Board*, **1710**, 58–68.

Morales, M. J. (1987). Analytical procedures for estimating freeway congestion. *ITE Journal*, **57**(1), 45–49.

Nam, D. H., & Drew, D. R. (1998). Analyzing freeway traffic under congestion: Traffic dynamics approach. *ASCE Journal of Transportation Engineering*, **24**(3), 208–212.

Nam, D. H., & Drew, D. R. (1996). Traffic dynamics: Method for estimating freeway travel times in real time from flow measurements. *ASCE Journal of Transportation Engineering*, **122**(3), 185–191.

Qi, Y., Teng, H., & Martinelli, D. R. (2007). An inverstigation of incident frequency, duration and lanes blockage for determining traffic delay. *Journal of Advanced Transportation*, **43**,(3), 275–299.

Rakha, H., & Zhang, W. (2005). *Consistency of shock-wave and queuing theory procedures for analysis of roadway bottlenecks*. Proceedings CD-ROM for the 84th Annual Meeting of Transportation Research Board, Washington, DC.

Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications* (2nd ed.). New York, NY: Springer.

Skabardonis, A., Petty, K., Noeimi, H., Rydzewski, D., & Varaiya, P. P. (1996). I-880 field experiment: Data-base development and incident delay estimation procedures. *Transportation Research Record: Journal of the Transportation Research Board*, **1554**, 204–212.

Skabardonis, A., Varaiya, P., & Petty, K. (2003). Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record: Journal of the Transportation Research Board*, **1856**, 118–124.

Sullivan, E. C. (1997). New model for predicting freeway incidents and incident delays. *Journal of Transportation Engineering*, **123** (4), 267–275.

Transportation Research Board. (2000). *Highway capacity manual 2000*. Washington, DC: Transportation Research Board of the National Academies.

van Woensel, T., Kerbache, L., Peremans, H., & Vandaele, N. (2008). Vehicle routing with dynamic travel times: A queuing approach. *European Journal of Operational Research*, **186**(3), 990–1007.

Wang, Y., Corey, J., Lao, Y., & Wu, Y. J. (2009). *Development of a statewide online system for traffic data quality control and sharing*. Research Report for TransNOW and Washington State Department of Transportation. Retrieved from http://www.transnow.org/files/final-reports/TNW2009-12.pdf

Willmott, C. (1981). On the validation of models. *Physical Geography*, **2**, 183–194.

Wirasinghe, C. S. (1978). Determination of traffic delays from shock wave analysis. *Transportation Research*, **12**, 343–348.