Capacity, mobility, and safety

Fall 2021

Xiao Qin, Ph.D., P.E.
Professor, University of Wisconsin-Milwaukee

Introduction

- Capacity, mobility, and safety are the three most important transportation system performance measures.
  - **Capacity** is the maximum traffic flow rate in vehicles per hour for a highway facility considering the prevailing traffic state, roadway conditions, and driver population.
  - **Mobility** can be monitored by the average vehicle operating speed of the highway facility.
  - **Safety** is directly measured by crash frequency and injury severity.

When traffic is light, drivers are less affected by other road users, so they feel less pressured to change lanes or slow down. But in a denser traffic condition, a driver's actions such as unexpected changes in speed or positioning may force other drivers to take evasive actions to avoid a collision.

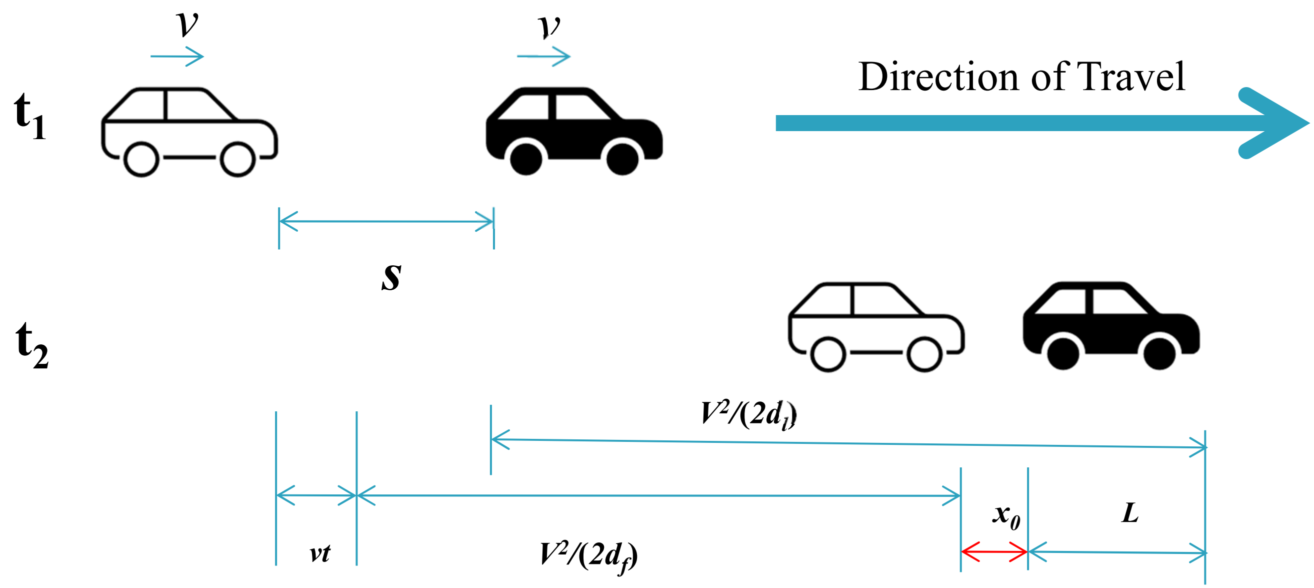
Introduction (cont'd)

- Capacity, mobility and safety are interconnected because of driver characteristics and behavior.
- Collecting, archiving, and processing traffic data in real-time presents opportunities for developing proactive safety management strategies.
- A real-time crash prediction (RTCP) model based on temporally and spatially proximal measurements (e.g., 100 m upstream within the most recent 5 min) can complement existing crash count models.

Learning Objectives

- Use car-following model to demonstrate the safety aspects of a classic driver behavior model.
- Develop relationships between crashes and traffic volume.
- Map crash typologies to a variety of patterns characterized by traffic variables.
- Apply Bayesian theory to predict crash probability.
- Use logistic regression to develop real-time crash prediction models given a real-time traffic input.
- Understand the motivation and develop RTCPM from simulated traffic data when actual traffic data are not available.

Modeling Space Between Cars

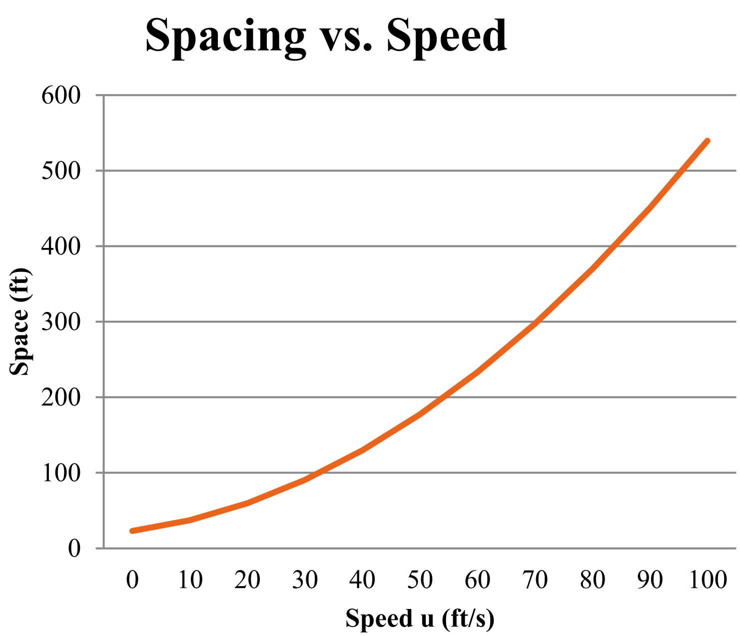


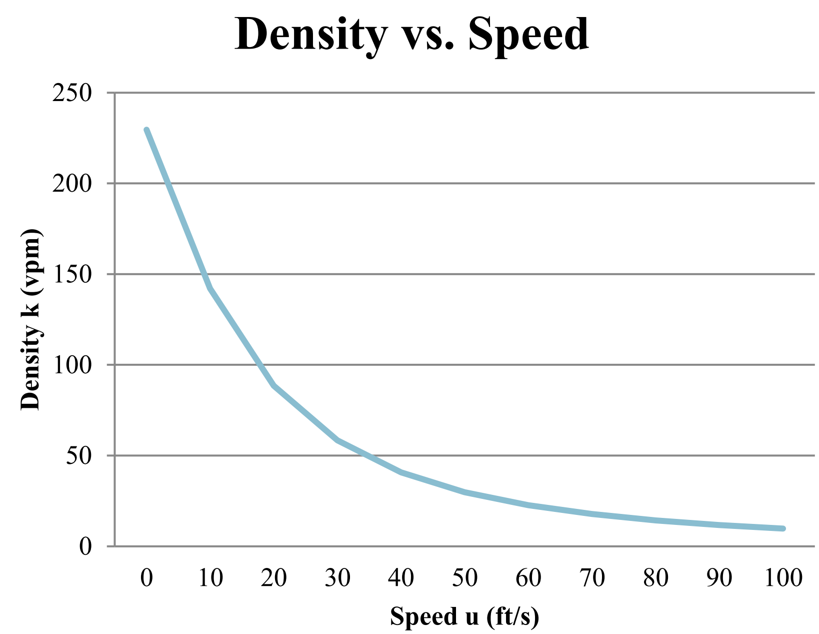
$$x_l = \frac{v^2}{2d_l} \tag{10.1a}$$

$$x_f = vt + \frac{v^2}{2d_f} \tag{10.1b}$$

$$s = x_f + L + x_0 - x_l = vt + \frac{v^2}{2d_f} + L + x_0 - \frac{v^2}{2d_l} \tag{10.1c}$$

Fundamental Diagrams (FD)





Assumption:

$L=20ft;\ x_0=3ft;\ \delta=1s;\ d_f=8ft/s^2;d_l=24ft/s^2$

6

Models for Flow-Speed-Density

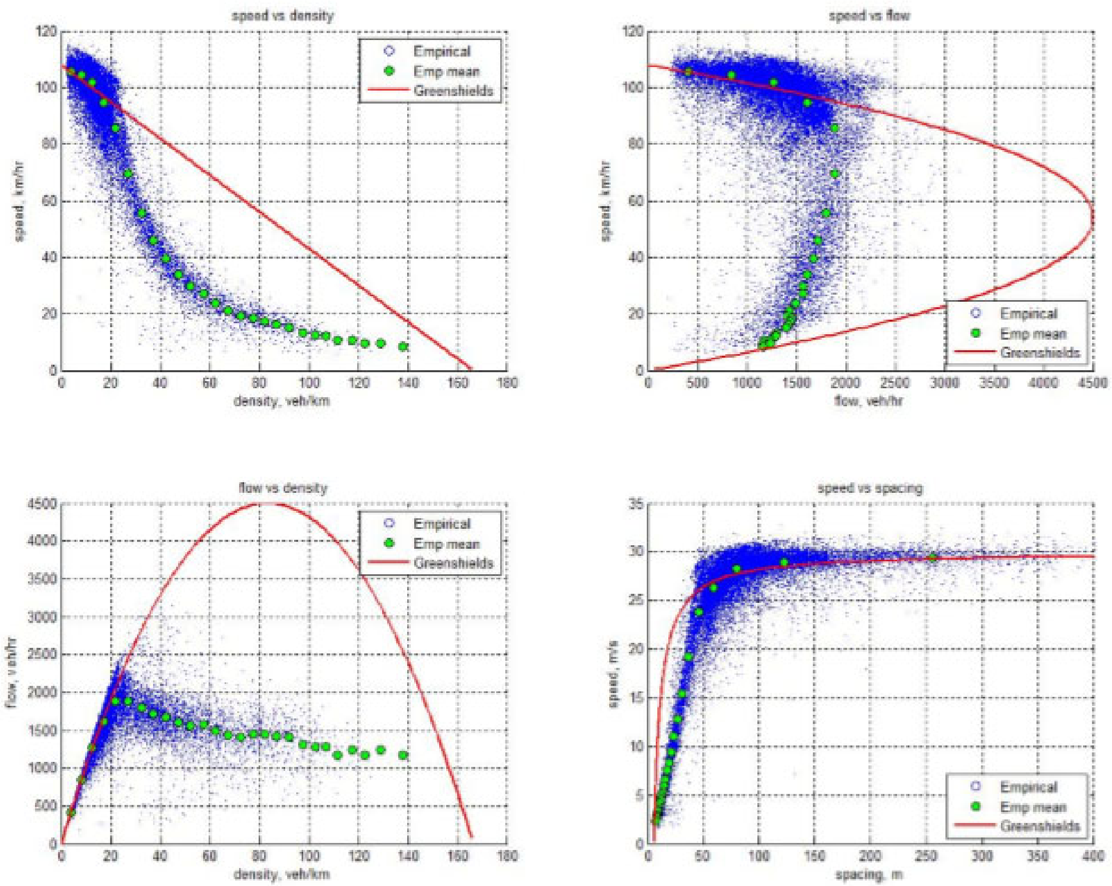


Figure 5.5: Fundamental implied by Greenshields model

$$\mu = \mu_f \left(1 - \frac{k}{k_j} \right)$$

Greenshields Model (1934)

Source: Traffic Flow Theory: Characteristics, Experimental Methods, and Numerical Techniques (1st Edition) by Daiheng Ni

Safety Considerations

Deceleration is relevant to safety: the safe level of operation occurs when the spacing between vehicles is such that the following car can safely stop by applying ***normal* deceleration**.

d_n is the normal or comfortable deceleration
d_e is the emergency deceleration
∞ is "instantaneous" or "stonewall" stop

Regime	Deceleration of lead vehicle	Deceleration of following vehicle
a	∞	d_n
b	d_e	d_n
c	∞	d_e
d	$d_l=d_f$	

8

Car-following Theory

Response=function (sensitivity, stimuli)

- Response is represented by the acceleration (deceleration) of the following vehicle.
- Stimuli is represented by the relative velocity of the lead and following cars.
- How to define sensitivity?

General Motors (GM) Models

GM 1

$$\ddot{x}_{n+1}(t+\Delta t) = \alpha(\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

GM 2

$$\ddot{x}_{n+1}(t+\Delta t) = \begin{cases} \alpha_1 \\ or \; (\dot{x}_n(t) - \dot{x}_{n+1}(t)) \\ \alpha_2 \end{cases}$$

GM 3

$$\ddot{x}_{n+1}(t+\Delta t) = \frac{\alpha_0}{x_n(t) - x_{n+1}(t)}(\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

GM 4

$$\ddot{x}_{n+1}(t+\Delta t) = \frac{\alpha(\dot{x}_{n+1}(t+\Delta t))}{x_n(t) - x_{n+1}(t)}(\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

GM 5

$$\ddot{x}_{n+1}(t+\Delta t) = \frac{\alpha_{l,m}(\dot{x}_{n+1}(t+\Delta t))^m}{(x_n(t) - x_{n+1}(t))^l}(\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

Connecting Micro- with Macro-scopic models

GM 3

$$\ddot{x}_{n+1}(t+\Delta t) = \frac{\alpha}{x_n(t) - x_{n+1}(t)} (\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

Greenberg Model

$$\Rightarrow \quad \boxed{\mu = \mu_o \ln\left(\frac{k_j}{k}\right)}$$

Traffic Stability

- Driver behavior: $C=\alpha(\Delta t)$
  - Unstable behavior: characterized by higher reaction times and higher sensitivity responses
    - Likely to induce unstable traffic conditions
  - Stable behavior: lower reaction times and lower sensitivity responses
    - Stable traffic conditions

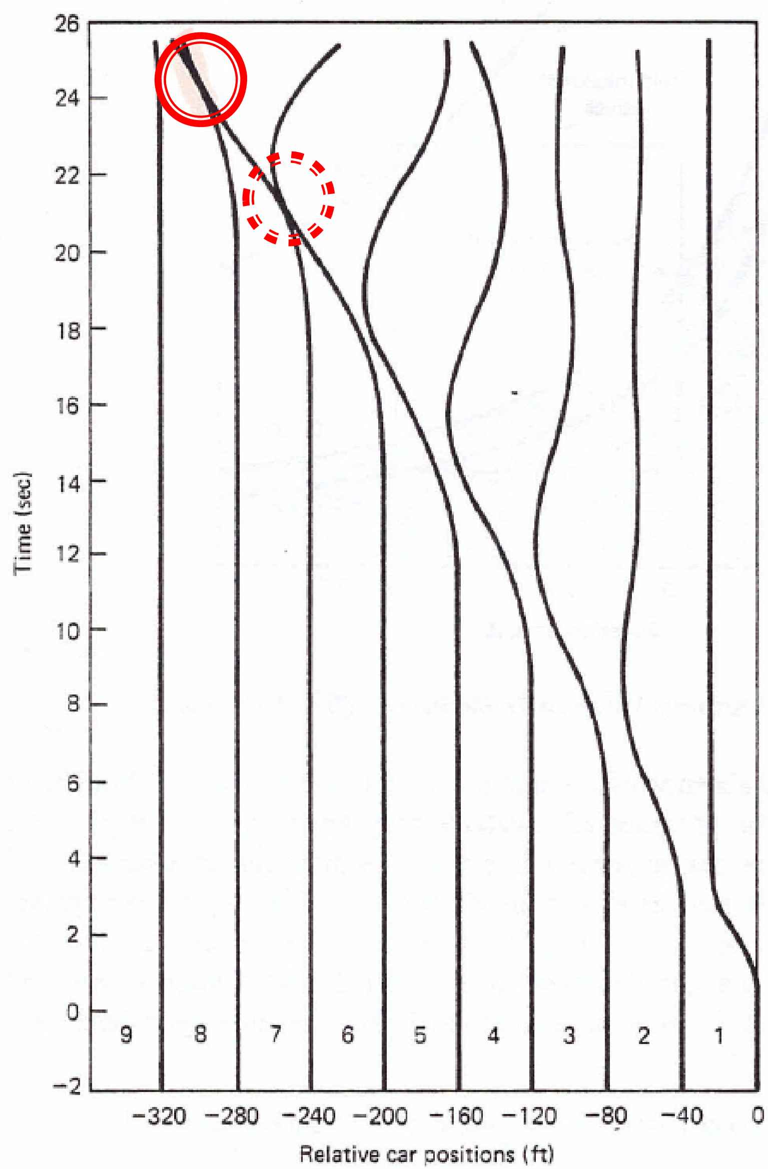
- Number of cars in the platoon
  - The driver in a single file of vehicles may behave in an erratic manner but not result in a collision.
  - But for other vehicles following in the line, it may not be the case.

Local Stability and Asymptotic Stability

Two types of traffic stability: by studying the change in car spacing as time goes by.

1. **Local stability**
   - Car-following behavior of just two cars
   - Local stability can be divided into three regions:
     - non-oscillatory: no oscillation will occur
     - damped oscillatory: the following vehicle oscillates but over time the oscillation decreases and the car following becomes stable.
     - increased oscillatory: oscillation will increase
2. **Asymptotic stability**
   - car-following behavior of a line of vehicles consisting of a lead vehicle and theoretically, an infinite number of following vehicles.

Example of Asymptotic Stability



"*the standard deviation of speed (thus, oscillations) is a significant variable, with an average odds ratio of about 1.08…. The likelihood of a (rear-end) crash increases by about 8% with an additional unit increase in the standard deviation of speed*".

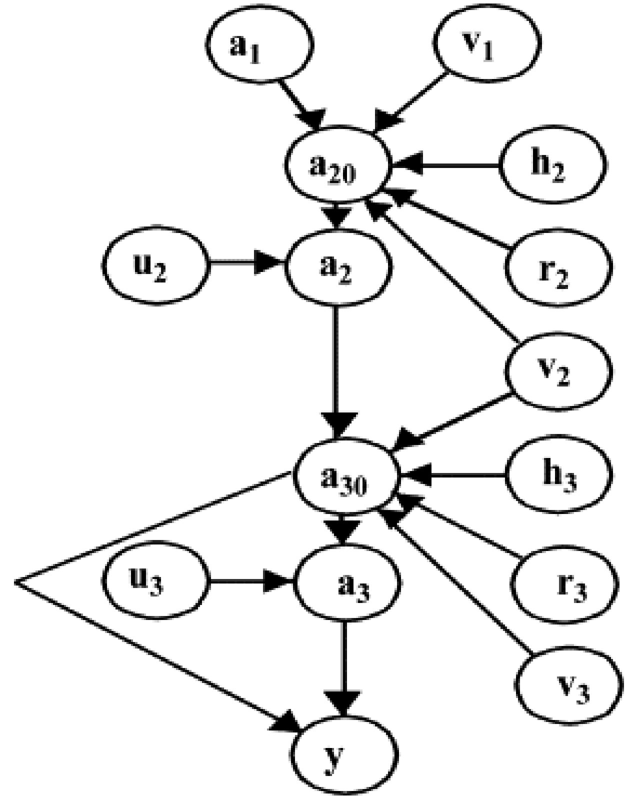
Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention*, *42*(2), 626-636.

Source: Traffic Flow Fundamentals by Adolf D. May

Figure 6.9 Example of Asymptotic Instability (From Reference 19)

A Causal Model of Crashes (Davis, et al.)

- The directed acyclic graph (DAG) displays a sequential braking model for a three-vehicle platoon.
- The nodes represent the model's variables, while the arrows indicate the presence and direction of causal dependencies.
- The variables a_{20} and a_{30} are the minimal decelerations needed, for vehicles 2 and 3, respectively.
- In the equations: a is a maximum achievable deceleration, and u_k accounts for the difference between the actual and the minimum deceleration. *y* is a collision indicator



$$a_k = \min(a_{k0} + u_k, a)$$

$$y = 0, \quad \text{if} a_{30} \leq a$$
$$y = 1 \quad \text{if} a_{30} > a$$

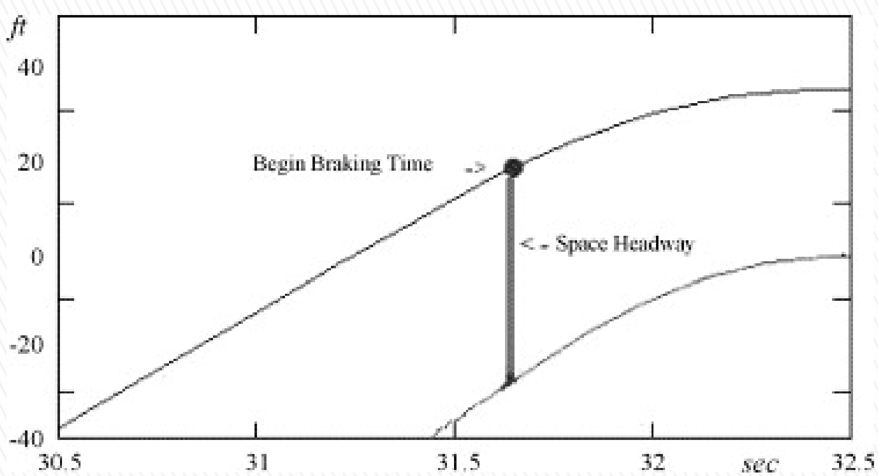
- Davis, G. A., & Swenson, T. (2006). Collective responsibility for freeway rear-ending accidents?: An application of probabilistic causal models. Accident analysis & prevention, 38(4), 728-736.
- Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. Accident Analysis & Prevention, 43(6), 1907-1919.

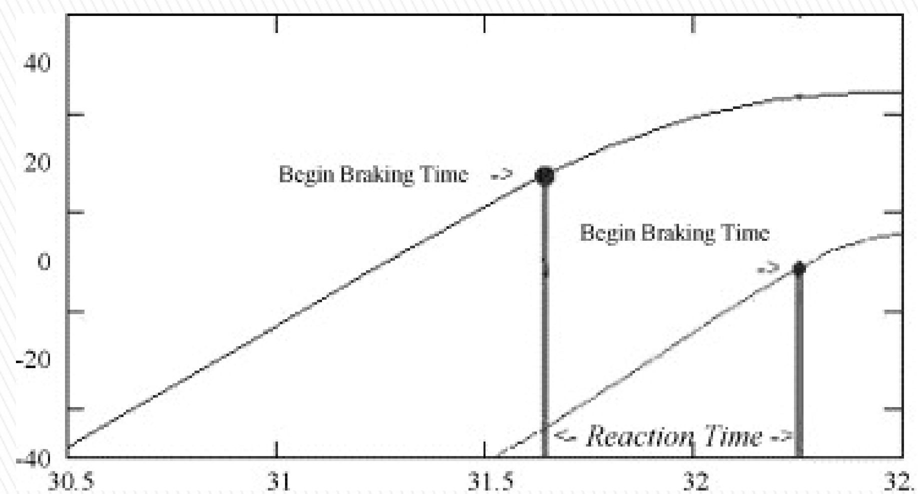
Causal Modeling and Estimation

- In practice, these values are uncertain. To allow for uncertainty, Pearl defines a probabilistic causal model as a causal model augmented with a probability distribution over the values taken on by the model's exogenous variables (e.g., individual speeds *v*, decelerations *a*, reaction times *r* and following headways *h*).
- To assess the possible causal contributions of the drivers in a platoon, first, need to determine values for the exogenous variables. This is done by assuming non-informative prior probability distributions for these parameters.
- Then treat the actual trajectory data as measurements from trajectory equation (next slide).
- Next, Bayes theorem can be used to compute posterior probability distributions for the trajectory model parameters.

Vehicle Trajectory

$$x_k(t) = \begin{cases} v_k t, & t \le t_{0k} \\ v_k t - 0.5 a_k (t - t_{0k})^2, & t_{0k} < t \le t_{0k} + \frac{v_k}{a_k} \\ v_k t_{0k} + \frac{v_k^2}{2a_k}, & t > t_{0k} + \frac{v_k}{a_k} \end{cases}$$

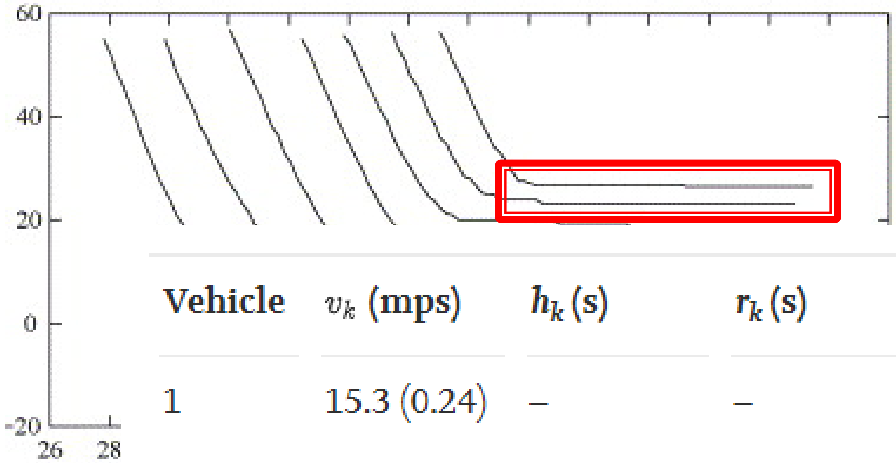




A space-time diagram: X-axis units are in seconds, Y-axis units are in feet.

Example computation of space headway from trajectories of leading and following vehicles

Example computation of reaction time from trajectories of leading and following vehicles



Trajectories of vehicles involved in accident on 30 December 2002, with collision between two right-most vehicles.

Vehicle	v_k (mps)	h_k (s)	r_k (s)	a_k (mps^2)	a_{k0} (mps^2)	t_{0k}	d_k (m)	$P[r_k > h_k]$
1	15.3 (0.24)	–	–	2.1 (0.03)	–	28.2 (0.1)	56.5 (2.0)	–
2	14.3 (0.10)	1.69 (0.02)	1.91 (0.14)	2.0 (0.02)	1.9 (0.02)	30.1 (0.1)	51.4 (0.8)	0.90
3	12.7 (0.13)	2.00 (0.02)	4.21 (0.16)	3.8 (0.27)	3.5 (0.20)	34.3 (0.2)	21.2 (1.8)	1.0
4	12.9 (0.08)	1.87 (0.03)	1.86 (0.17)	4.3 (0.16)	3.9 (0.13)	36.1 (0.1)	19.2 (0.9)	0.47
5	12.0 (0.07)	1.21 (0.02)	1.44 (0.10)	4.9 (0.28)	4.4 (0.20)	37.6 (0.1)	14.8 (0.9)	0.99
6	12.9 (0.19)	1.17 (0.03)	1.07 (0.14)	5.3 (.48)	5.2 (0.44)	38.7 (0.1)	15.9 (1.5)	0.19
7	12.7 (0.12)	1.24 (0.03)	1.65 (0.15)	6.2 (.33)	7.6 (0.56)	40.3 (0.1)	13.1 (0.9)	0.99

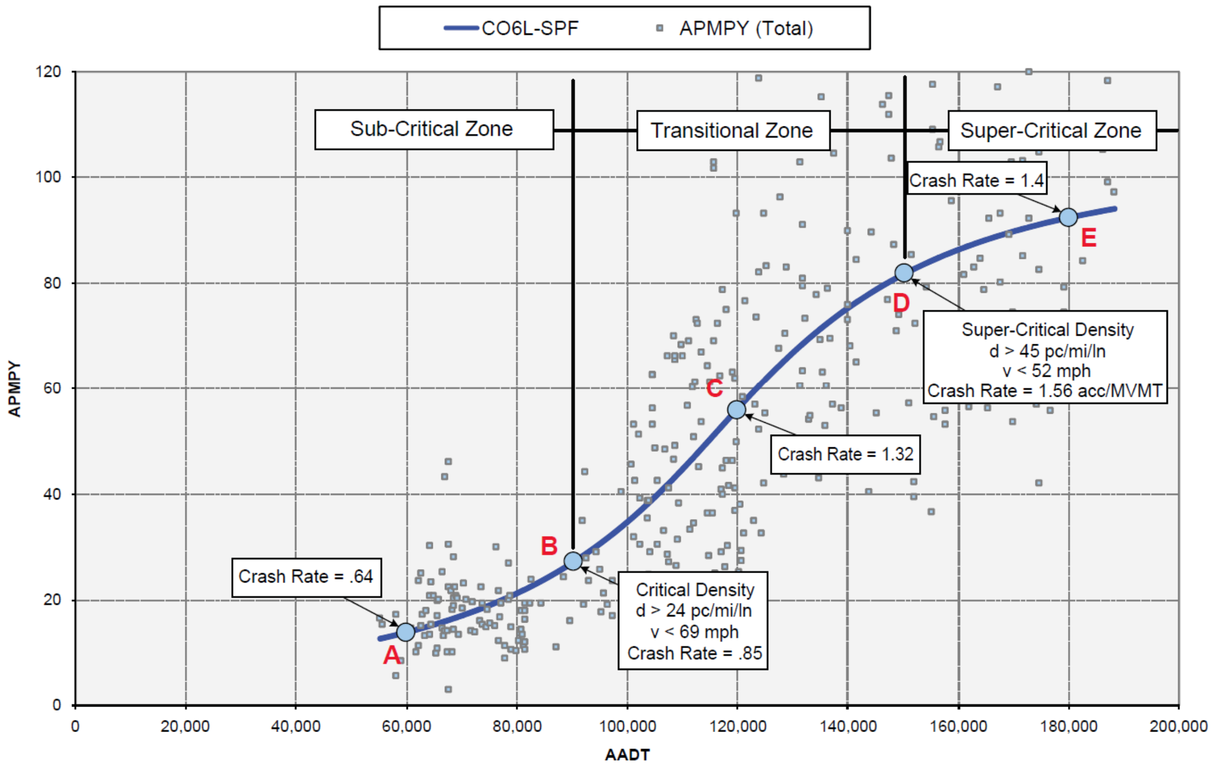
Driver 7, whose reaction time was approximately 0.4 s longer than his/her headway, the minimum deceleration jumped to about 7.6 mps^2, which exceeded the 6.2 mps^2 observed to have been used by driver 7, and a collision resulted.

Findings: A Probabilistic Causal Model of Traffic Conflicts & Crashes

- Findings are: (1) short following headways by the colliding drivers were probable causal factors for the collisions, (2) for each collision, at least one driver ahead of the colliding vehicles probably had a reaction time that was longer than his or her following headway, and (3) had that driver's reaction time been equal to his or her following headway, the rear-end collision probably would not have happened.
- The risk arising from a driver's headway choice has both an internal component, and an external component which falls primarily on other drivers.
- If not possible to increase headways, it may be possible to decrease reaction times.
- Applications: advanced driver assistance system; coordinated adaptive cruise control system.

Safety as a Function of Traffic Flow

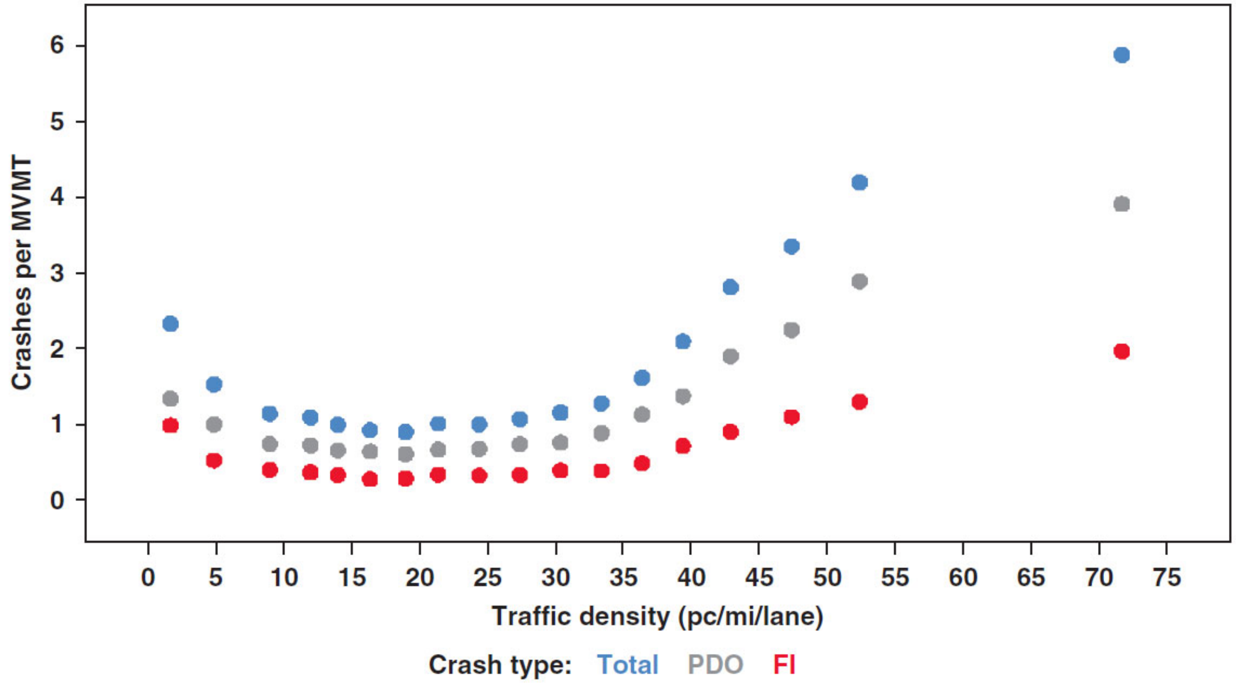
- One of the common approaches to exploring the relationship between crashes and corresponding traffic flow characteristics such as speed, flow rate, and density is to model crashes as a function of prevailing traffic parameters.
- The extensive examination of traffic flow characteristics has led to the concept of the level of safety service (LOSS). Kononov and Allery (2003) defined four LOSS levels as follows:
  - LOSS-I: low potential for crash reduction;
  - LOSS-II: better than expected safety performance;
  - LOSS-III: less than expected safety performance; and
  - LOSS-IV: high potential for crash reduction.



Freeway LOS (HCM):
A: d<11 pc/mi/ln
B: 11-18 pc/mi/ln
C: 18-26 pc/mi/ln
D: 26-35 pc/mi/ln
E: 35-45 pc/mi/ln
F: >45 pc/mi/ln

Kononov et al. (2011) identified six critical points along the SPF which is a non-linear relationship between accident per mile per year (APMPY) and AADT.

- A is crash rate ≤ 0.64 crashes/MVMT;
- B is crash rate between 0.64 and 0.85 crashes/MVMT;
- C is crash rate between 0.85 and 1.32 crashes/MVMT;
- D is crash rate between 1.32 and 1.56 crashes/MVMT; and
- E is crash rate ≥ 1.4 crashes/MVMT.



Crash rate versus density (Harwood et al. (2013)).

Total number of crashes:

$$Total\ per\ MVMT = 2.636 - 0.2143 \times D + 0.00708 \times D^2 - 4.80 \times 10^{-5} \times D^3 \tag{10.2}$$

Fatal and injury crashes:

$$FI\ per\ MVMT = 1.022 - 0.0842 \times D + 0.00264 \times D^2 - 1.79 \times 10^{-5} \times D^3 \tag{10.3}$$

Property damage only crashes:

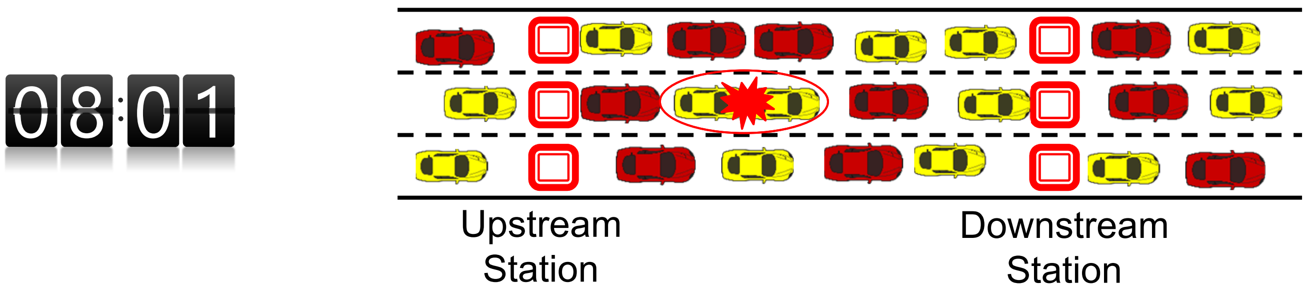
$$PDO\ per\ MVMT = 1.614 - 0.1301 \times D + 0.00444 \times D^2 - 3.01 \times 10^{-5} \times D^3 \tag{10.4}$$

Characterizing Crashes by Real-time Traffic

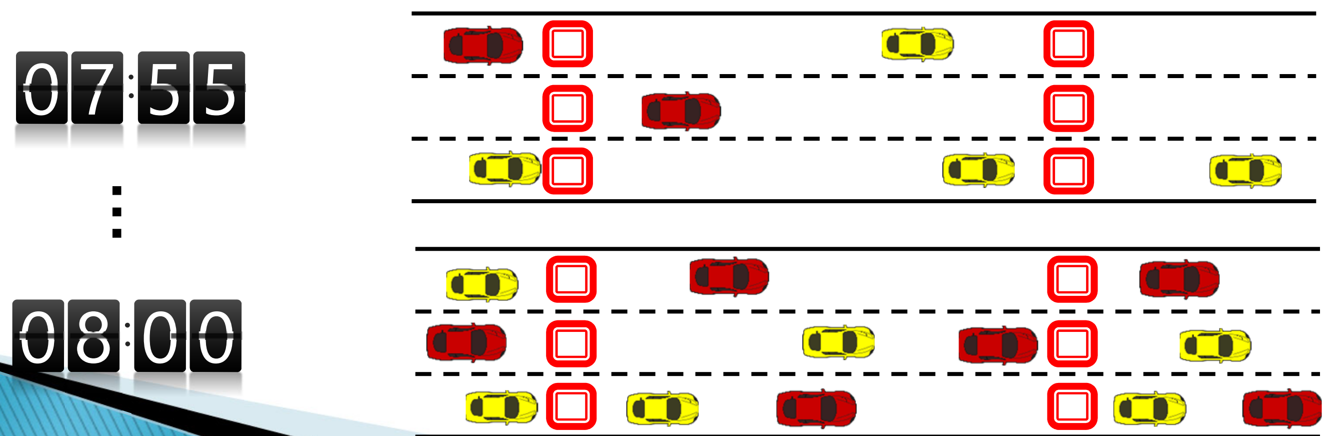
- Variations in traffic flow are strong indicators of the safety status of a traveling environment.
- Real-time crash studies use traffic data immediately before a crash to characterize the instantaneous traffic conditions under which crashes are more likely to occur.
- These studies are to detect disruptive traffic conditions and measure their impacts on safety.
- Each crash record has its date and time and location description. With a known location and time, crashes can be associated with traffic sensors located in their proximity, making it possible to link crash data with real-time traffic data. Crash-related traffic patterns are summarized by statistics such as the mean, median, variance, difference in percentile values, as well as the trends within extended intervals (i.e., 5 min).
- Monitoring two or more consecutive sensors in the same travel direction at the same time allows analysts to classify the traffic state as congested, free-flow, traffic bottleneck, or back-of-the-queue.

Loop Detector Data

- In order to know why a crash took place at 8:01am



- We need to know what happened between 7:55 and 8am.



Linking Traffic with Crashes

Inductive Loop Detector (ILD) Traffic Data

Region	Detector ID	Date	Time	Volume (VPL)	Speed (MPH)	Occupancy (%)
SE	5513	4/16/2014	9:00	20	67.17	5.67
SE	5513	4/16/2014	9:01	20	62.85	7.17
SE	5513	4/16/2014	9:02	14	65	4.33
SE	5513	4/16/2014	9:03	14	68.9	3.83
SE	5513	4/16/2014	9:04	15	67.67	5.83
SE	5513	4/16/2014	9:05	13	64.2	4.17

Crash Data

DOCTNMBR	ACCDDATE	ACCDTIME	REGION	COUNTY	CNTYCODE	MUNICIPALITY	ONHWY	ONHWYDIR	WTHRCOND	WISLR_LATDECDG	WISLR_LONDECDG
C311WBS	1/1/2012	20:40	SE	WAUKESHA	67	BROOKFIELD	94	E	SNOW	43.0289308	-88.1420277
DHLT79F	1/7/2012	13:12	SE	WAUKESHA	67	BROOKFIELD	94	E	CLR	43.0289302	-88.1416941
DHLT799	1/10/2012	17:31	SE	WAUKESHA	67	BROOKFIELD	94	E	CLDY	43.0244371	-88.1063855
C2ZKG06	1/17/2012	9:55	SE	WAUKESHA	67	BROOKFIELD	94	E	SNOW	43.0244062	-88.1010924
C2ZKG07	1/18/2012	8:20	SE	WAUKESHA	67	BROOKFIELD	94	E	CLR	43.0247588	-88.1132117
C2ZL670	1/20/2012	12:20	SE	WAUKESHA	67	BROOKFIELD	94	E	SNOW	43.0244062	-88.1010924
C3030L0	1/24/2012	17:55	SE	WAUKESHA	67	BROOKFIELD	94	E	CLR	43.0289276	-88.14036
C2ZVBQR	1/29/2012	7:23	SE	WAUKESHA	67	BROOKFIELD	94	E	SNOW	43.0259169	-88.087746
DHSWFX9	2/7/2012	10:55	SE	WAUKESHA	67	BROOKFIELD	94	E	CLDY	43.0289219	-88.1373583

Exploratory data analyses

- Exploratory data analyses are regularly performed as a preliminary step to screen traffic flow patterns that have a strong association with crashes.
- Changes in flow rate, speed, traffic density, or combinations of them can be measured by small time steps such as 1-min intervals.

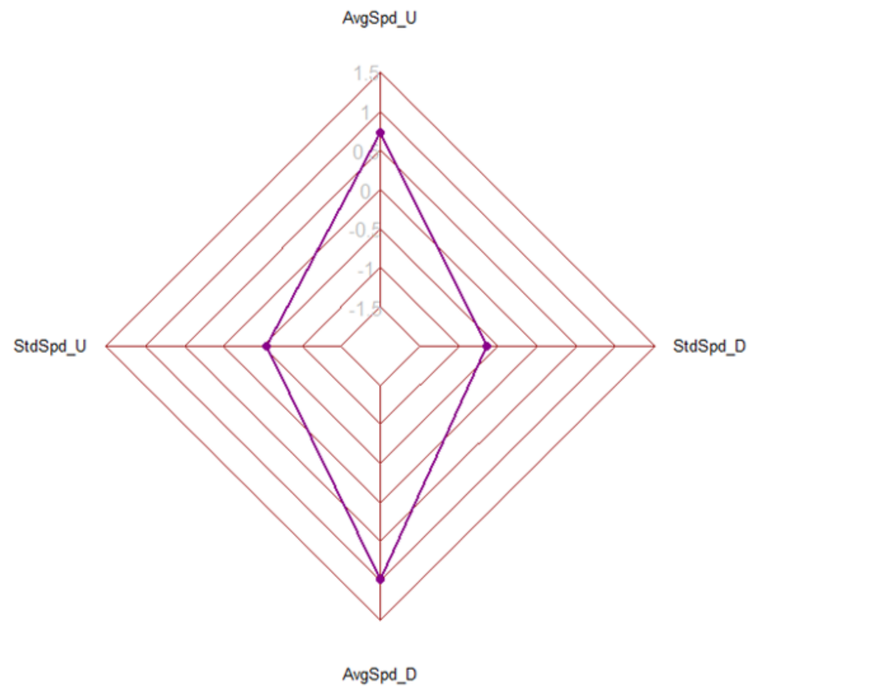
TABLE 10.2 Input traffic flow variables for statistical measures.

Performance measures	Variable
Central tendency	Mean volume/speed/occupancy
	Median volume/speed/occupancy
Variation	Standard deviation of volume/speed/occupancy
	Difference between 90th and 50th percentiles of volume/speed/occupancy
Temporal trend	Time series measures between data points for volume/speed/occupancy
Spatial consistency	Speed and density differentials measured at the same time.

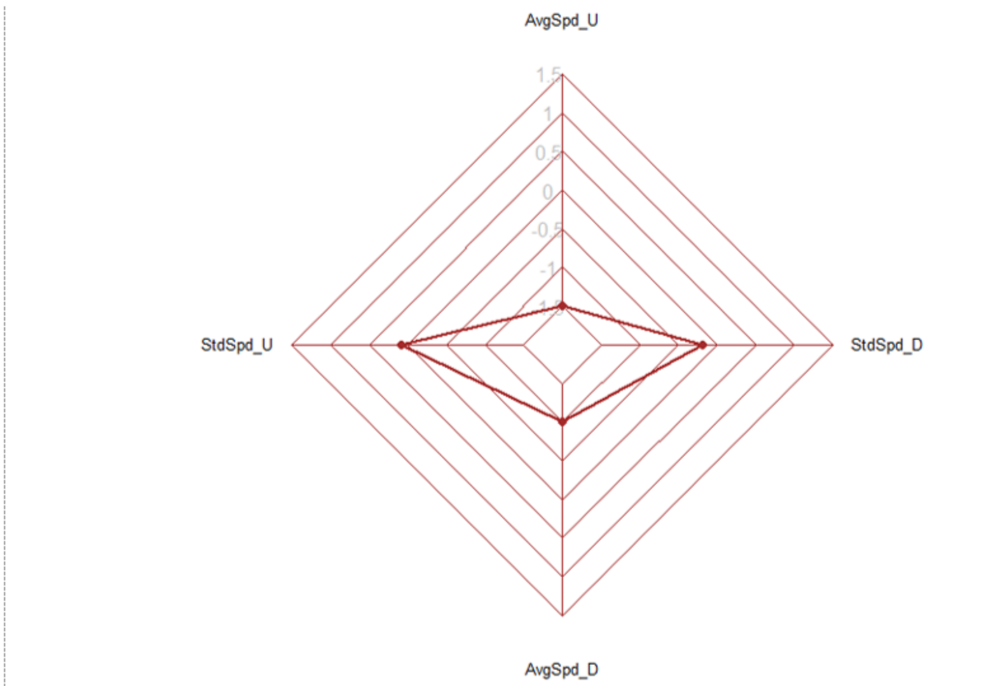
Exercise 12.1 Use Cluster Analysis to Characterize High Crash Risk Traffic Conditions

- Four speed-related parameters are used to characterize the travel safety level:
  - average 1-minute speed within 5-minute interval at upstream detector (AvgSpd_U);
  - average 1-minute speed within 5-minute interval at downstream detector (AvgSpd_D);
  - standard deviation of 1-minute speed within 5-minute interval at upstream detector (StdSpd_U); and
  - standard deviation of 1-minute speed within 5-minute interval at downstream detector (StdSpd_U).
- The K-means cluster analysis method can now be used to: a) create clusters of crash cases as the combinations of speed parameters, b) describe the clusters, c) compare speed characteristics from upstream and downstream of the crash location, and d) identify the traffic conditions with high crash risk.

Exercise 12.1: Characterizing Crashes by Real-time Traffic (cont'd)

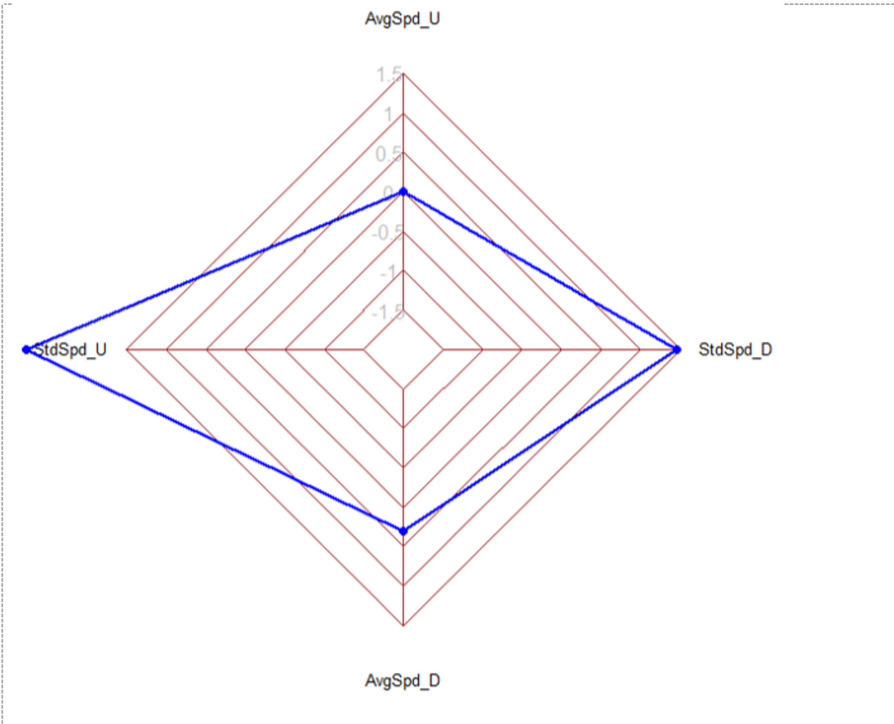


A. high speed and low speed variation at both upstream and downstream

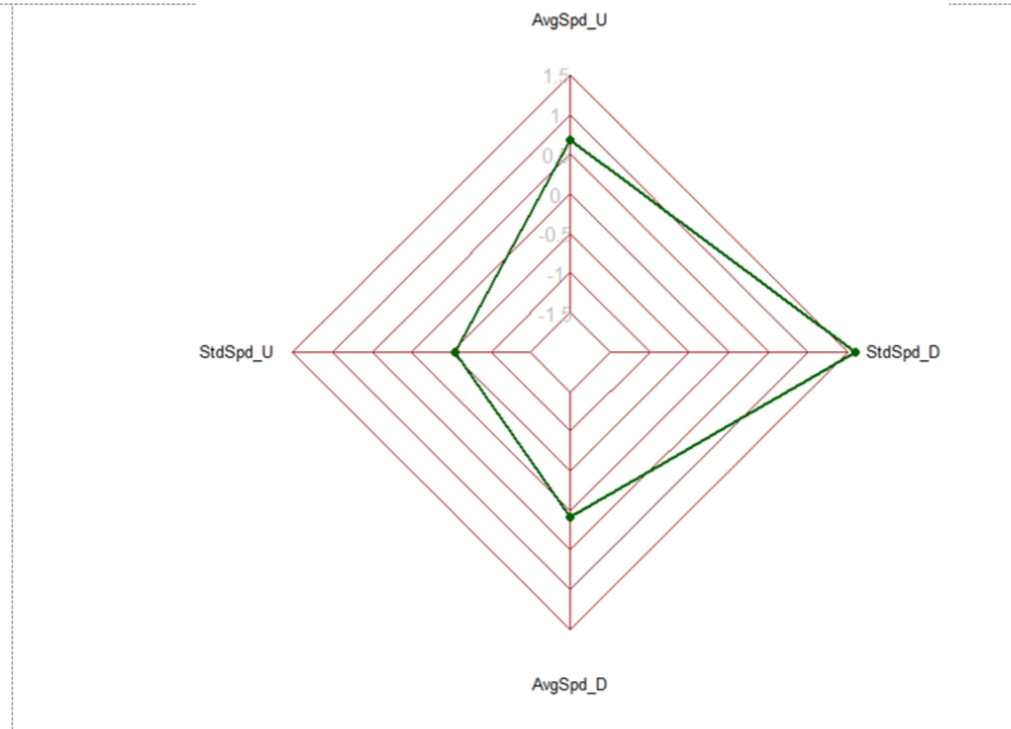


B. low speed and low speed variation at both upstream and downstream

Exercise 12.1: Characterizing Crashes by Real-time Traffic (cont'd)

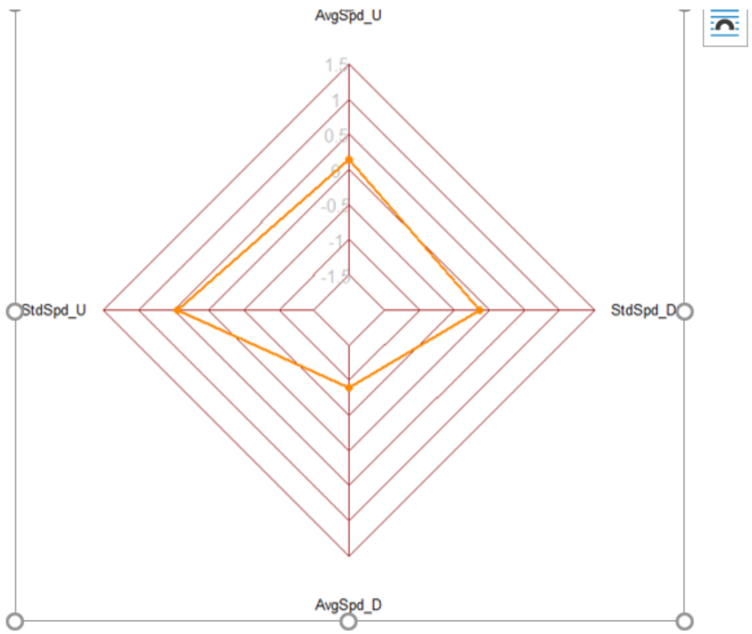


C. average speed and high speed variation at both upstream and downstream



D. average speed with low speed variation at upstream and average speed with high speed variation at downstream

Exercise 12.1: Characterizing Crashes by Real-time Traffic (cont'd)



E. average speed with moderately speed variation at upstream and low speed and speed variation at downstream

Predicting Immiment Crash Likelihood

Crash propensity can be calculated for these circumstances based on their appearance and duration in the traffic stream.

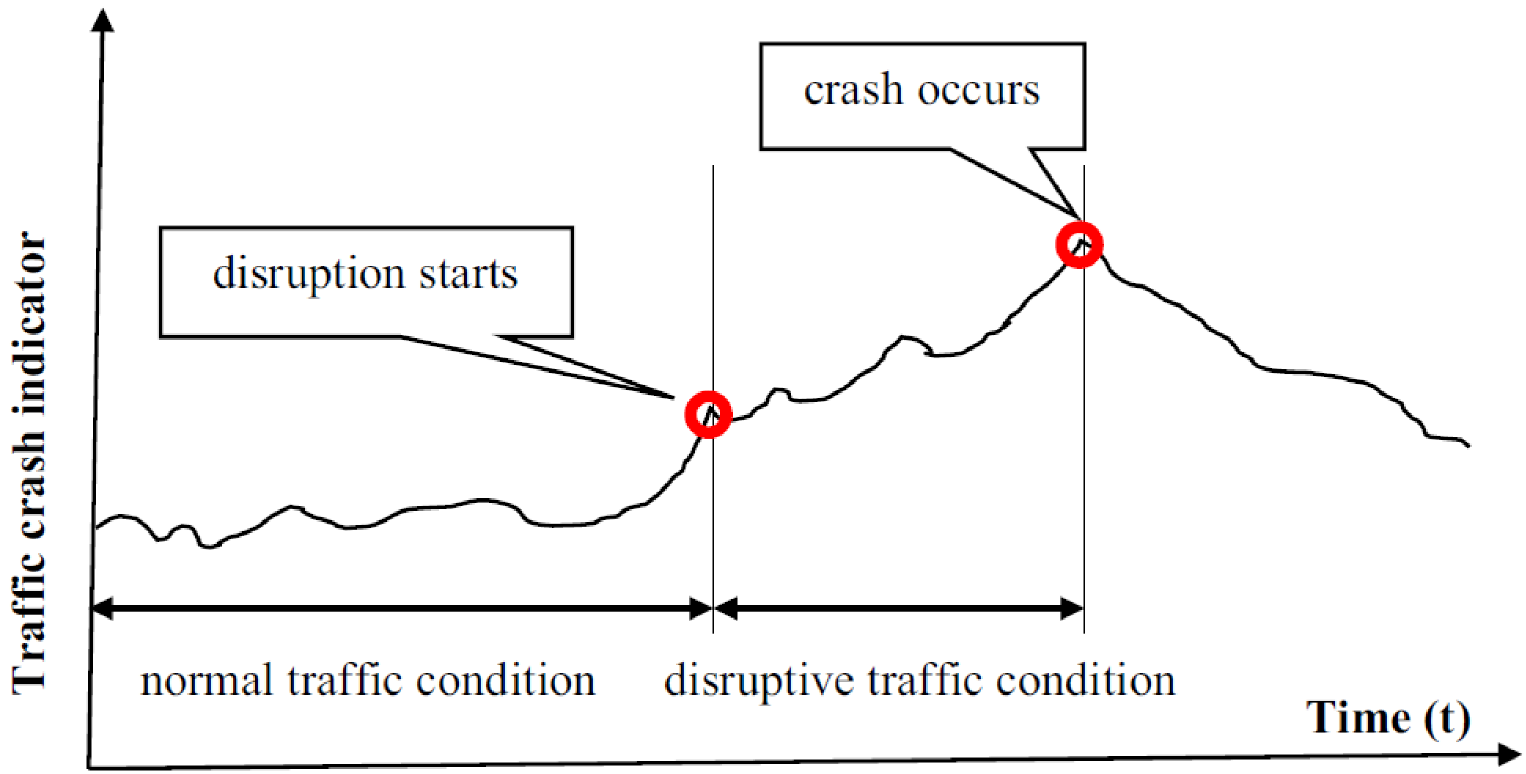


FIGURE 10.5 Traffic dynamics and crashes.

Bayesian Method: Predicting Imminent Crash Likelihood (cont'd)

- The Bayesian method can help predict future crash likelihood through the observation of current traffic state and the historical traffic profile.
- If we assume event 1 to be a crash event C and event 2 to be an observable traffic event X (e.g., speed variation within 5 minutes), the conditional probability of C given the speed variation measurement X can be formulated

$$P(C|X) = \frac{P(X|C)P(C)}{P(X|C)P(C) + P(X|N)P(N)} = \frac{P(C) \times f_{crash}(X)}{P(C) \times f_{crash}(X) + P(N) \times f_{non-crash}(X)} \quad (10.6)$$

where P(C|X) is the posterior probability that the given traffic measurement would lead to a crash occurrence; P(C) is the prior probability that given traffic measurement belongs to crash cases; P(N) = 1-P(C) is the prior probability that given traffic measurement belongs to non-crash cases; f_{crash} (X) is the probability density function of the traffic measurement leading to a crash; and, $f_{non\text{-}crash}$ (X) is the probability density function of the traffic measurement not related to any crash.

Predicting Imminent Crash Likelihood (cont'd)

- Normal traffic conditions are defined as a 5-minute period 30 minutes before a crash, and disruptive traffic conditions are the 5-minute period preceding a crash. Flow, density and speed data were collected from upstream detector stations.
- The probability density function of standard deviation of speed can be constructed for both normal and disruptive traffic conditions, respectively using nonparametric kernel smoothing.

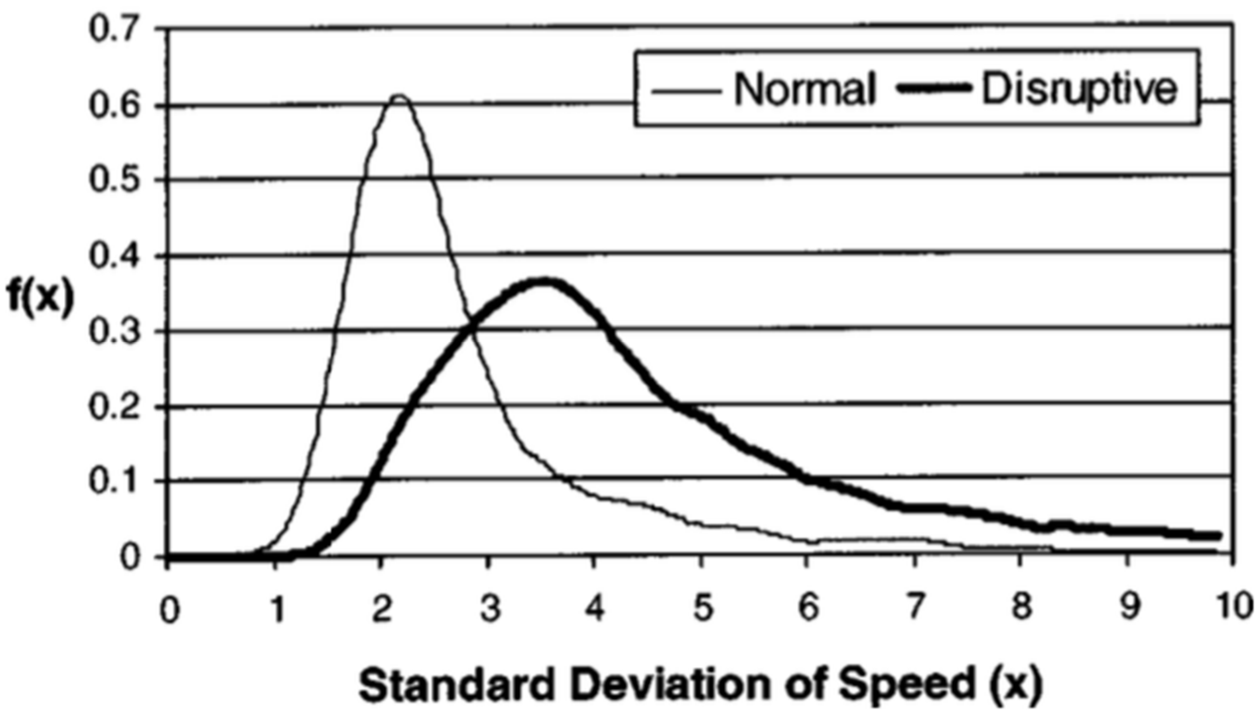


Figure: Probability density function curves of standard deviation of speed. (Oh et al., 2005)

Predicting Imminent Crash Likelihood (cont'd)

- The prior probability (P(C)) that given speed variation belongs to crash cases can be approximated as: $P(C) = \frac{\text{number of 5−min intervals of crash cases}}{\text{total number of 5−min intervals}}$.
- Probability of accident occurrence $P(C|X)$ calculated with Equation 10.6 within the next 5-minutes with respect to the standard deviation of speed implies that a higher standard deviation leads to a higher probability of crashes.

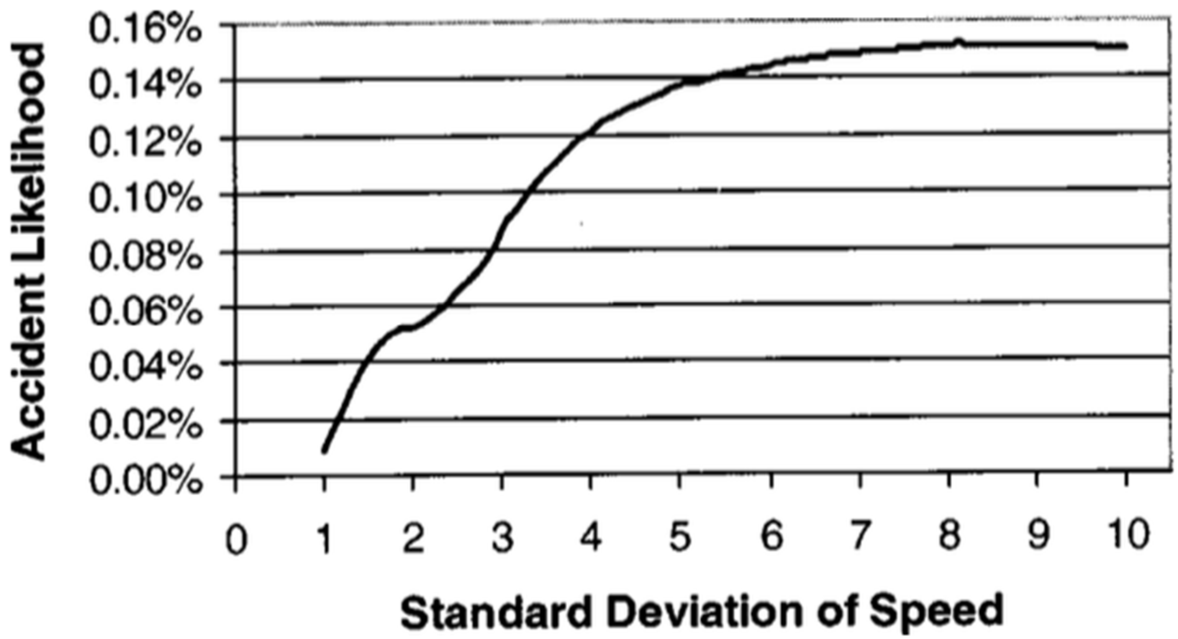


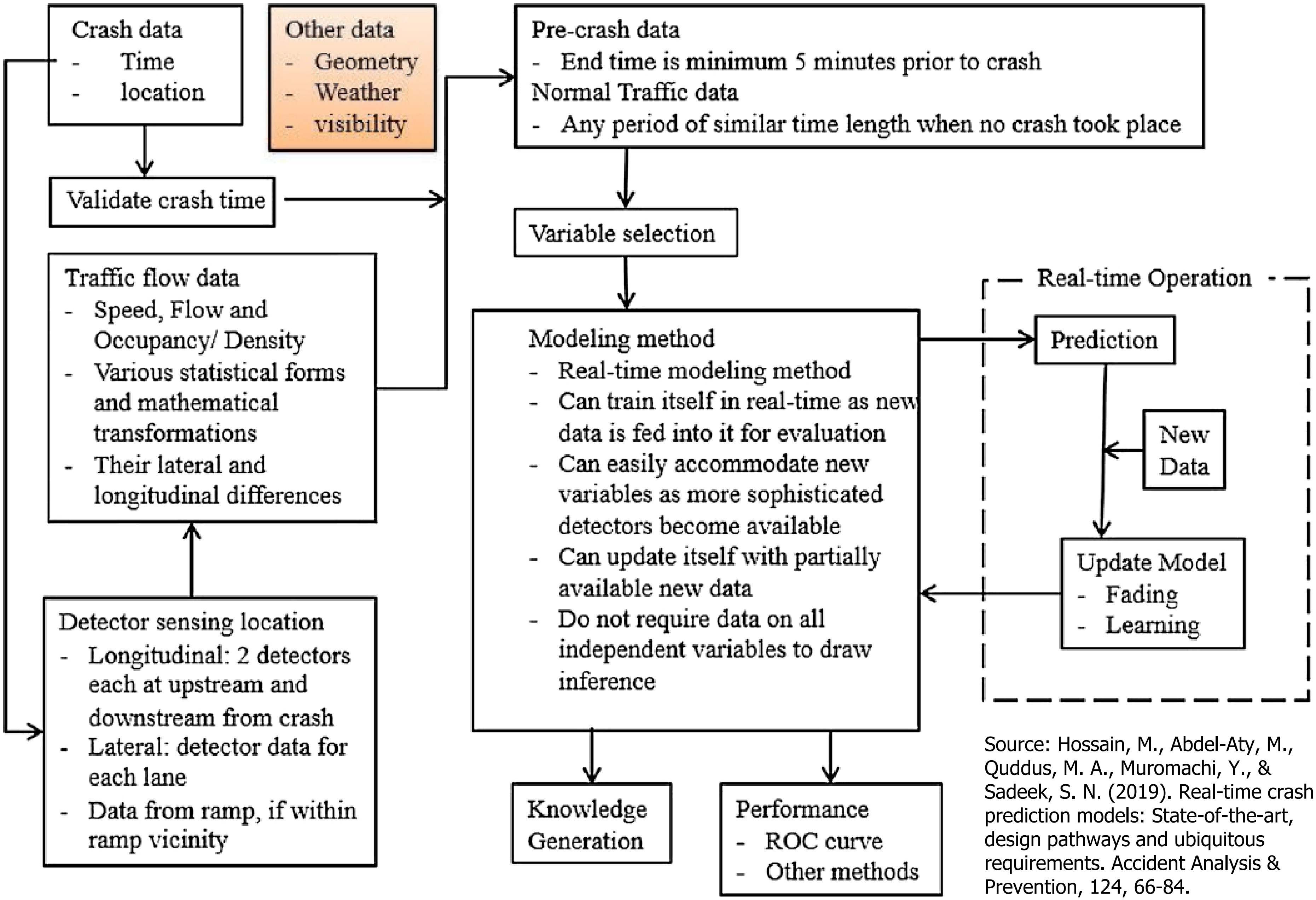
Figure: Estimated probability of a crash given standard deviation of speed (Oh et al., 2005).

Real-time Predictive Analysis of Crashes

Real-time Crash Analysis and Prediction

- The use of statistical models to construct meaningful relationships between historical crash data and traffic flow variables, quantify their impact on crash occurrence, and predict future crash occurrence when new traffic observations are available, so called the real-time crash prediction model (RTCPM).
- The underpinning assumption of real-time predictive crash analysis is that a certain combination of traffic conditions is likely to lead to a crash, aka "crash precursors."

Statistical approach	Statistical approach	AI/Data mining
a. logistic regression i. matched case control, ii. simple, iii. conditional, iv. sequential, v. Bayesian conditional parameter, vi. Bayesian random parameter, vii. Bayesian, viii. Multivariate, ix. Bayesian matched case-control, x. Multilevel, xi. Multilevel Bayesian, xii. Random parameter, xiii. Mixed, xiv. ordinal b. Aggregated log linear model c. Multivariate Probit d. Bayesian classifier e. Generalized estimating equations (GEE) f. Non-linear Canonical Correlation Analysis g. Bayesian Statistics h. Seemingly unrelated negative binomial	h. Seemingly unrelated negative binomial i. Poisson, Negative binomial, Zero-hurdle Poisson, Zero hurdle negative binomial j. Bayesian Structural Equation Modelling k. Binary response logit model l. NRBF, m. Binary Logit, n. Bayesian Bivariate Poisson-lognormal model o. UFC p. Bayesian Hierarchical Poisson Model q. Poisson log-normal Model r. Multinomial Logit Model s. Random Parameter Negative Binomial	a. Neural network i. Simple, ii. Probabilistic, iii. Bayesian, iv. Deep, v. Others b. Bayesian Network i. Static, ii. Dynamic c. Classification trees i. CART, ii. SVM, iii. Rule based classifier, iv. RVM d. Genetic algorithm e. Stochastic Gradient Boosting f. k-NN g. PCA Source: Hossain, M., Abdel-Aty, M., Quddus, M. A., Muromachi, Y., & Sadeek, S. N. (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. Accident Analysis & Prevention, 124, 66-84.



Source: Hossain, M., Abdel-Aty, M., Quddus, M. A., Muromachi, Y., & Sadeek, S. N. (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. Accident Analysis & Prevention, 124, 66-84.

Optional

General Framework of RTCPM

Logistic Regression Models

- Study Design
  - Unmatched studies
  - Matched studies
- Regression Model
  - Binary logistic regression model
  - Conditional logistic regression model

Binary Logistic Regression Model

The binary logistic regression model is a good choice for a dichotomous (binary) dependent variable such as crash versus non-crash.

$$\text{logit}\, p_n = \log\left(\frac{p_n}{1-p_n}\right) = \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \ldots + \beta_k x_{n,k} \qquad (10.7)$$

where p_n represents the crash probability given $\mathbf{x}_n$ $(n = 1, N)$; $\mathbf{x}_n = (x_{n,1}, x_{n,2}, \ldots, x_{n,k})$ is a set of k explanatory variables for sample n; βs are the coefficients for $(x_{n,1}, x_{n,2}, \ldots, x_{n,k})$; k is the number of parameters; and N is the number of cases.

Eq. (10.8) shows how the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ can be estimated by maximizing the log-likelihood function:

$$lnL(\boldsymbol{\beta}, \boldsymbol{x_n}) = \sum_{n=1}^{N}\left[\left(\beta_0 + \beta_1 x_{n,1} + \ldots + \beta_k x_{n,k}\right) - \ln(1 + e^{-(\beta_0 + \beta_1 x_{n,1} + \ldots + \beta_k x_{n,k})}\right] \qquad (10.8)$$

Exercise 10.2: a binary logit model to estimate the safety impact of traffic

TABLE 10.3 Modeling results.

Variable	Estimate	Std. error	*P* value	Odds ratio
Intercept	−1.159	0.533	0.030	—
Average volume	0.000179	0.000090	0.047	1.000
std. dev. volume	0.001182	0.000497	0.017	1.001
Average density	−0.058750	0.006554	<0.001	0.943
std. dev. density	0.066090	0.018720	<0.001	1.068

- The estimated coefficients are presented as the log odds ratio between the probability of a crash case and a non-crash case given the change of an independent variable.
- The odds ratio can be estimated by taking the exponential of the estimated coefficient.
- With other variables being held constant, if average traffic density increases by one unit, the odds ratio of a crash is 0.943 (or $e^{-0.05875}$). In other words, the predicted odds of a crash are 5.7% lower than they are before the one-unit increase.

Example: Analysis of Crash Precursors (Lee, C., et al., 2002)

- Explore factors contributing to changes in crash rate for individual vehicles on an urban freeway.
- Develop a probabilistic model relating significant crash precursors to changes in crash risk.
- References:

  - Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Transport. Res. Rec. J. Transp. Res. Board 1784, 1–8.
  - Lee, C., Hellinga, B., Saccomanno, F., 2003. Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. Transport. Res. Rec. J. Transp. Res. Board 1840, 67–77.

Variables Used in the Model (Lee, C., et al., 2002)

- Crash precursors represent the traffic flow conditions prior to a crash.
- Two types of crash precursors are often considered: (a) variation of speed and (b) traffic density.
  - Variation of speed is measured by the coefficient of variation of speed (CVS) computed over an observation period with fixed duration.
  - Two types of CVS are estimated: a) CVS1, measures the average speed variation experienced by drivers on each lane as they travel from upstream to downstream; and b) the average variation of speed difference across lanes

Crash Prediction Model (Lee, C., et al., 2002)

- Crash prediction model was developed using a log-linear analysis because it represents the relationship between crash precursors and crash frequency adjusted by traffic exposure. The distribution of crash frequency is assumed to follow the Poisson distribution.
- Alternatively, a binary logit model can be used for crash prediction. However, there are two problems associated with the use of a logit model in real-time crash prediction.
    - the overrepresentation of no-crash data.
    - the difficulty with determining exposures for continuous traffic flow measures.

$$\ln(F) = \theta + \lambda_{\mathrm{CVS}_1(i)} + \lambda_{\mathrm{CVS}_2(j)} + \lambda_{D(k)} + \lambda_{P(l)} + \lambda_{R(m)} + \lambda_{W(n)} + \beta \cdot \ln(\mathrm{EXP}) \tag{4}$$

where

F = expected number of crashes over the analysis time frame,

θ = constant,

$\lambda_{\mathrm{CVS}_1(i)}$ = effect of the coefficient of variation of speed along the road section (CVS_1) having i levels,

$\lambda_{\mathrm{CVS}_2(j)}$ = effect of the coefficient of variation of speed difference across lanes (CVS_2) having j levels,

$\lambda_{D(k)}$ = effect of density having k levels,

$\lambda_{P(l)}$ = effect of time of day having l levels,

$\lambda_{R(m)}$ = effect of road geometry having m levels,

$\lambda_{W(n)}$ = effect of weather having n levels,

β = parameter for exposure, and

EXP = exposure in vehicle-kilometers of travel.

Findings (Lee, C., et al., 2002)

TABLE 2 Estimated Parameters of Log-Linear Model

Parameters	Estimate	Standard Error*	Z-Value**
Constant (θ)	1.3386	.1985	6.74
$\lambda_{CVS_1=1} (CVS_1 \le 0.075)$	-1.4307	.1903	-7.38
$\lambda_{CVS_1=2}$ $(0.075 < CVS_1 \le 0.1)$	-1.2750	.1826	-6.98
$\lambda_{CVS_1=3}$ $(CVS_1 > 0.1)^{***}$	0		
$\lambda_{CVS_2=1} (CVS_2 \le 0.35)$	-1.8417	.2300	-8.01
$\lambda_{CVS_2=2}$ $(0.35 < CVS_2 \le 0.5)$	-.8271	.1508	-5.48
$\lambda_{CVS_2=3}$ $(CVS_2 > 0.5)^{***}$	0		
$\lambda_{D=1}$ $(D \le 15)$	-1.3911	.2097	-6.63
$\lambda_{D=2}$ $(15 < D \le 25)$	-.4289	.1498	-2.86
$\lambda_{D=3}$ $(D > 25)^{***}$	0		
$\lambda_{P=0}$ $(P = 0)$	-.6049	.1437	-4.21
$\lambda_{P=1}$ $(P = 1)^{***}$	0		
$\lambda_{W=0}$ $(W = 0)$	.9648	.1765	5.47
$\lambda_{W=1}$ $(W = 1)^{***}$	0		
$\lambda_{R=0}$ $(R = 0)$	-.5271	.1356	-3.89
$\lambda_{R=1}$ $(R = 1)^{***}$	0		
β (Exposure)	.2870	.1869	8.26

* A measure of the dispersion of the coefficient.
** A standardized measure of the parameter coefficient.
*** Aliased cells.

Crashes are more likely to occur

1. as the variation of speed along the section and the variation of speed difference across lanes increase.
2. as the density increases
3. during the peak period than off-peak period
4. in the road sections with more frequent lane changes (i.e., merging and diverging sections) than in the straight sections.
5. in normal weather conditions
6. as the exposure increases

Conditional Logistic Regression Model

- A matched case-control study focuses on variables of interest by controlling for nuisance factors.
- A "case" is defined as the representative traffic conditions occurring right before a crash, while "control" represents the non-crash traffic conditions. Each crash case involves several non-crash events that are selected as controllers such that the nontraffic-flow variables (e.g., location, time, season) of non-crash cases match the corresponding variables of crash cases.
- Each case and its controllers constitute a stratum. The controlled, nontraffic-flow variables are the same within each stratum, but are different across strata.

$$\mathrm{logit}(p_{nj}) = \propto_i + \sum_{k=1}^{K} \beta_k x_{njk} \tag{10.9}$$

where α is a stratum-specific parameter and is treated as nuisance parameter (not estimable); x_{njk} is the kth traffic flow variable for the case (j = 0) or the jth control in the nth stratum; n=1,2, .., N; j=0, 1, …, J; and k=1, 2, …, K. N is the number of strata, J denotes the number of controls, and K represents the number of explanatory variables.

Parameter Estimation

$$l_n(\beta) = \frac{\exp(\sum_{k=1}^{K} \beta_k x_{n0k})}{\sum_{j=0}^{J} \exp(\sum_{k=1}^{K} \beta_k x_{njk})} \quad (10.10)$$

$$l(\beta) = \prod_{n=1}^{N} l_n(\beta) \quad (10.11)$$

- If stratum-specific parameters α s are treated as nuisance parameters, a conditional likelihood of β could be created. The maximum likelihood estimators are expressed in Eq. (10.10) (Hosmer and Lemeshow, 2004).
- The full conditional likelihood is the product over N strata
- The full conditional likelihood is independent of stratum-specific parameters αs and thus, αs cannot be estimated. <u>So, the purpose of the conditional logistic model is not to calculate the crash probability of a specific case, but to estimate the effects of variables of interest on a crash case through the slope coefficients βs.</u>

Example 10.1: A conditional logistic model to estimate the safety impact of traffic (Aty et al., 2004)

TABLE 10.4 Model results for real-time freeway crash prediction[a].

Variable		AO		Log (CVS)	
Station	Time slice	$P_r > \chi^2$	Hazard ratio	$P_r > \chi^2$	Hazard ratio
A	1	0.008	1.021	0.007	1.862
A	2	0.008	1.021	0.024	1.682
A	3	0.005	1.022	0.015	1.740
A	4	0.039	1.016	0.017	1.728
A	5	0.083	1.014	0.069	1.520
A	6	0.034	1.017	0.186	1.355

[a]The 30-min time period before a crash is divided into six 5-min intervals called Time Slices, with Slice 1 being the 5 min before the crash.

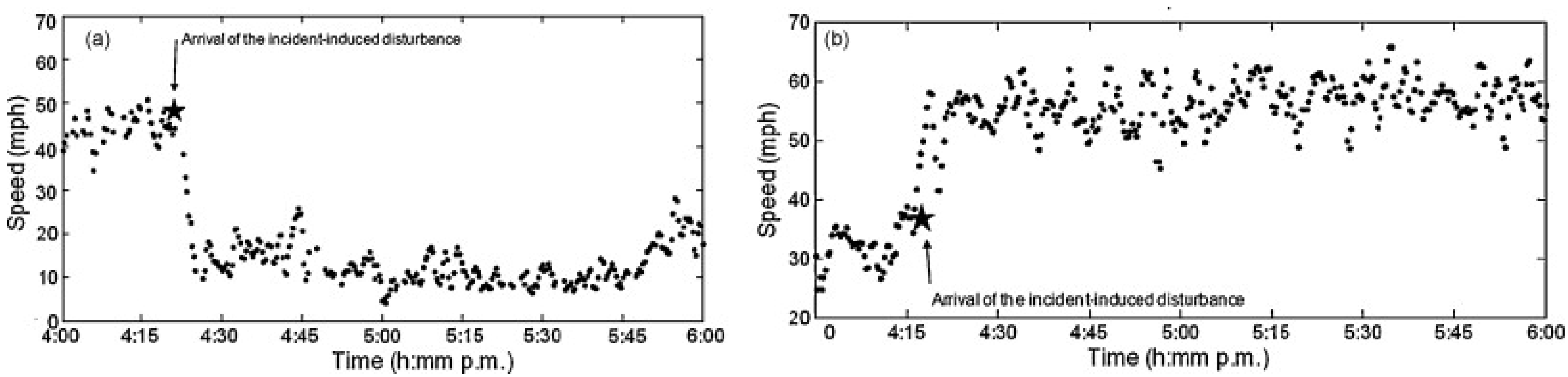
The hazard ratio (i.e., exp(β)), as called odds ratio, is an estimate of the expected change in the risk ratio of having a crash versus non-crash per unit change in the corresponding factor. A value of greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the factor increase, the odds of having a crash increase. For example, a hazard ratio of 2.5 corresponding to CVS means the risk for a crash increases 2.5 times for each unit increase in CVS.

Example: Impact of traffic oscillations on freeway crash occurrences (Zheng et al., 2010)

- Traffic oscillations are characterized by recurring decelerations followed by accelerations (stop-and-go) during congestion.
- A matched case-control study using traffic and crash data from a freeway segment. Traffic conditions prior to each crash were taken as cases, while traffic conditions during the same periods on days without crashes were taken as controls.
- Conditional logistic regression models.
- Number of samples (82 crashes) were obtained and the control-to-case ratios (4:1–7:1) were tested to assess consistency. The evaluation results show that the amplitude of oscillations was consistently significant for different control samples and control-to-case ratios.

Source: Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention*, *42*(2), 626-636.

Crash Time Estimation



- Time-series speeds at the detector stations immediately upstream (Macadam Ave.) and downstream (Morrison Bridge) of a crash that occurred.
- Data from this upstream station taken prior to 4:20 p.m. are used to represent pre-crash traffic conditions.

Source: Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention*, *42*(2), 626-636.

Table 2. Basic statistics for the potential explanatory variables (4:1 control-to-case ratio).

Potential Variable	Mean	Std. Dev.	Min	25% Percentile	75% Percentile	Max
Average speed[a]	24.743	11.237	2.303	16.542	30.448	49.938
Average count[b]	5.294	1.656	0.703	4.072	6.444	9.559
Average occupancy[c]	20.955	14.913	0.486	4.968	33.401	64.294
Std. dev. of speed[d]	4.982	3.057	0.327	2.928	6.015	19.957
Std. dev. of count	0.894	0.402	0.0993	0.595	1.098	2.460
Std. dev. of occupancy	4.364	3.829	0.380	1.233	6.386	30.609

The mean amplitude of traffic oscillations is measured by the standard deviation of speed is 5 mph (8 km/h). The standard deviation of oscillations (also in terms of standard deviation of speed) is 3 mph (5 km/h), ranging from 0.3 mph (0.5 km/h) to nearly 20 mph (32 km/h)

Source: Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention*, *42*(2), 626-636.

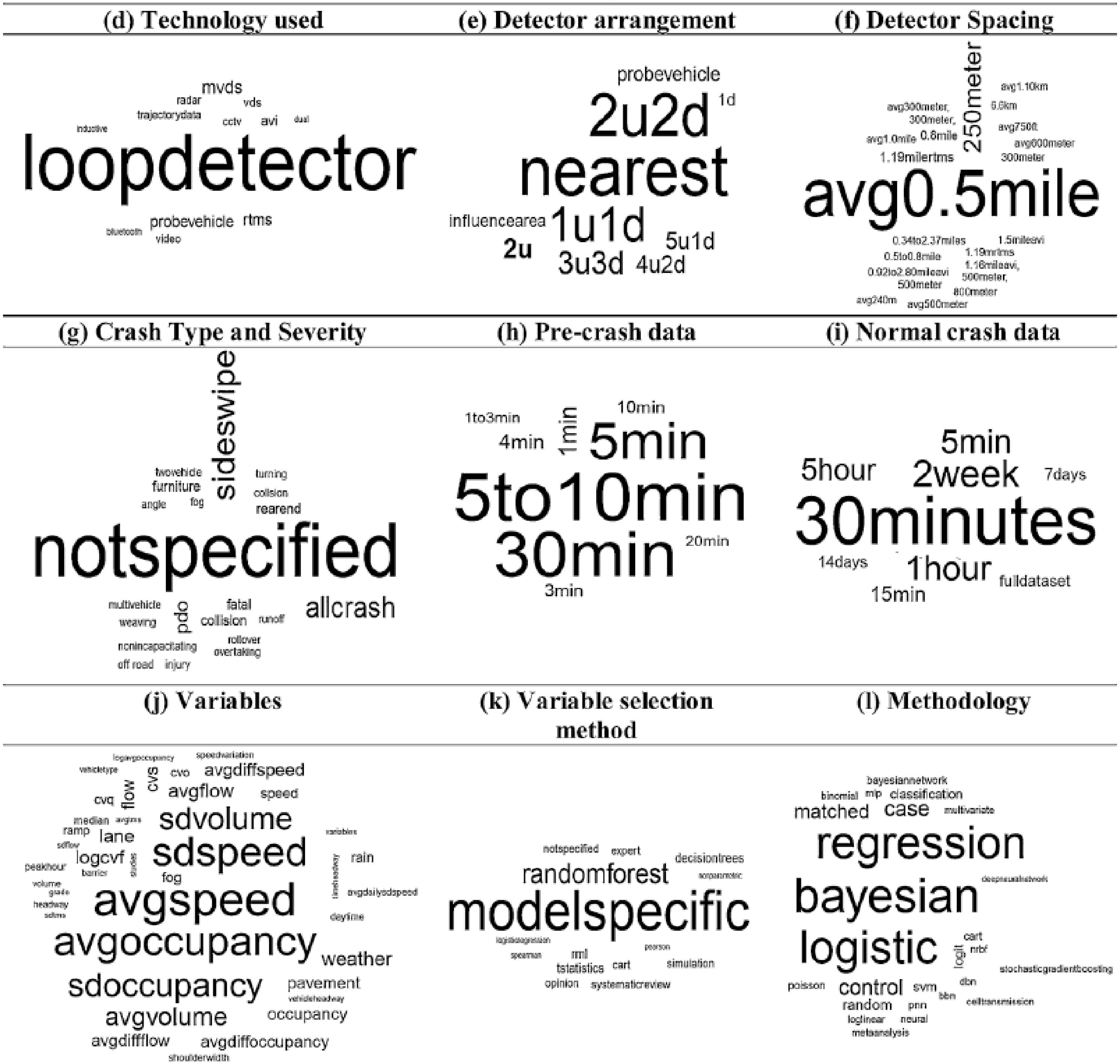
Modeling and Results (Zheng et al., 2010)

- Since flow, density and speed are highly correlated in congested traffic, only one of them is used to represent average traffic state or oscillations in the modeling process. Therefore, two variables (one for oscillations and one for average traffic state) were included.
- Removing outliers potentially can lead to over-fitting, which diminishes the effectiveness of the model; so, all data points were included.
- Interaction terms are not considered for the lack of a strong theoretical insight.
- The results reveal that the standard deviation of speed (thus, oscillations) is a significant variable, with an average odds ratio of about 1.08. The average traffic states prior to crashes were less significant than the speed variations in congestion.

Source: Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention*, *42*(2), 626-636.

A Note about Binary Logit and Conditional Logistic Regression Models

- Both **matched** and **unmatched** designs have been applied in the prediction of real-time crash risk, depending on the study's purpose.
- In an **unmatched** study design, non-crash cases are randomly selected. Confounding factors do not need to be matched in an unmatched design, so it requires less effort to identify non-crash cases. However, a large sample size is required to ensure accurate estimation, especially when the variable number is high.
- The **unmatched** design predicts the probability of a crash given the input variables.
- In the **matched** case-control study, however, the conditional logistic model does not provide the estimates of the nuisance factor, and thus it cannot directly predict the crash risk of given traffic conditions.
- Another issue with **matched** case-control studies is that a randomly chosen "control" is very likely to share similar traits with crash-prone conditions because crash cases constitute only a very small fraction of all data.
- Generally speaking, the **unmatched** case-control study handles the effects of confounding factors in the modeling stage by including them as variables in the regression equation; while the **matched** case control study tackles confounding factors in the data sampling stage.

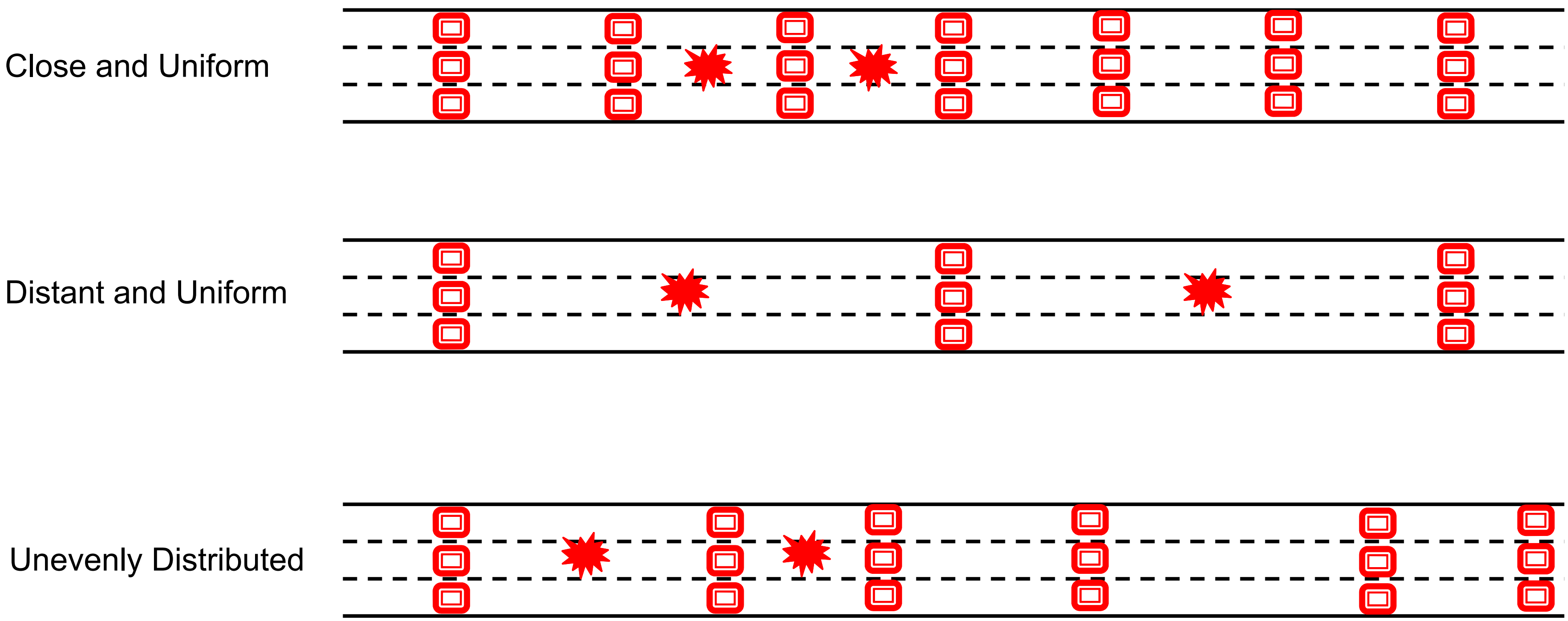


Text cloud summary of RTCPM

Source: Hossain, M., Abdel-Aty, M., Quddus, M. A., Muromachi, Y., & Sadeek, S. N. (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. Accident Analysis & Prevention, 124, 66-84.

Issues with Real-time Crash Prediction Methods

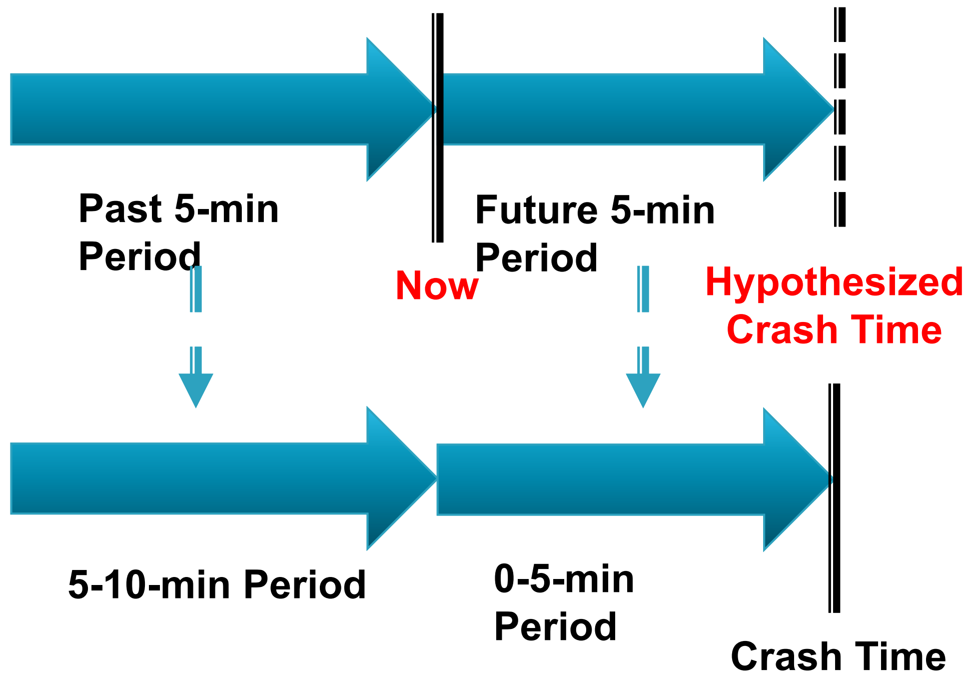
Spatial Issue



Temporal Issue

Crash Prevention

Crash Modeling



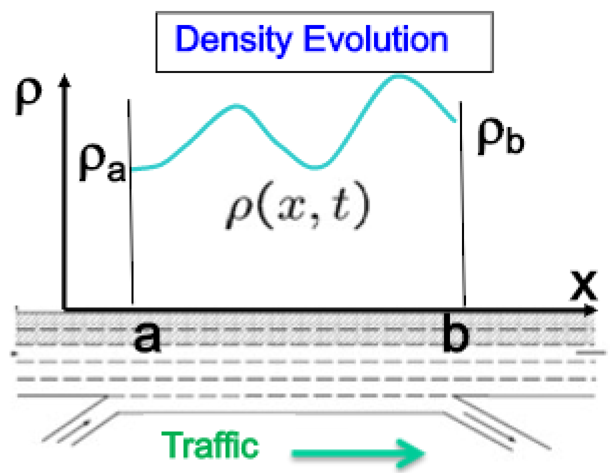
Using Traffic Simulation To Predict Crashes

- The presence of spatial-temporal discrepancies among loop detectors will undermine the accuracy of crash prediction that depends on quality traffic flow data from loop detector stations.
- Simulated traffic data may be used as an alternative data source in the event that discrepancies become a concern.
- The macroscopic traffic flow model is a popular consideration because it formulates the relationship among traffic flow characteristics like flow, density, and speed of a traffic steam.
- The cell transmission model (CTM) is a popular macroscopic traffic flow model that can capture many important traffic flow behaviors (i.e., queue formation and dissipation and shockwave propagation) on a given corridor (Daganzo, 1994).

Why CTM

- Cell transmission model (CTM) is a macroscopic simulation approach first proposed by Dr. Carlos Daganzo in his paper: "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory". *Transportation Research Part B: Methodological* 28.4 (1994): 269-287.
- CTM is a discretized framework for solving the kinematic wave theory, the Lighthill-Whitham-Richards (LWR) Model (Lighthill and Whitham 1955, Richards 1956), to model the traffic flow characteristics.
- LWR partial differential equation

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial q(\rho)}{\partial x} = 0$$



Fundamental Diagrams

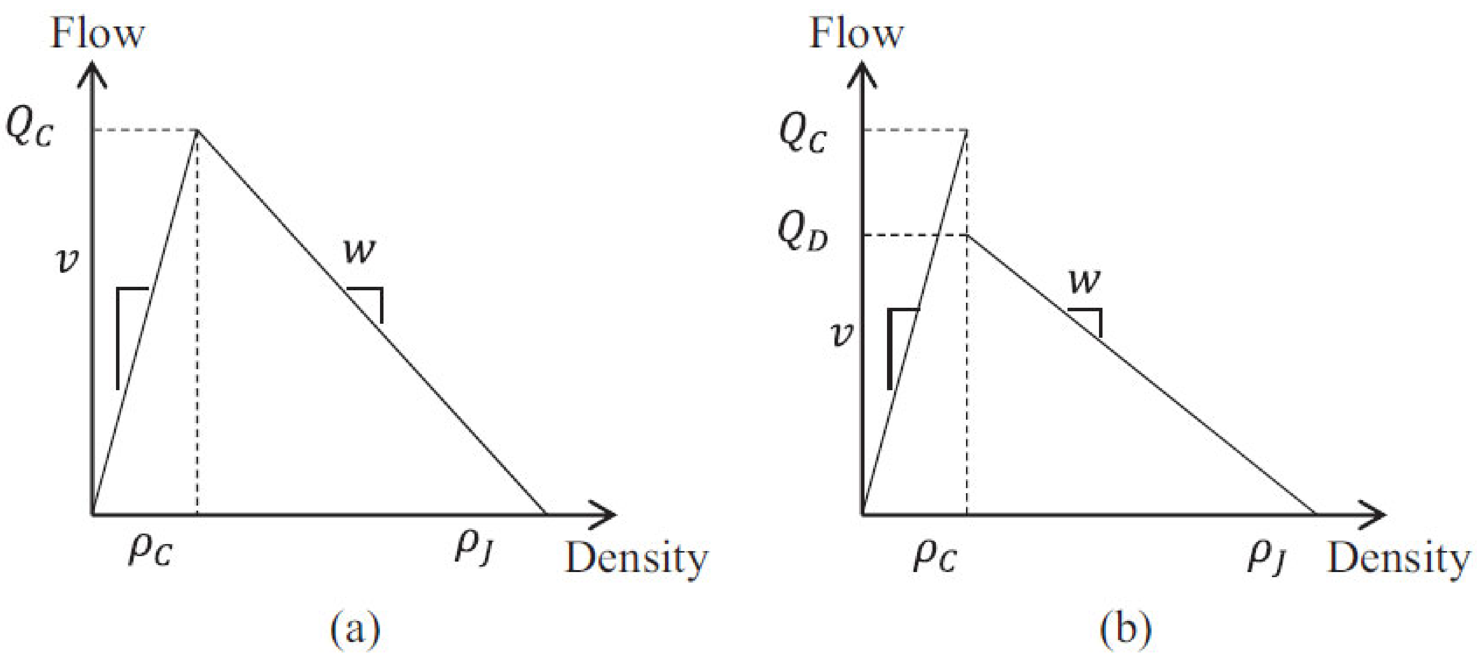


FIGURE 10.7 (A) Triangular fundamental diagram; (B) Fundamental diagram with capacity drop.

CTM Configuration

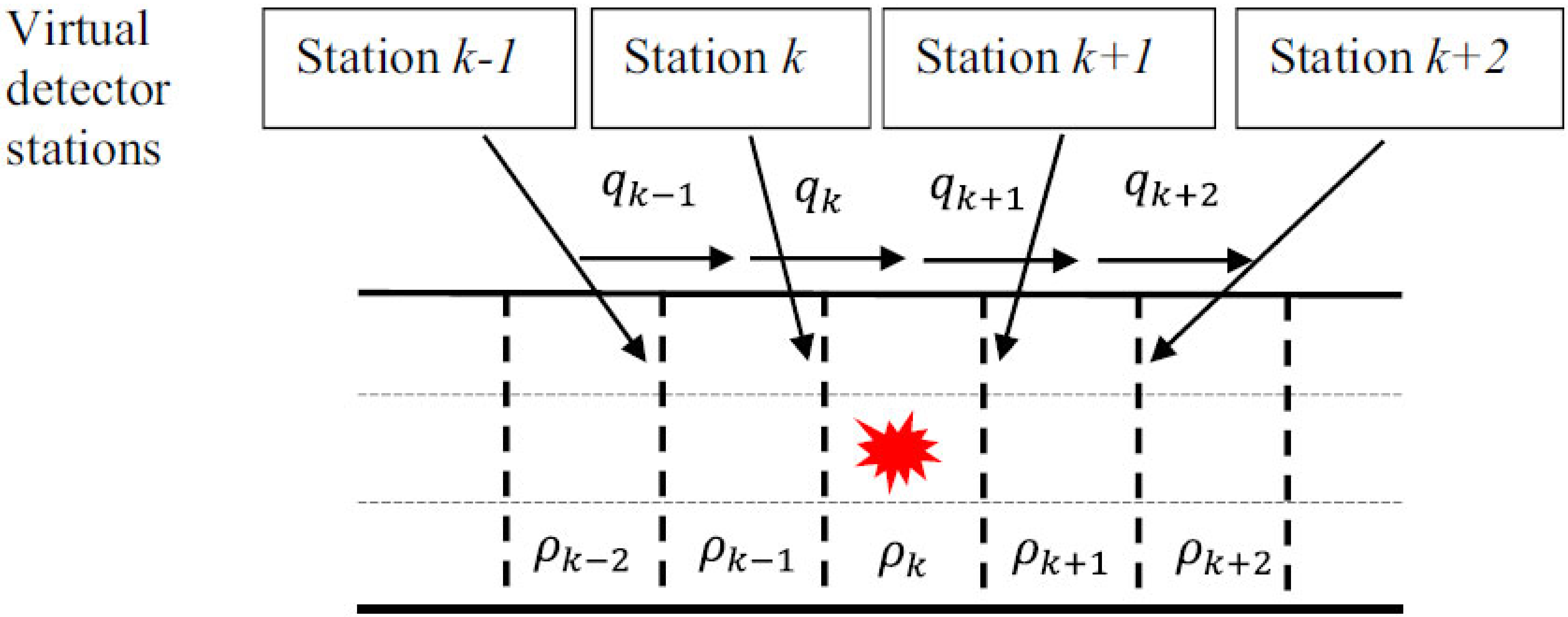
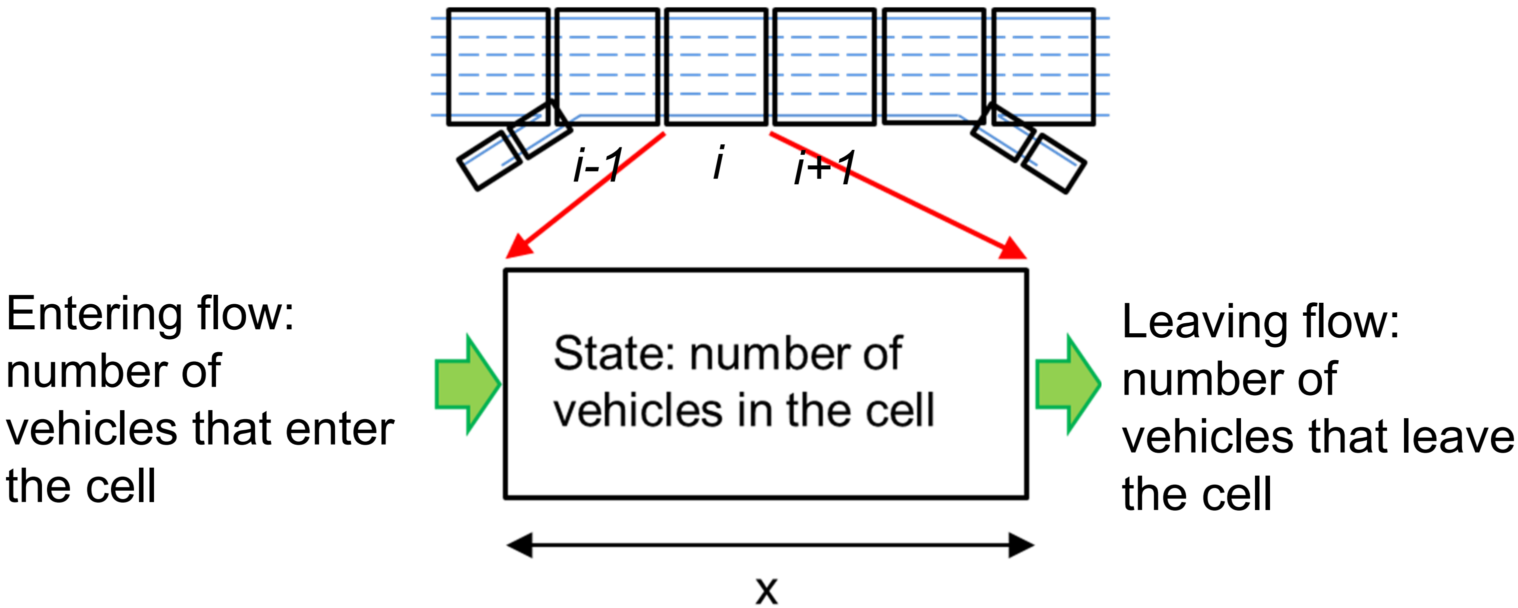


FIGURE 10.8 Layout of virtual loop detector stations.

Traffic Evolution in CTM



- Equation of the vehicle count of Cell *i* at time step *t+1*:

$$n_i(t+1) = n_i(t) + y_i(t) - y_{i+1}(t) = n_i(t) + R_i(t) \cdot T - S_i(t) \cdot T$$

$y_i(t)$: the number of vehicles entering Cell *i* in next time step, *t*

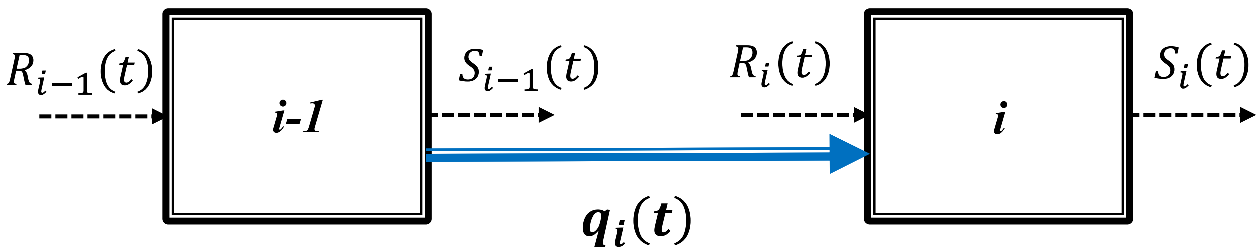
- Equation of traffic density of Cell *i* at time step *t+1*:

$$\rho_i(\mathrm{t}+1) = \rho_i(t) + \frac{T}{l_i}\left(q_i(t) - q_{i+1}(t)\right), \quad (10.14)$$

$q_i(t)$: the flow rate

Sending and Receiving Functions

- Receiving function ($R_i(t)$): the maximum flow rate that can be received by Cell *i* in next time step, *t*
- Sending function ($S_i(t)$): the maximum flow rate that can be supplied by Cell *i* in next time step, *t*



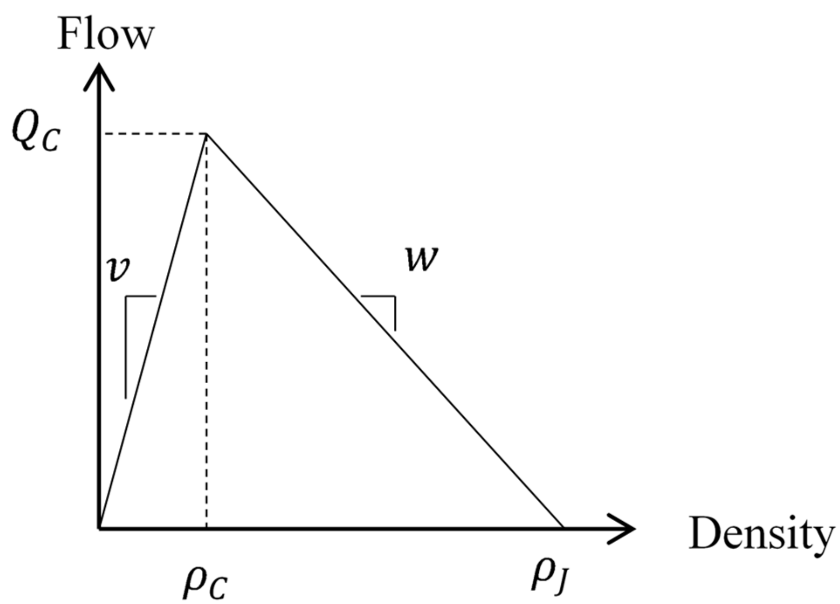
$q_i(t)$=min{$S_{i-1}(t)$, $R_i(t)$} for $1<i<N$

$q_1(t)$=min{$q_U(t)$, $R_i(t)$} for the first cell

$q_N(t)$=min{$S_{i-1}(t)$,$q_D(t)$} for the last cell

U: upstream
D: downstream

Fundamental Diagram (FD)



Q_C: capacity flow

ρ_C: critical density

ρ_J: jam density

v: free-flow speed

w: shockwave speed

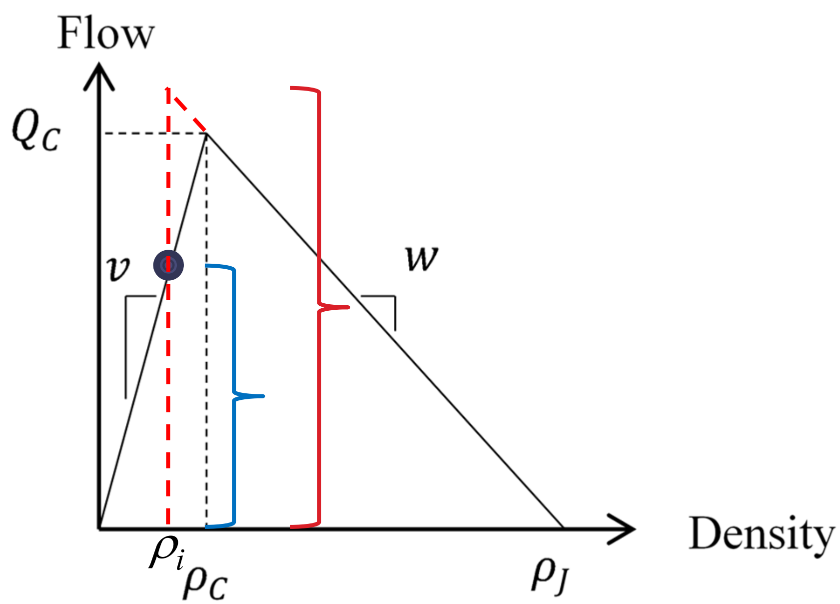
(10.15, 10.16, 10.17)

$$S_i(t) = \min\{v_i \rho_i(t), Q_{C,i}\}$$

$$R_i(t) = \min\{Q_{C,i}, w_i\left(\rho_{J,i} - \rho_i(t)\right)\}$$

$$q_i(t) = \min\{S_{i-1}(t),\ R_i(t)\} = \min\{v_{i-1}\rho_{i-1}(t), Q_{C,i-1},\ Q_{C,i}, w_i\left(\rho_{J,i} - \rho_i(t)\right)\}$$

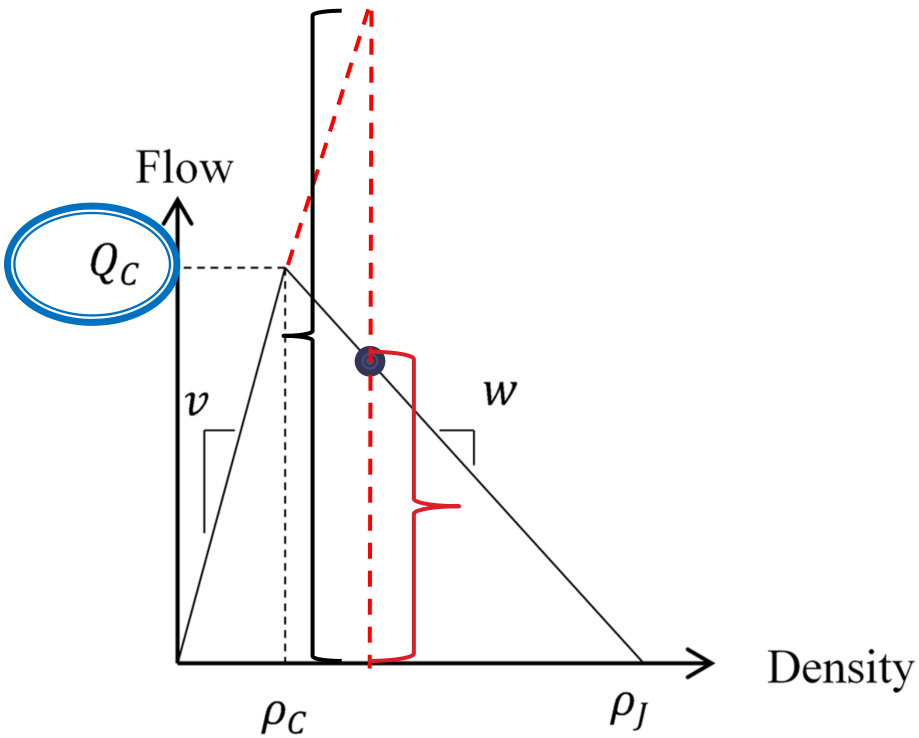
Fundamental Diagram (FD) (cont'd)



$$S_i(t) = \min\{v_i\rho_i(t), Q_{C,i}\}$$

$$R_i(t) = \min\{Q_{C,i}, w_i\left(\rho_{J,i} - \rho_i(t)\right)\}$$

Fundamental Diagram (FD) (cont'd)



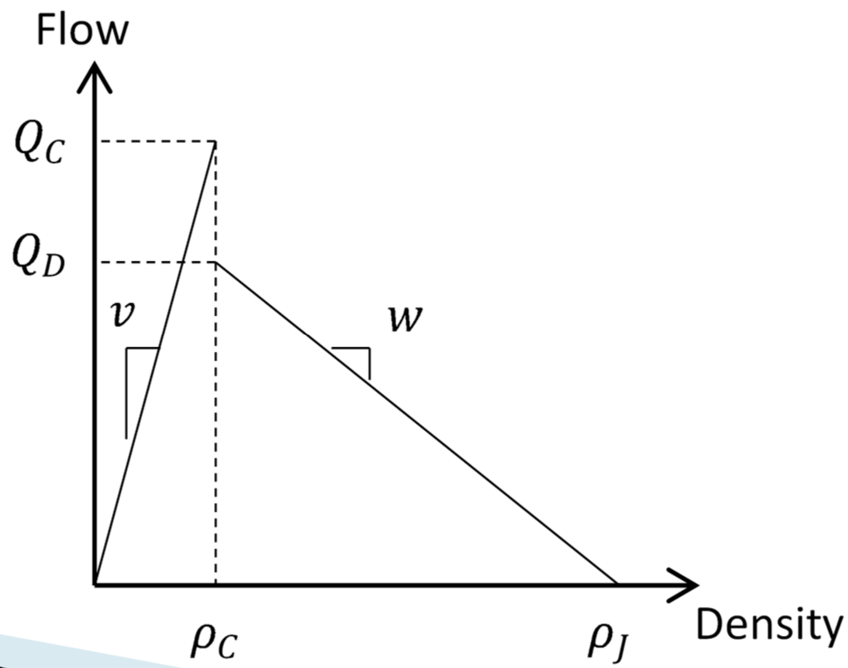
$$S_i(t) = \min\{v_i\rho_i(t), Q_{C,i}\}$$

$$R_i(t) = \min\{Q_{C,i}, w_i\left(\rho_{J,i} - \rho_i(t)\right)\}$$

- If $\rho_i(t) > \rho_C$ (congested), $S_i(t) = Q_{C,i}$, $R_i(t) = w_i\left(\rho_{J,i} - \rho_i(t)\right)$

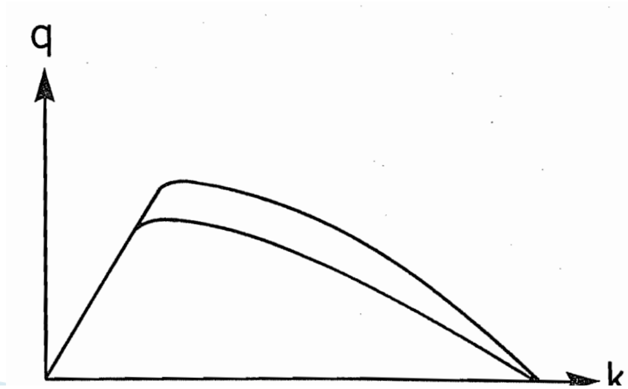
Modelling Traffic Phenomena in CTM – Capacity Drop

Studies of freeway bottlenecks have shown that **discharge flow** decreases after queues form just upstream (Hall and Agyemang-Duah, 1991; Banks, 1991; Cassidy and Bertini, 1999).



Modelling Traffic Phenomena in CTM – Oscillation/ Stop-and-go Traffic

- Traffic oscillations, also known as the "stop-and-go" traffic, refer to the phenomenon that congested traffic tends to oscillate between slow-moving and fast-moving states rather than maintain a steady state.
- Newell (1963) proposed a possible explanation: "*an instability would arise if drivers catching up with denser/slower traffic ahead were to delay braking, perhaps in the hope that traffic would clear up before they had to slow down. This behavior would result in average spacings shorter than usual when traffic was decelerating and would cause an instability*"



Crash Prediction with Simulated Traffic

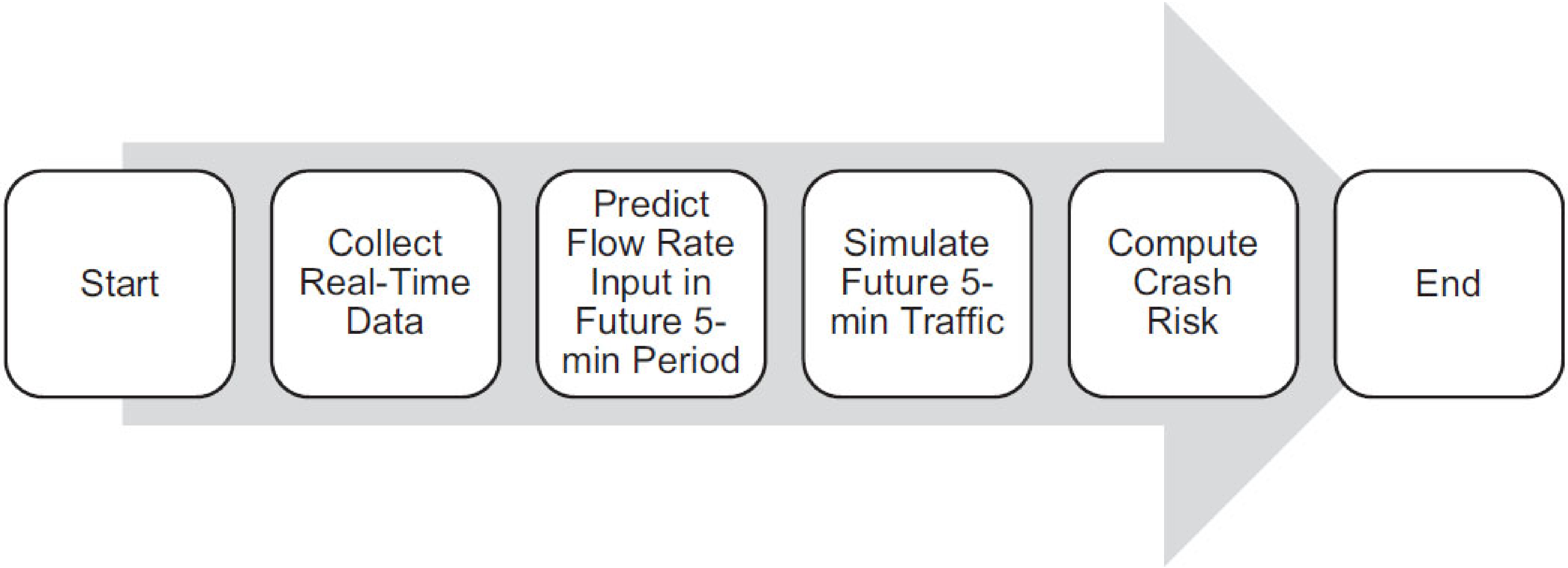


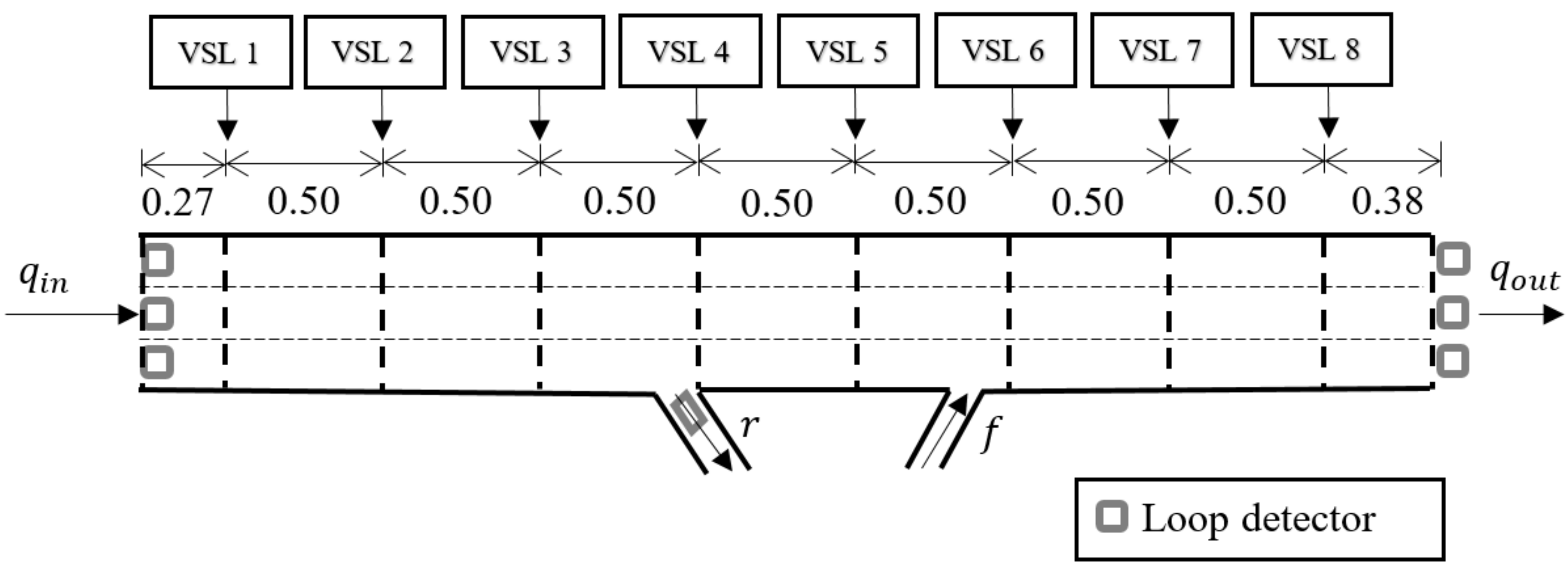
FIGURE 10.10 The procedures of predicting crashes with simulated traffic.

Case Study: Variable Speed Limit (VSL)

Variable Speed Limit (VSL)

Data

- The corridor was divided into 41 0.1-mi cells with 42 virtual detector stations
- Eight VSL signs are 0.5-mi apart
- Traditional CTM simulation environment



Variable Speed Limit Alternatives

- Target speed drop: 1) 10 MPH drop; 2) 20 MPH drop
- Speed limit change rate: 10 MPH per 30-s
- Maximum speed difference between neighboring VSL signs: 10 MPH (all VSL signs are coordinated)
- Activation rule: if the predicted crash probability of one cell in the future 5-min interval exceeds the pre-specified threshold, the nearest upstream VSL sign will be activated

$$p_i = \frac{e^{\pi}}{1+e^{\pi}}$$

$$\pi = -4.406 + 1.990 * BN + 1.764 * CT + 0.452 * (FF \times \ StdTsdDen_d)$$

$$+0.903 * (FF \times \ StdTsdSpd_d) - 1.049 * (FF \times OnRamp) + 1.146 * (FF \times Snow)$$

$$+0.530 * (BQ \times \ StdTsdDen_d) + 3.111 * (BQ \times Curve) + 0.00824 * (CT \times AvgDen_u)$$

Performance Assessment

- **Crash Risk (*R*):**

$$R = \sum_{i=1}^{I} r_i$$

$$r_i = \begin{cases} p_i - p_{thre}, & if\ p_i > p_{thre} \\ 0, & if\ p_i \leq p_{thre} \end{cases}$$

- **Total Travel Time (TTT):** the total of travel time that all vehicles spent on the corridor.
- **Relative change in *R* and TTT:**

$$\Delta M = \frac{\sum_{k=1}^{K} M_{k,CPS} - \sum_{k=1}^{K} M_{k,Non}}{\sum_{k=1}^{K} M_{k,Non}} \times 100\%$$

Comparison Results

Control Strategy	Count	ΔR	ΔTTT
(1) Non-activated VSL	59	0%	0%
(2) VSL: 10 MPH Drop	37	-26.9%	7.2%
(3) VSL: 20 MPH Drop	8	-10.5%	2.5%
Total	104	-21.2%	5.8%

- 104 out of 113 crash cases triggered the crash prevention module;
- The control strategy yielding the minimum crash risk is deployed

Conclusion: proposed strategy proves to be promising in improving safety without compromising mobility.

Future of the Real-time Crash Prediction Methods

RTCPMs by Purpose

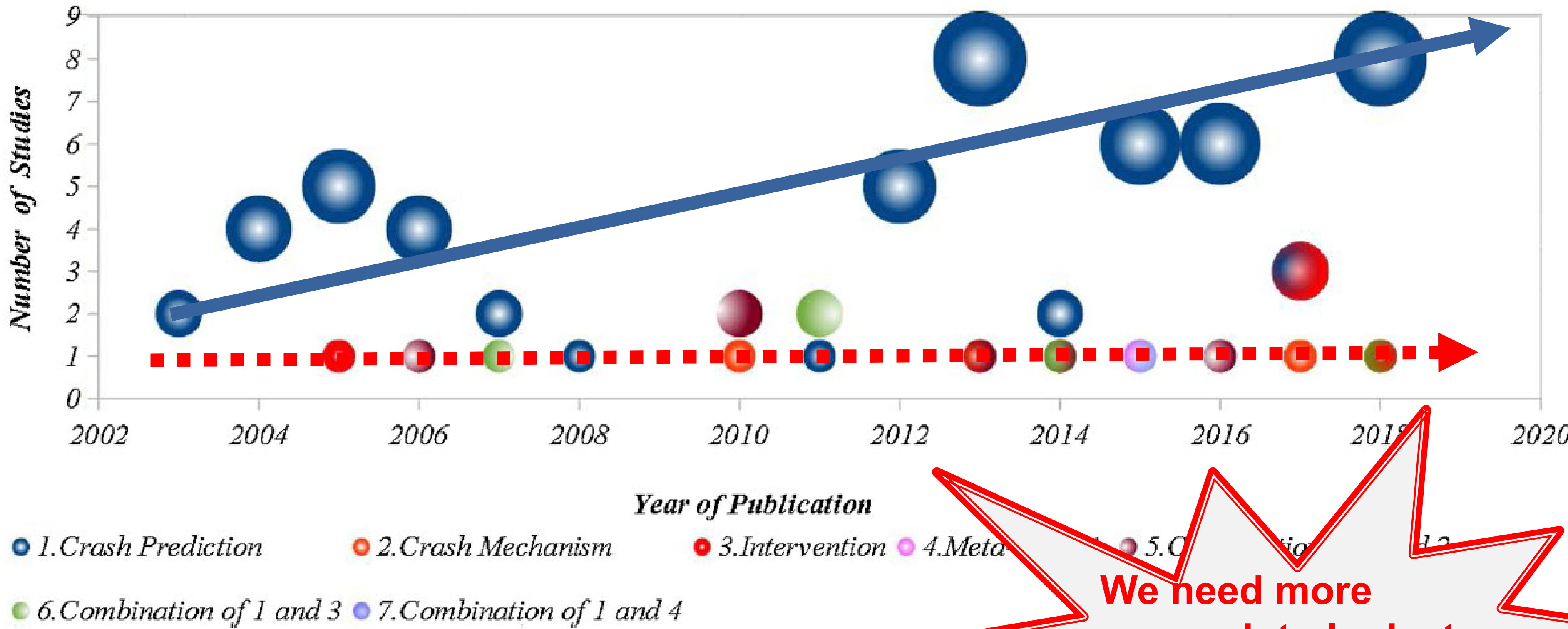


Fig. 3. RTCPMs constructed with various ma

We need more research to look at crash mechanism and effectiveness of intervention!

Source: Hossain, M., Abdel-Aty, M., Quddus, M. A., Muromachi, Y., & Sadeek, S. N. (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. Accident Analysis & Prevention, 124, 66-84.

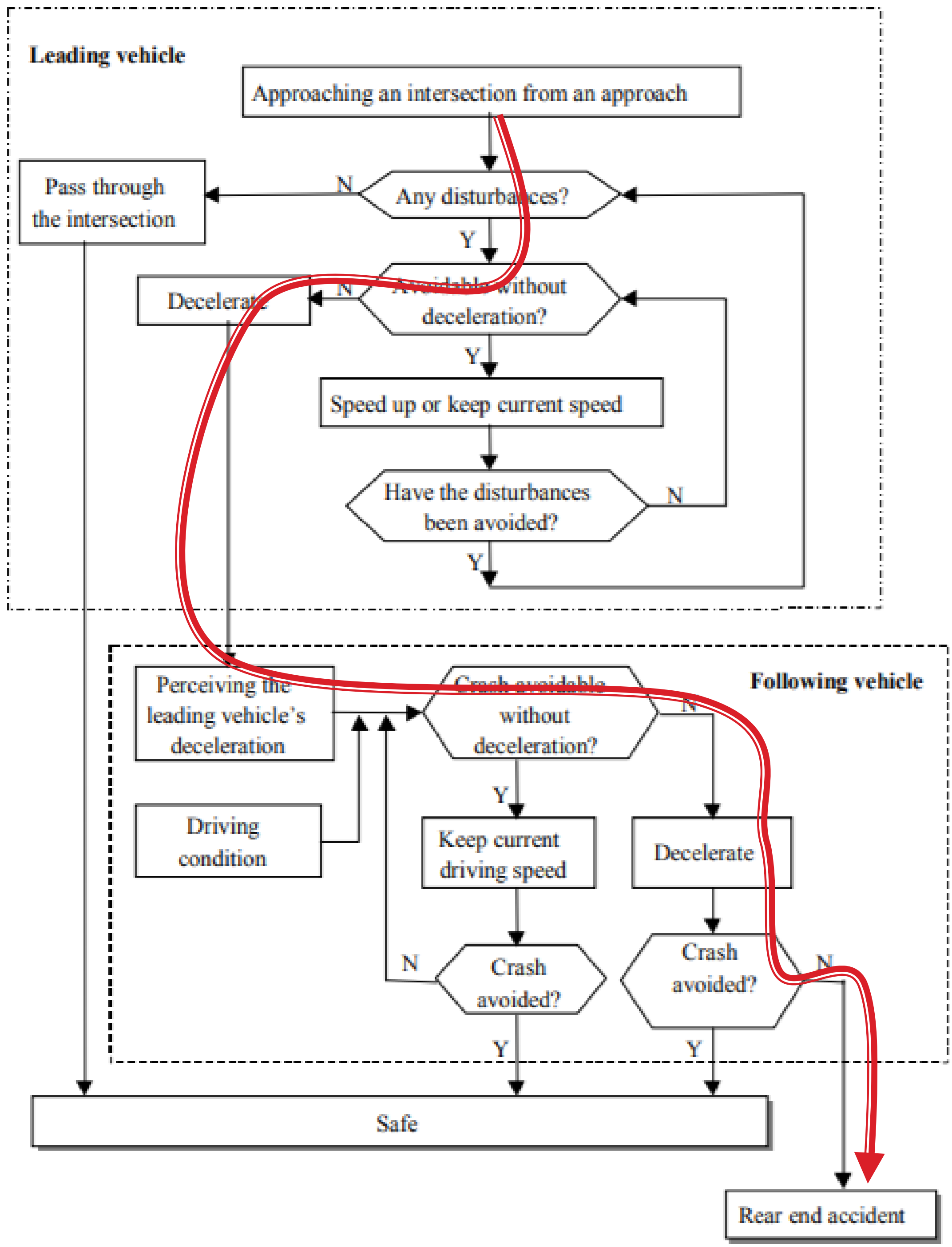
Example: an Occurrence-Mechanism Approach (Wang et al. 2002)

- A mathematical representation for a rear-end crash generating mechanism by expressing crash probability as the product of the probability of the lead vehicle decelerating and the probability of the following car's failure to respond properly.
- The probability of encountering an obstacle vehicle is assumed to be a function of the frequency of disturbances that cause the driver of a leading vehicle in a vehicle pair to decelerate.
- The probability of the trailing vehicle's driver failing to respond is the probability that this drivers' needed perception/reaction time is less than the available perception/reaction time.

References:

- Y Wang, H Ieda, F Mannering. Estimating rear-end accident probabilities at signalized intersections: occurrence-mechanism approach, Journal of Transportation engineering 129 (4), 377-384, 2002
- JK Kim, Y Wang, GF Ulfarsson, Modeling the probability of freeway rear-end crash occurrence,, Journal of transportation engineering 133 (1), 11-19, 2006

Flow chart of crash causation



Methodology

- Rear-end accidents are the result of a lead vehicle's deceleration (P_o) and the ineffective response of the following vehicle's driver (P_f): $P = P_o P_f$
- Assumed occurrence of disturbances follow a Poisson process, the probability of the driver of a leading vehicle encountering a disturbance j in time t as: $P_j = \int_0^t \eta_j e^{-\eta_j t} dt = 1 - e^{-\eta_j t}$
- Probability of the driver of a following vehicle encountering an obstacle vehicle is equal to the probability that at least one disturbance occurs within some specified time period: $P_o = 1 - \prod_j (1 - P_j) = 1 - e^{-\sum_j \eta_j t} = 1 - e^{-e^{\beta x}}$ (if $\sum_j \eta_j t = e^{\beta x}$ is assumed)

Methodology (cont'd)

- To incorporate perception/reaction time into a model of drivers' failure probability (P_f), consider available perception/reaction time (APRT) and needed perception/reaction time (NPRT).
- If NPRT is greater than APRT, crash happens. If APRT and NPRT are assumed to be random variables with some assumed distribution, drivers' failure probability can be determined.
- Assume APRT and NRPT follow the Weibull distribution:

$$P_f = \int_0^\infty \int_{t_a}^\infty f(\alpha,\lambda,t) f(\alpha,\gamma,t_a) dt dt_a = \int_0^\infty e^{-\lambda t_a^\alpha} \alpha\gamma t_a^{\alpha-1} e^{-\gamma t_a^\alpha} dt_a = \frac{1}{1+\lambda/\gamma}$$

- Assume $\frac{\lambda}{\gamma} = e^{-\varphi Z}$, then $P_f = \frac{1}{1+e^{-\varphi Z}}$ where φ and Z are vectors of estimable coefficients and explanatory variables affecting P_f, respectively

Methodology (cont'd)

- Given this, the number of accidents for some vehicle flow, v_{ik}, can be viewed as following a binomial distribution with the probability of having n_{ik} accidents; where

$$P_{ik} = P_{oik} \cdot P_{fik} = \frac{1 - e^{-e^{\beta \mathbf{X}_{ik}}}}{1 + e^{-\varphi \mathbf{Z}_{ik}}}$$

- Assume a Poisson-gamma model, the probability of n_{ik} crashes is a negative binomial distribution.

$$P(n_{ik}) = \frac{\Gamma(n_{ik} + \theta)}{\Gamma(n_{ik} + 1)\Gamma(\theta)} \left(\frac{\theta}{v_{ik} P_{ik} + \theta} \right)^{\theta} \left(\frac{\lambda_{ik} P_{ik}}{v_{ik} P_{ik} + \theta} \right)^{n_{ik}} \quad v_{ik}$$

where $\theta = 1/\delta$, or inversed-overdispersion parameter;

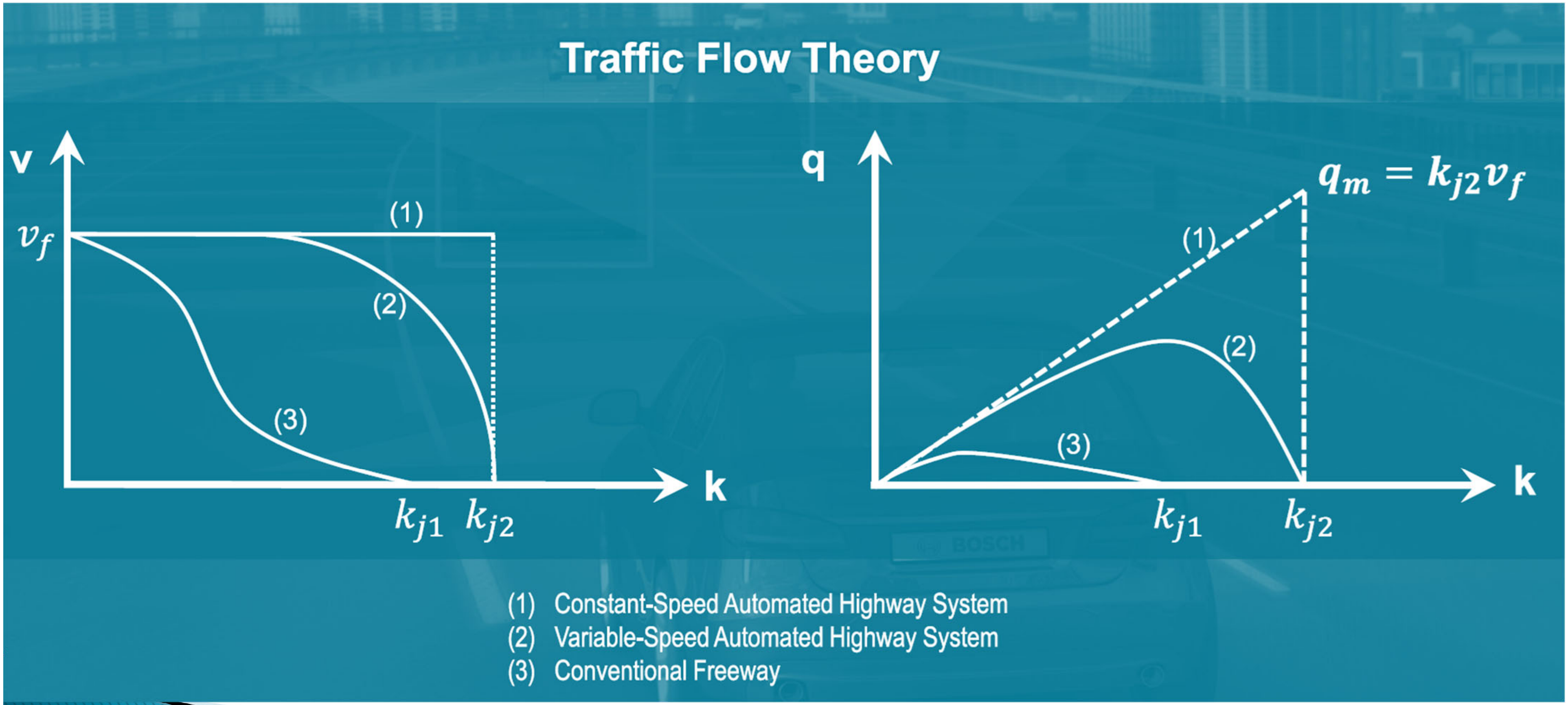
Table 1. Summary statistics for continuous variables.

Variable	Mean	Standard Deviation	Minimum	Maximum
Rear end accidents per approach	1.61	1.99	0.00	11.00
Daily left-turn traffic volume of the entering approach	3,109	2,347	0	20,467
Daily through traffic volume of the entering approach	12,844	9,088	142	51,660
Daily right-turn traffic volume of the entering approach	3,139	2,244	0	11,928
Speed limit of the entering approach (km/h)	49.37	8.83	30	60
Total number of approach lanes (including both entering lanes and exiting lanes)	4.96	1.88	1.00	11.00
Total number of lanes in left approach (including both entering lanes and exiting lanes)	5.02	1.94	1.00	12.00
Night-to-day traffic flow ratio	0.49	0.06	0.33	0.78

Findings (Wang, et al. 2002)

- By considering the occurrence mechanism of rear-end accidents, the model can explicitly account for human factors.
- Moreover, by considering factors that affect both the probability of encountering an obstacle vehicle and the probability of a driver failing to react quickly enough to avoid a collision with the obstacle vehicle, explanatory variables can be significant in both functions and sometimes affect overall accident probability in different directions.
- An example is the effect of the speed limit that was found to decrease the probability of encountering an obstacle vehicle, but increase the probability of a driver failure. Existing models with just canonical linear or log-linear link functions are unable to account for such dual impacts of important explanatory variables.

Automated and Connected Vehicles?



References

1. AASHTO., 2010. Highway Safety Manual. American Association of State Highway and Transportation Officials (AASHTO), Washington, D.C..
2. Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transport. Res. Rec. J. Transp. Res. Board 1897, 88–95.
3. Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Statistician 46 (3), 175–185.
4. Brodsky, H., Hakkert, A.S., 1983. Highway accident rates and rural travel densities. Accid. Anal. Prev. 15 (1), 73–84.
5. Cassidy, M.J., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. Transp. Res. Part B Methodol. 39 (10), 896–913.
6. Cedar, A., Livneh, M., 1982. Relationship between road accidents and hourly traffic flows-I and II. Accid. Anal. Prev. 14 (1), 19–44.
7. Chen, Z., Qin, X., 2019. A novel method for imminent crash prediction and prevention. Accid. Anal. Prev. 125, 320–329.
8. Chen, Z., Qin, X., Shaon, M.R.R., 2018. Modeling lane-change-related crashes with lane-specific real-time traffic and weather data. J. Intell. Transp. Syst. 22 (4), 291–300.
9. Courant, R., Friedrichs, K., Lewy, H., 1967. On the partial difference equations of mathematical physics. IBM J. Res. & Devel. 11 (2), 215–234.
10. Daganzo, C.F., 1994. The cell transmission model: network traffic. Transp. Res. Part B Methodol. 29 (2), 79–93.
11. Davis, G. A., & Swenson, T. (2006). Collective responsibility for freeway rear-ending accidents?: An application of probabilistic causal models. Accident analysis & prevention, 38(4), 728-736.
12. Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. Accident Analysis & Prevention, 43(6), 1907-1919.
13. Dervisoglu, G., Gomes, G., Kwon, J., Horowitz, R., Varaiya, P., 2009. Automatic calibration of the fundamental diagram and empirical observations on capacity. In: Transportation Research Board 88th Annual Meeting Proceedings.
14. Forbes, T.W., Zagorski, H.J., Holshouser, E.L., Deterline, W.A., 1958. Measurement of driver reactions to tunnel conditions. Highw. Res. Board Proc. 37, 345–357.
15. Frantzeskakis, J.M., Iordanis, D.I., 1987. Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece. Transport. Res. Rec. J. Transp. Res. Board 1112, 29–38.
16. Garber, N.J., Gadiraju, R., 1989. Factors affecting speed variance and its influence on accidents. Transport. Res. Rec. 1213, 64–71.
17. Gazis, D.C., Herman, R., Rothery, R.W., 1961. Nonlinear follow-the- leader models of traffic flow. Oper. Res. 9, 545–567.
18. Gwynn, D.W., 1967. Relationship of Accident Rates and Accident Involvements with Hourly Volumes. Traffic Quarter, pp. 407–418.
19. Hall, F.L., Agyemang-Duah, K., 1991. Freeway capacity drop and the definition of capacity. Transport. Res. Rec. J. Transp. Res. Board 1320, 91–98.
20. Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136 A K-Means Clustering Algorithm. Appl. Stat. 28, 100–108.
21. Harwood, D.W., Bauer, K.,M., Potts, I.B., 2013. Development of relationships between safety and congestion for urban freeways, transportation research record. J. Transp. Res. Board 2398, 28–36.
22. HCM., 2010. Highway Capacity Manual. Transportation Research Board (TRB), Washington, D.C..
23. Hosmer Jr., D.W., Lemeshow, S., 2004. Applied Logistic in Applied Logistic Regression, first ed.
24. Kononov, J., Allery, B., 2003. Level of service of safety: conceptual blueprint and analytical framework. Transport. Res. Rec. 1840 (1), 57–66.
25. Kononov, J., Lyon, C., Allery, B., 2011. Relation of flow, speed, and density of urban freeways to functional form of a safety performance function. Transport. Res. Rec. J. Transp. Res. Board 2236, 11–19.
26. Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Transport. Res. Rec. J. Transp. Res. Board 1784, 1–8.
27. Lord, D., Manar, A., Vizioli, A., 2005. Modeling crash-flow-density and crash-flow-V/C ratio for rural and urban freeway segments. Accid. Anal. Prev. 37 (No. 1), 185–199.
28. Mensah, A., Hauer, E., 1998. Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. Transport. Res. Rec. J. Transp. Res. Board 1635, 37–43.
29. Oh, J.S., Oh, C., Ritchie, S.G., Chang, M., 2005. Real-time estimation of accident likelihood for safety enhancement. J. Transport. Eng. 131 (5), 358–363.
30. Pande, A., Abdel-Aty, M., Hsia, L., 2005. Spatiotemporal variation of risk preceding crashes on freeways. Transport. Res. Rec. J. Transp. Res. Board 1908 (1), 26–36.
31. Persaud, B.N., Dzbik, L., 1993. Accident prediction models for freeways. Transport. Res. Rec. J. Transp. Res. Board 1401, 55–60.
32. Pipes, L.A., 1953. An operational analysis of traffic dynamics. J. Appl. Phys. 24 (No 3), 274–287.
33. Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid. Anal. Prev. 36 (2), 183–191.
34. Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., 2005. Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov chain Monte Carlo modeling. J. Transport. Eng. 131 (5), 345–351.
35. Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., Tepas, D., 2006. Bayesian estimation of hourly exposure functions by crash type and time of day. Accid. Anal. Prev. 38 (6), 1071–1080.
36. Roshandel, S., Zheng, Z., Washington, S., 2015. Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis. Accid. Anal. Prev. 79, 198–211.
37. Xu, C., Liu, P., Wang, W., 2016. Evaluation of the predictability of real-time crash risk models. Accid. Anal. Prev. 94, 207–215.
38. Xu, C., Liu, P., Wang, W., Li, Z., 2012. Evaluation of the impacts of traffic states on crash risks on freeways. Accid. Anal. Prev. 47, 162–171.
39. Xu, C., Liu, P., Wang, W., Li, Z., 2014. Identification of freeway crash-prone traffic conditions for traffic flow at different levels of service. Transport. Res. Pol. Pract. 69, 58–70.
40. Y Wang, H Ieda, F Man[illegible] Estimating rear-end accident probabilities at signalized intersections: occurrence-mechanism approach, Journal of Transportation engineering 129 (4), 377-384
41. Zheng, Z., Ahn, S., & Monsere, C. M. ([illegible]pact of traffic oscillations on freeway crash occurrences. Accident Analysis & Prevention, 42(2), 626-636.
42. Zhou, M., Sisiopiku, V.P., 1997. Relationship betwe[illegible]pacity ratios and accident rates. Transport. Res. Rec. J. Transp. Res. Board 1581, 47–52.