Adaptive Prediction of Traffic Incident Duration Using Change Detection and Bayesian Networks

Zilin Bian[a], Dachuan Zuo[a], Kaan Ozbay[a] and Jingqin Gao[a]

[a]Tandon School of Engineering, New York University, 6 MetroTech Center, Brooklyn, NY, USA;

ARTICLE HISTORY
Compiled February 3, 2025

ABSTRACT
Real-time incident management reduces non-recurrent congestion in transport systems. Accurate incident duration prediction is pivotal for traffic incident management and ensuring roadway safety. Recent advancements in prediction methods utilize machine learning with historical incident data and other relevant sources. However, incident duration is influenced by many dynamic factors within the transport system, such as varying traffic congestion and resources of the response team/system. Traditional machine learning models with fixed parameters may experience model drift issues, which degrades its prediction power due to changes in the environment. This paper introduces a traffic incident duration prediction model that incorporates a change detection method and Bayesian network model. The objective is to provide an accurate predictive traffic incident duration model which can automatically detect and adapt to system changes over varying temporal scales. Using New York City as a case study, our findings reveal systematic changes in the distribution of traffic incident duration during 2015-2021. The adaptability analysis also demonstrates that our proposed model is able to capture these temporal changes, update the learned parameters in real-time, and minimize the need for constant off-line calibrations.

KEYWORDS
Incident duration prediction; traffic evolution; change detection; Bayesian network

1. Introduction

The National Highway Traffic Safety Administration (NHTSA) (NHTSA 2022) estimated that 42,915 people died in motor vehicle traffic crashes in 2021, a 10.5 percent increase from 2020 and in fact the highest number of fatalities since 2005. To enhance road safety, traffic incident management (TIM) is often used. Effective TIM has been long proven to save the lives of emergency responders, reduce the occurrence of secondary crashes, and relieve congestion on the nation's roadways (Owens et al. 2009).

The duration of a traffic incident, abbreviated as TID in the paper (i.e., detection, response, clearance, and recovery time), has a well-documented relationship to the likelihood of crash survival (Owens et al. 2009), particularly the clearance time. At the operational level of a non-recurrent traffic incident, a total of four distinct stages are

CONTACT Zilin Bian. Email: zb536@nyu.edu

typically involved. These stages are closely aligned with the incident timeline (shown in Fig.1), with more information becoming available as the incident progresses from start to finish. Each stage marks an important step in the TIM process, providing insight into the current state of the incident and the necessary actions to be taken by transportation agencies. These operational-level changes in traffic incidents describe the present moment during the presence of the incident and operational actions are needed from transportation agencies in a real-time manner.

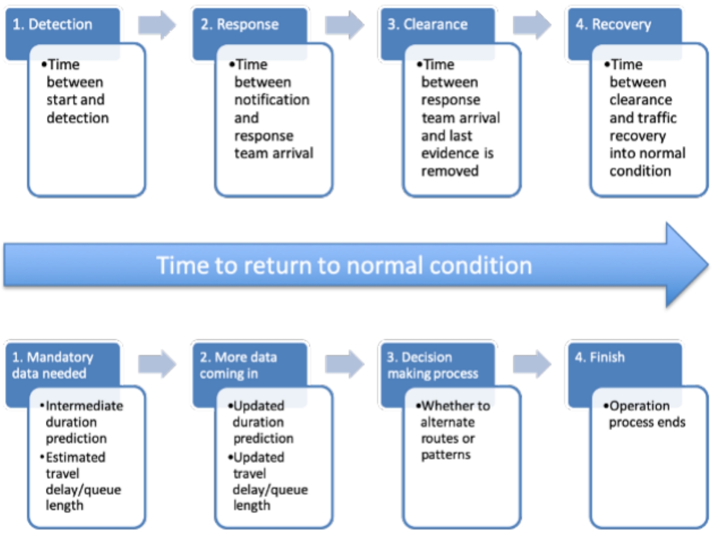


Figure 1. Operational level changes of a non-recurrent traffic incident

While the operational level changes of the traffic incident are necessary to be studied, the systematic changes for traffic incidents are also noteworthy. The systematic changes highlight the broader perspective of understanding the incident as a product of complex interactions between multiple factors of the transportation system (Dimitriou and Vlahogianni 2015), including the environmental factors (e.g., roadway geometry, weather condition), human factors (e.g., timely arrival of the emergency response team to the incident scene), traffic condition (e.g., upstream and downstream traffic flow) and incident related factors (e.g., incident type, fatalities and injuries).

Among these factors, few are constant, and most experience significant changes over time. For example, the travel demand and congestion conditions in New York City in 2019 increased by 10% compared to 2017 and have evolved continuously year by year (Reed 2019). Another example is the onset of the COVID-19 pandemic, which significantly changed traffic demand, travel behavior, available resources, and labor requirements for response teams, and so on. These changes, expressed in terms of individual factors as well as the web of systemic interactions, in turn, introduce challenges in accurately estimating the duration of traffic incidents in real time. There is therefore a need to either build a new predictive incident duration model or update the parameters of the existing model every time an evolutionary change is discovered.

Nevertheless, most studies dealing with incident duration prediction in the existing literature (Knibbe et al. 2006; Wei and Lee 2007; Lin, Wang, and Sadek 2016; Kuang et al. 2019) are modeled and estimated using historical data, with the assumption of temporal stability within the system. Therefore, these data-driven predictive models may no longer be suitable once evolutionary changes are incorporated using new data. This is called "model drift" (or concept drift), a term often used in streaming learning (Žliobaitė 2010) that describes when an original model does not adapt to the up-to-date situation and therefore can no longer provide satisfactory predictive results. In the case of the prediction of traffic incident duration, or TID prediction, model drift manifests when the previously trained model cannot adapt to the characteristics of the most recent traffic incidents and thus may no longer provide good predictions. The rapid advancement of traffic sensing technologies has significantly transformed the collection of traffic incident information. For example, Traffic cameras equipped with computer-vision technology can now automatically detect incidents and relay detailed reports to transportation management agencies (Zuo et al. 2023). This acceleration in data collection not only increases the volume of information available but also underscores the need for an adaptive model capable of handling rapid technological evolution and the resulting data accumulation. These advanced sensing technologies are changing the landscape of traffic incident management by altering the nature of collected data. This evolution affects both the numerical values of Traffic Incident Duration (TID) and the characteristics of associated traffic incident attributes. As these technologies continue to develop, they introduce variations in how incidents are measured and in the types of attributes that are collected and analyzed. Consequently, it is essential for predictive models of TID to adapt not only to fluctuations in numerical values but also to the broader changes in the characteristics of traffic incidents. Models that can dynamically accommodate these shifts in data capture will be better positioned to provide accurate and timely predictions, aligning with the capabilities of modern traffic management systems.

To address these challenges, our study develops a Traffic Incident Duration (TID) predictive model that integrates a Change Detection (CD) method with Bayesian Networks (BNs). This combination equips the model to adapt to both evolutionary and long-term changes within the traffic system. BNs are particularly useful at managing changes in input attributes by dynamically adjusting their learned knowledge structure. They are effective at making predictions even with incomplete input information, thereby ensuring their effectiveness in both immediate and progressive changes in traffic incidents. Demiroluk and Ozbay (2014); Kuang et al. (2019) demonstrated that the BNs can handle limited incident information as the incident progresses through different stages, and can update the prediction results as new information becomes available. The CD method enables the proposed model to adapt to the systematic and long-term changes by detecting changes that signals potential model drift. A novel strategy is designed to acquire such signals, verify when model drift occurs, and send messages to the BNs to update the learned parameters with the new data pattern. The results show that the proposed adaptive TID prediction model can detect significant changes in the long-term temporal development of TIDs and adapt to up-to-date situations without losing prediction power. To this end, the contributions of this paper are as follows:

- We conduct a descriptive analysis to study the temporal characteristics of the TIDs. We decompose the distribution of TIDs and extract the seasonal and trend patterns of TIDs and verify that there are obvious seasonal patterns and trend changes in the TIDs.
- We propose a hybrid framework using CD method and BN model to predict the TID in an adaptive manner. This framework is constructed upon popular CD method to detect changes and the BN model composing of various structure and parameter learning approaches which make it flexible to adapt to different model drifts.
- By comparing our proposed model with different update strategies, we demonstrate that our proposed model is able to provide satisfactory prediction results without the need of frequently/repetitively re-training the modeling and updating the parameters.
- Our findings provide insights for researchers and decision makers that the TIDs may also evolve and change at a long-term dimension over seasons and years. Capturing such changes, learning the development and adapting to the current situations are very important to the applicable usage of the TID prediction model.

2. Literature review

2.1. *Traffic incident duration prediction model*

A large amount of studies focus on estimating and predicting TID using plethora approaches, such as statistical methods and data-driven methods. Statistical methods, such as regression (Khattak, Schofer, and Wang 1995; Garib, Radwan, and Al-Deek 1997; Peeta, Ramos, and Gedela 2000; Khattak et al. 2016), parametric methods using probabilistic distribution analysis (Golob, Recker, and Leonard 1987; Chung and Yoon 2012; Alkaabi, Dissanayake, and Bird 2011), and hazard-based models (Qi and Teng 2008; Chung 2010) are popular approaches used by researchers to predict the TID.

Apart from statistical and parametric approaches, researchers also tended to use data-driven models such as machine-learning techniques to investigate the TIDs. Given the availability of massive data collected and generated within transportation systems, machine learning models are able to borrow the power of data fusion and explore the relationship between the TIDs and other incident-related attributes, without requiring any predetermined and explicit assumptions. Machine learning models used in existing studies to predict the TIDs include tree-based models (Ozbay and Kachroo 1999; Knibbe et al. 2006),(Zhan, Gan, and Hadi 2011), Artificial neural network (ANN) models (Wei and Lee 2007), (Yu et al. 2016), Bayesian networks (BNs) (Ozbay and Noyan 2006; Demiroluk and Ozbay 2014), support vector machine (SVM) (Yu et al. 2016), and hybrid models (He et al. 2013; Lin, Wang, and Sadek 2016; Kuang et al. 2019). For example, (Ozbay and Noyan 2006; Demiroluk and Ozbay 2014) adopted the BNs to model the TIDs with the consideration of the stochastic variation and the presence of missing information of incident data. There are also hybrid models that incorporated multiple approaches, He et al. (2013) adopting a tree-based quantile regression method which incorporates the merits of both quantile regression and tree-structured modeling. Kuang et al. (2019) combined the BNs with k-nearest neighbor (KNN) to predict the TIDs according to the incident severity.

2.2. *Temporal analysis in traffic incident duration*

A traffic incident is a streaming process whose status evolves through different stages. Existing studies apply different approaches to deal with the temporal changing information but they broadly follow the same logic–that is, training the separate models for different phases/stages of a traffic incident. For instance, Wei and Lee (2007) adopted the ANN and introduced a multi-period forecast framework to encode the temporal changes. They proposed two types of incident duration models in which Model A performs the preliminary forecast at the detection stage, while Model B takes over as more information becomes available. Qi and Teng (2008) introduced a time-sequential procedure for online prediction of TIDs using hazard-based models. The procedure contains multiple stages in the development of the incident and applies an individual hazard-based model for each stage with different variables upon the available information.

However, the changes exist not only in the development of each individual traffic incident but also in the changing environment, with a long-term temporal dimension. As previously stated, factors such as environmental factors and traffic flow can evolve and change over seasons and years, thereby the predictions of TIDs could also evolve and change. Nevertheless, almost all current studies estimated and trained their models in a one-time manner and did not revisit those models when new data became available in later years; nor have they investigated whether their models can adapt to changing environments and continue to provide satisfactory results.

Therefore, it is necessary to investigate if the TIDs evolve and change along with the environment over time. Moreover, the duration prediction model must capture, adapt to, and update based on such changes, in order to provide satisfactory prediction results.

2.3. *Available change detection methods*

To adapt to the changing environment and meet the challenges brought on by disruptive changes, CD methods have attracted attention due to their capability to detect interruptions and changes in time-series data. The CD methods can be classified into two methodologies: online change detection and offline change detection.

Online change detection methods are also referred to as drift detection methods (DDM). The traditional statistical approaches of DDM report detection results by comparing the difference between the statistical distributions (Harrou, Sun, and Madakyaru 2016; Salem, Naït-Abdesselam, and Mehaoua 2012). Other popular statistical methods, such as the early drift detection method (EDDM),(Baena-Garcıa et al. 2006) and the cumulative-sum based method (CUMSUM) (Mohanty, Pradhan, and Routray 2007), can provide different signal levels, such as warnings and alarms when a change is detected. There are also window-based approaches, such as the adaptive window-based (ADWIN) method (Hassani 2019), which finds changes in time-series data by comparing the averages in two subject time windows. The online CD methods are appropriate for the usage of exhaustive systems, which require monitoring any immediate changes constantly.

The offline CD methods usually consist of cost functions and search methods. The cost functions are usually estimated using a likelihood ratio estimation (Truong, Oudre, and Vayatis 2020) approach and a regression structure (Bai and Perron 2003) approach, where the search methods include exact search methods and approximate search methods. Exact search methods usually have greater computational complexity than approximate search methods but are more accurate since they seek more optimal results. Pruned exact linear time (PELT) (Killick, Fearnhead, and Eckley 2012) and the segment neighborhood search method (Auger and Lawrence 1989) are commonly used exact search methods. The popular approximate search methods include binary segmentation (Scott and Knott 1974) and the window-based method (Truong, Oudre, and Vayatis 2020). The details of model comparison and selection will be described in Section.4.

3. Traffic incident data and descriptive analysis

The data used in this paper was provided by the New York City Department of Transportation (NYCDOT). The dataset consists of 96,414 non-recurrent traffic incidents that occurred in the area of New York City from 2015 to 2021, of which two major incident types were included: crashes and disabled vehicles. This dataset compiles traffic incident information gathered from a variety of sources, including detections by the Transportation Management Center (TMC), reports from police departments, and data collected through crowdsourcing services such as Waze (Amin-Naseri et al. 2018). It is important to note that the data provided by NYCDOT contain various types of traffic incidents such as crashes, disabled vehicles, construction activities, and so on, in this study we only focus on the crash- and disabled-vehicle-related events. Moreover, in this paper we adopt the definition of TID as mentioned in (Li, Pereira, and Ben-Akiva 2018), which specifies TID as the time period between the detection of a traffic incident (Stage 1) and the clearance of the incident (Stage 3), as illustrated in Fig.1.

Table.1 provides a description of the available traffic incident-related variables extracted from the dataset that will be used for TID modeling purposes. The characteristics related to the traffic incident include the incident features (the involvement of injury, lane closure type (e.g., number of lanes affected), the involvement of heavy vehicle and fire), the spatial features (direction, county), and the temporal features (peak hour or off peak, weekday or weekend, season and year).

Table 1. Description of the available incident-related variables

Variable	Type	Description
TID	Continuous	The traffic incident duration in minutes
Injury involved	Categorical	0: no history, 1: Injuries involved
Lane closure type	Categorical	0-3, 0: zero travel lane, 1: one travel lane, 2: more than two travel lanes, 3: all travel lanes
Heavy vehicle involved	Categorical	0: no history, 1: heavy vehicle involved
Fire involved	Categorical	0: no history, 1: fire observed
Direction	Categorical	0-5, 0: both directions, 1: east, 2: west, 3: south, 4: north, 5: no information
County	Categorical	0-4, 0: manhattan, 1: kings, 2: queens, 3: bronx, 4: richmond
Peak hour	Categorical	0: off peak, 1: peak hour
Workday	Categorical	0: weekday, 1: weekend
Season	Categorical	0: spring and fall, 1: summer, 2: winter
Year	Categorical	0-6, 0: 2015, 1: 2016, 2: 2017, 3: 2018, 4: 2019, 5: 2020, 6: 2021

The TID is calculated based on the reported start and end timestamp based on the obtained data records. We aggregated the number of traffic incidents by every month between 2015 and 2021, the distribution of traffic incident numbers for crash and disabled vehicle are shown in Fig.2 (e). We obtained and calculated the duration of each individual traffic incident to investigate whether temporal patterns and changes exist in the distribution of the TID in this 7-year-long dataset. To be specific, we adopted a conventional time series decomposition method called Seasonal and Trend decomposition using LOESS (STL) (Cleveland et al. 1990), to decompose the distribution of the TID and investigate if any seasonal patterns and trends exist. The STL is able to decompose the time series data and extract three types of components: trends, seasonality, and residual with outliers.

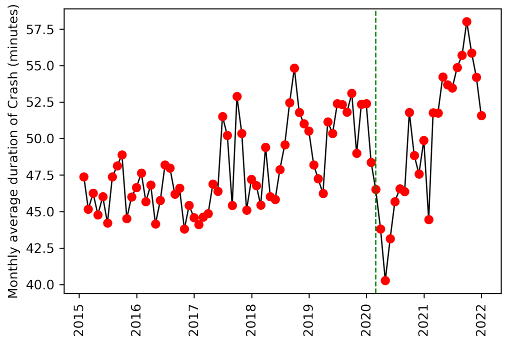
For each individual type of traffic incident, we conducted temporal analysis by mapping the distribution of the monthly averaged TIDs and extracting the decomposed features through the use of STL method. The distribution of the TIDs and the decomposed results (trend, seasonality, and outlier) are presented in Fig.2. Fig.2 (a) and (b) presents the distribution of crash TIDs and its decomposed features. By looking at the trend and distribution of the TIDs, it is easy to observe that the TIDs remained a relatively constant trend and ranged between 42 minutes and 48 minutes from 2015 to 2017. The trend of TIDs started to increase after 2017 until it saw the first drop in the first half year of the COVID-19 pandemic in 2020. Then the trend met a rebound from the second half of 2020 to the end of 2021. The TIDs maintained a relatively unstable behavior ranging between 40 minutes and 54 minutes from 2017 to 2021. The green dashed line in Fig.2 represents the start time of the COVID-19 pandemic in New York City in 2020. It is then easy to notice that the TIDs met the most significant changes before and after the COVID-19 pandemic: a 21.5% drop within 5 months and a (25%) increase in the next 6 months. The seasonal patterns of crash TIDs also met similar phenomena: the seasonality shares similar patterns between 2015 and 2017 and changed into a different pattern from 2017 to 2021. This indicates that the incident-related factors such as the environmental factors and traffic flow conditions varied and evolved, causing different traffic patterns between 2015-2017 and 2017-2021.

TIDs of disabled vehicle shared similar patterns with crash: 1) the seasonal patterns are similar between 2015 and 2017 then changed to another pattern between 2017 and 2020, 2) the trend of the TIDs of disabled vehicle met a decrease in 2015-2017, an increase in 2017-2019, a decrease and then a rebound from 2019 to 2021. The COVID-19 pandemic also caused the most significant variant, a 41% increase between the COVID-19 pandemic years 2020 and 2022. Moreover, the number of traffic incidents demonstrated similar seasonal patterns in Fig.2 (e) for both crashes and disabled vehicles, with a significant decrease observed during the COVID-19 period.

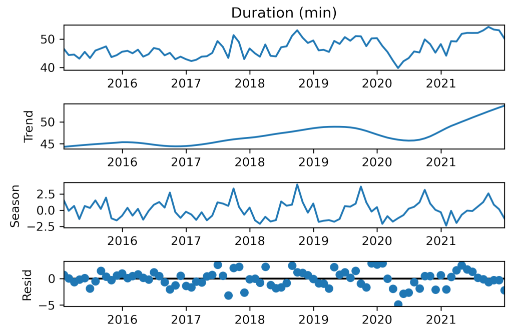
These findings confirm the aforementioned assumptions that the TIDs may meet gradual and sudden changes and also seasonal patterns along the long-term temporal dimensions. Therefore, one cannot ensure that a one-time model can capture such changes and adapt to provide satisfactory results.

4. Methodology

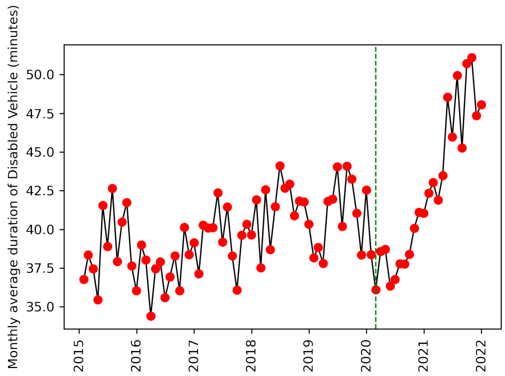
In this study we propose a hybrid framework that combines use of the change detection (CD) methods and Bayesian networks (BNs) to adaptively predict the TIDs. Fig.3 shows the workflow of the proposed methodological framework applied to detect both trend changes as well as sudden changes in the distribution of the TIDs and provide the prediction results using a novel update strategy scheme. The proposed framework contains three stages: 1) monitoring, 2) modeling, and 3) evaluating.



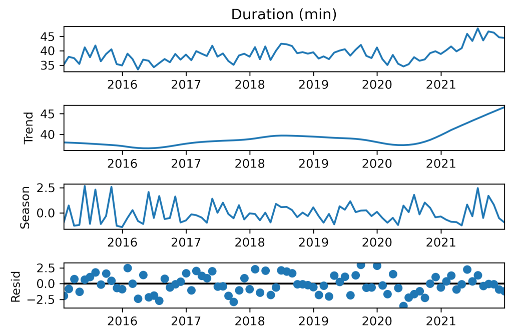
(a) TID for Crash



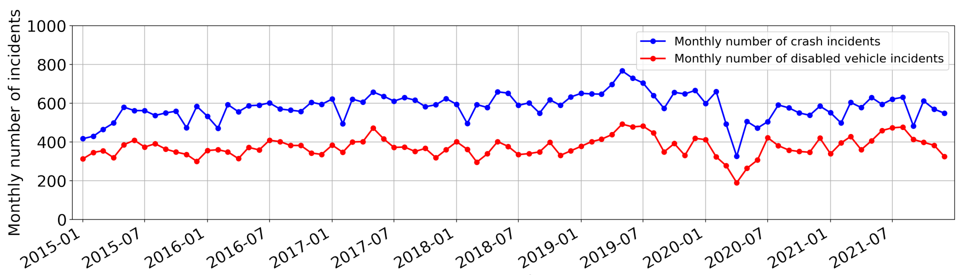
(b) Decomposed features for Crash



(c) TID for Disabled Vehicle



(d) Decomposed features for Disabled Vehicle



(e) Monthly number of incidents for Crash and Disabled Vehicle

Figure 2. TID distribution, decomposed features, and monthly number of Crash and Disabled Vehicle incidents.

4.1. *Problem definition*

Stamped by the occurrence time of each traffic incident, we consider the TIDs as a time-series sequence which can be aggregated at different temporal levels such as daily, weekly and monthly levels. Denote the time-series of the TIDs as a non-stationary sequence $y = y_1, ..., y_T$ that takes value in $\mathbb{R}^d (d >= 1)$ and has T observations. The problem formulation of this study will be divided into two parts: 1) the objective of the CD method is to extract those timestamps in y where sudden changes or trend changes are detected, and 2) the initial learned BNs will test and validate if any model drift exist at those detected timestamps and update the learned parameters based on

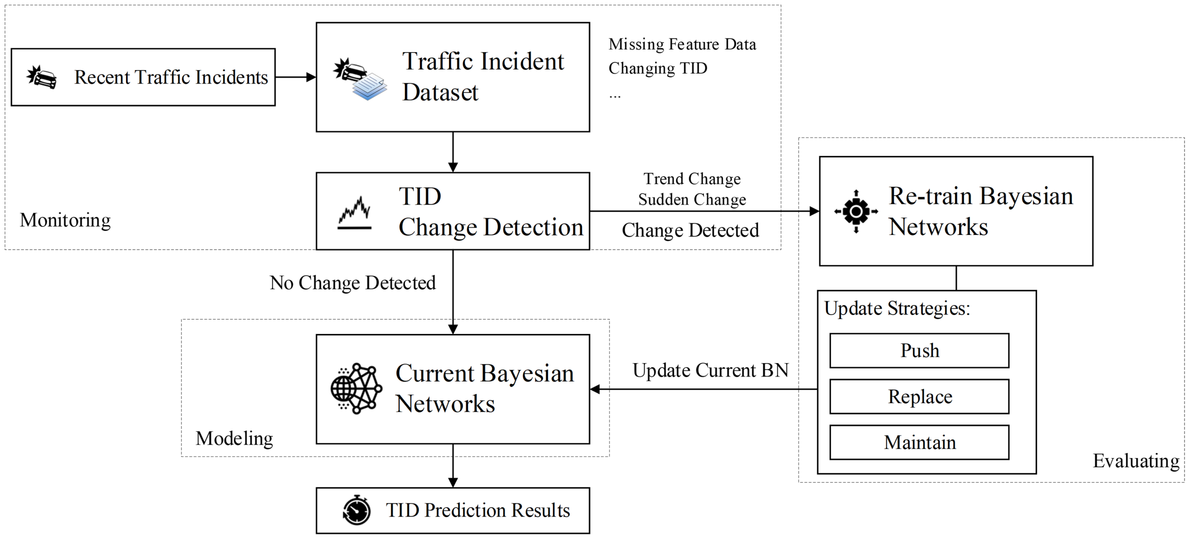


Figure 3. Workflow of proposed hybrid framework for adaptive TID prediction

the proposed update strategy.

To be specific, the objective of the CD method is to minimize the objective function $V(\tau)$ by selecting the appropriate segmentation τ. The objective function of the CD methods can be then formulated as (Truong, Oudre, and Vayatis 2020):

$$min_{\tau} V(\tau) + penalty(\tau) \tag{1}$$

The $V(\tau)$ is set to a sum of the cost function $c(\cdot)$ for a specific segmentation τ, and $penalty(\tau)$ serves as a regularization term which determines complexity of a certain segmentation. By optimizing $V(\tau)$, the CD method can provide the optimal segmentation with the corresponding timestamps.

Denote $t_k, (k = 1, 2, ..., K)$ as the detected timestamps/change points by the CD method. t_k will serve as an alarm to tell the pre-trained BNs when to re-train the models and update the learned parameters. The decision-making process will be served by the update strategy in the evaluating stage, which will be introduced in the following subsections. It is important to note that the aggregated time-series sequence of TID is utilized solely for Change Detection (CD) purposes and to determine whether to initiate the evaluating stage. Once the BNs in the modeling stage are updated, TID predictions are made for each individual traffic incident, based on their specific input features.

4.2. *Monitoring stage*

As illustrated in Fig.3, the monitoring stage continuously updates the traffic incident dataset with recent information. Every month, a CD method analyzes this dataset to identify both sudden change points (changes in mean and variance) and trend changes in the TID distribution, to accompany the findings from the previous sections (Sec.3) that the TID distributions actually experience not only the short-term drop or jump in the TIDs, but also the long-term changes residing in the TID trend. The monitoring stage uses the results from the CD to determine the subsequent steps in the workflow. If no changes are detected in the traffic incident dataset, the process directly advances to the modeling stage. However, if changes are detected, it transitions to the evaluating stage for further analysis and appropriate actions.

The CD method is composed using the search method and the cost function as mentioned in (Bian et al. 2021). The cost function monitors two consecutive intervals of a time-series data and identifies the candidate change point where those before and after intervals are significantly different. The cost function is then optimized using the search method via adjusting the penalty function.

This paper adopts a commonly used exact search method, PELT, to optimize the cost function. As compared to other exact search methods such as the segment neighborhood search method (Auger and Lawrence 1989), which has the computational complexity as ($\mathbb{O}(n^3)$), PELT maintains a relatively smaller computational complexity ($\mathbb{O}(n)$). By assuming a linear relationship between the number of change points and the size of the time-series data, PELT adopts linear penalty function to minimize $V(\tau)$ (Killick, Fearnhead, and Eckley 2012). Given the two time indexes t and s and a total of T samples, where $t < s < T$, the PELT partitions the sequence y between t and s following the rule:

$$\begin{aligned} if[min_\tau V(\tau, y_{0,t}) + \beta|\tau|] + c(y_{t,s}) >= [min_\tau V(\tau, y_{0,s}) + \beta|\tau|] holds, \\ t \ cannot \ be \ the \ last \ change \ point \ prior \ to \ T \end{aligned} \tag{2}$$

Where $V(\tau)$ is the objective function, $\beta|\tau|$ is the penalty function, $c(y_{t,s})$ is the cost function of the sub-sequence $y_{t,s}$.

With the help of PELT, this paper employs the likelihood ratio estimation as the cost function (Truong, Oudre, and Vayatis 2020), which can obtain the optimal partitioning rules and identifies the change points by comparing the piece-wise mean and variance of the corresponding intervals.

4.3. *Modeling stage*

In the modeling stage, Bayesian Networks (BNs) are utilized to predict TID. Initially, the BN model is pre-trained using early incident data (e.g., from 2015) and subsequently reassessed based on the latest CD results. If no changes are detected, the BN model retains its existing structure and parameters. Otherwise, the current BN model will be updated based on the feedback received from the evaluating stage.

The BN model used in the modeling stage is a directed acyclic graphical (DAG) model which can represent the joint probabilistic distributions of a set of random variables and their conditional dependencies. One key feature of BN is to represent the causal relationship of the random variables and predict the consequences of such a relationship. To be specific, the BN can provide the probabilistic inference of unknown values given the observations of a set of random variables. The causal relationship of these variables can be represented in the DAG, including the conditional dependencies (both local and global dependencies). The BN can then parameterize the conditional probability distributions of each node associated with the DAG.

In the context of predicting TID, BNs could be used to model the relationships between different factors that affect the TID, such as the injury severity, the number of closed lanes and other variables mentioned in Table.1. Denoting all these variables as random variables $X_i, i \in 1, 2, ..., N$, in DAG, the joint probabilistic distribution of $X_1, X_2, ..., X_N$ can be calculated based on the chain rule of probability, converting to the product of the conditional probability of X_i and its parent node:

$$P(X_1, X_2, ..., X_N) = \prod_{i=1}^{N} P(X_i|\pi(X_i)) \tag{3}$$

Where X_i is the ith node in the DAG and $\pi(X_i)$ is its parent node.

The CPD of each node in DAG can then be parameterized and estimated based on the observation variables. By marginalizing the posterior distributions across the estimated joint distribution of all nodes, the inference of dependent variable can be supplied with probabilistic scores.

The process of constructing a BN involves three major tasks: 1) creating a DAG to represent the causal relationships between variables; 2) parameterizing and estimating the conditional probability distributions of each node based on the DAG; and 3) making inferences about a dependent variable by marginalizing the posterior distributions of relevant nodes.

When the DAG is not available as prior knowledge, it can be learned directly from data through a process known as structure learning. The general idea of structure learning is to identify the best-fitting DAG by searching among all possible configurations of the network. Such a search requires:

- A *scoring function* to measure the goodness-of-fit of each candidate DAG. Commonly used scoring functions include the Bayesian Dirichlet equivalent (BDeu) (Chickering, Geiger, and Heckerman 1995), K2 (Cooper and Herskovits 1992), and the Bayesian Information Criterion (BIC) (Schwartz n.d.).
- A *search algorithm* to navigate the space of all possible DAGs and report the best-scoring one.

Mathematically, given a dataset $D = \{d_1, d_2, \ldots, d_N\}$ of N records over variables $\{X_1, \ldots, X_{N_{\text{vars}}}\}$, the scoring function $S(M; D)$ for a candidate DAG structure M often takes the form of a penalized likelihood using BIC score, which can be written:

$$S(M; D) = L(M \mid D) - \frac{\lambda}{2} \log(N) \tag{4}$$

where λ is a scaling factor and $L(M \mid D)$ is the log-likelihood:

$$L(M \mid D) = \sum_{d \in D} \log P(d \mid M) = \sum_{d \in D} \sum_{i=1}^{N_{\text{vars}}} \log P(X_i^{(d)} \mid \pi(X_i)^{(d)}, M). \tag{5}$$

Structure learning then becomes the optimization problem:

$$M^* = \arg\max_{M} S(M; D). \tag{6}$$

To find the DAG that best optimizes the chosen scoring function, various candidate search algorithms can be employed. An exhaustive search, although conceptually simple, is generally computationally infeasible for large networks due to the super-exponential growth of the search space. Greedy strategies, such as hill-climbing search, iteratively add, remove, or reverse arcs to improve the score until a local maximum is reached. Additionally, tree-based methods such as Naive Bayes, Chow-Liu (Chow and Liu 1968), and Tree-Augmented Naive Bayes (TAN) (Keogh and Pazzani 1999) treat the structure as a tree and use weighted edges to guide the structure search.

Once the structure is learned, the next step is parameter learning, where the conditional probability distributions associated with each node are estimated. This can be done using maximum likelihood estimation (MLE) or Bayesian methods (e.g., maximum a posteriori). For MLE, the parameter $\hat{\theta}$ that maximizes the likelihood is chosen as follows:

$$\hat{\theta} = \arg\max_{\theta} L(M_{\theta} \mid D). \tag{7}$$

The resulting conditional probability distributions specify how each node's probability distribution depends on its parents in the DAG. One of the key strengths of BNs is the flexibility provided by DAG structure learning. Unlike deep learning methods that often require a rigid format of input attributes, DAG structure learning allows for flexibility in the input feature set. This adaptability makes it possible for the learned DAG structure to vary and adjust to different types of traffic incidents and diverse sets of input attributes.

4.3.1. Feature selection through structure learning and parameter estimation

BNs inherently perform feature selection through the interplay of structure learning and parameter estimation. In the context of TID prediction, the structure learning phase determines which input features are directly connected to TID and which have subsequent connections, thus identifying those variables that best explain the target. By examining the presence or absence of edges between variables, the learned network structure encodes their dependencies, ensuring that only features providing meaningful information about TID remain connected. This process functions similarly to traditional feature selection. Once the structure is established, parameter estimation further refines our understanding of each selected feature's importance. Using MLE, we can derive parameters $\hat{\theta}$ that represent how variations in the parent features influence the probability distribution of TID. These resulting conditional probability distributions reveal both the magnitude and direction of each feature's effect on TID, thus quantifying the contributions of selected features.

In this study, three tree-based search algorithms, including Naive Bayes, Chow–Liu, and TAN, are evaluated simultaneously. We select the algorithm producing the highest BIC score, and adopt its learned DAG for feature selection. The final DAG thereby visualizes the dependencies and highlights the relationships among the features chosen for TID prediction.

4.3.2. Handling real-time TID prediction with partial information

BNs are effective in managing uncertainties and dependencies between variables, updating predictions as new data becomes available. This capability is essential in traffic incident management, where information of incidents are usually not available at one time but will come in sequentially. BNs are particularly suited to handling situations with incomplete or evolving data, making them highly effective for tasks like predicting traffic incident duration when all information isn't available immediately.

For example, denote input variables $X_i, i \in 1, 2, ..., N$ to include traffic incident duration (TID), number of lanes closed (L), time of the incident (T), severity of the incident (S), and presence of a heavy vehicle (H). Suppose that initially when the traffic incident is reported to traffic management center (TMC), only L and T are observed. Following Eq.3, BNs can provide an initial prediction of TID by marginalizing over the unobserved S and H. The marginal probability of TID given L and T can be written as:

$$P(TID|L,T) = \sum_S \sum_H P(TID|L,T,S,H) \cdot P(S|L,T) \cdot P(H|L,T) \tag{8}$$

Once new information S and H are observed/reported to TMC, the BNs update the existing estimate by integrating the new information without re-estimating the entire model. This updating process can be expressed as:

$$P(TID|L,T,S,H) = \frac{P(TID|L,T) \cdot P(S|L,T,TID) \cdot P(H|L,T,TID)}{P(S,H|L,T)} \tag{9}$$

The above process demonstrates the capability of Bayesian networks (BNs) to handle real-time TID prediction, particularly in situations where traffic incident information becomes available incrementally. This highlights the need for BNs that can produce reliable forecasts despite incomplete initial data, and subsequently refine their predictions as additional information arrives.

4.4. *Evaluating stage*

The evaluating stage determines the need to update the current BN model, based on a novel update strategy. The message of how to update the current BN will be provided once the evaluating stage is finished and the proper update strategy (e.g., push, replace or maintain) is selected.

The CD method detects the changes in the distribution of TIDs, the update strategy acquires the detected time-stamps and informs the BN model whether it should update the learned parameters. As mentioned above, we denote the detected timestamps from CD as $t_k, (k = 1, 2, ..., K)$, indicating the detected changes in the distribution of the TIDs. We denote t_0 as the initial timestamp where we start to evaluate the prediction results from the pre-trained BN model ($BN_{[Pre]}$). Fig.4 presents the process of the designed update strategy. The objective of the update strategy is to supply different partitions of the training data to BN models based on the proposed strategies, compare their prediction performances, and update the model parameters upon the decisions. To be specific, when a change t_k is reported by the CD method, the sequence between timestamps $[t_k - 1, t_k]$ will supply as the input of testing stage, the sequence prior to $t_k - 1$ will be partitioned based on different update strategies to re-train the current BN model:

- **Push** ($BN_{[P]}$), stores all of data up to the prior of detected timestamp t_k in the queue as training data, sequenced as $[0, t_k - 1]$, marked in red rectangle in Fig.4.
- **Replace** ($BN_{[R]}$), stores only the most recent data partitioned by the up-to-date

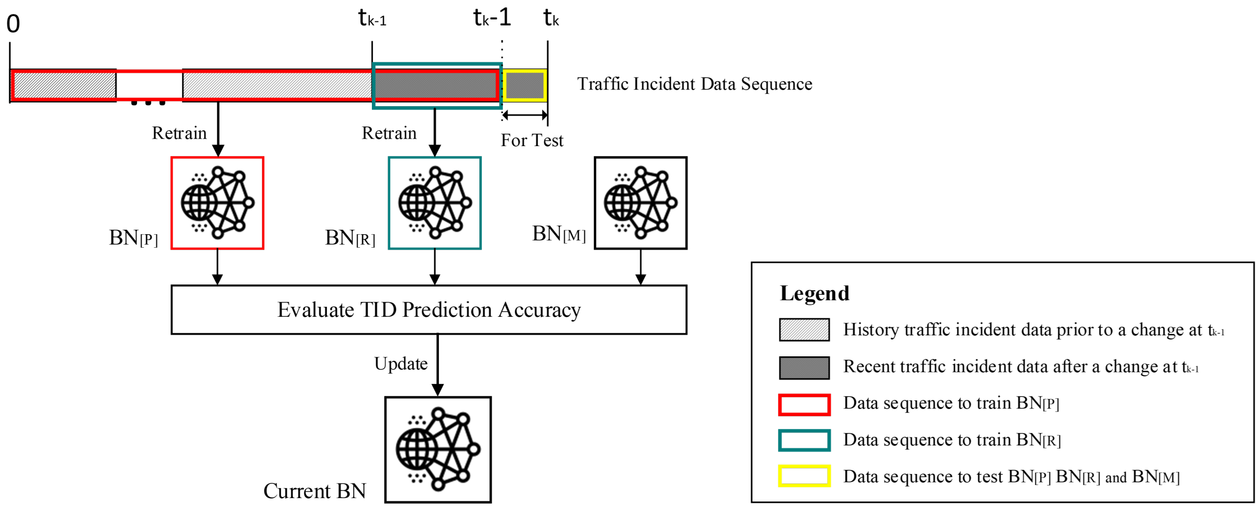


Figure 4. Update strategy for the proposed framework.

detected timestamp t_k and the previous detected timestamp t_{k-1} in the queue as training data, sequenced as $[t_{k-1}, t_k - 1]$, marked in green rectangle in Fig.4.

- **Maintain** $(BN_{[M]})$, current BN, no parameters will be updated.

The prediction performances of $BN_{[P]}, BN_{[R]}, BN_{[M]}$ will be compared using the testing data $[t_k - 1, t_k]$. The model drift occurs when the predicted accuracy of either $BN_{[P]}$ or $BN_{[R]}$ is better than $BN_{[M]}$. The model with the best-predicted accuracy in $BN_{[P]}, BN_{[R]}$ will supply as the decision-making process of how to update the current BN parameters.

4.5. *Workflow of TID prediction in traffic management*

Fig.3 illustrates the proposed workflow for predicting TID within traffic management operations. This workflow comprises three interconnected stages: monitoring, evaluating, and modeling, each serving a distinct function to ensure that TID predictions remain accurate, timely, and adaptive to evolving conditions.

In the monitoring stage, the system continuously evaluates ongoing traffic incidents for significant shifts in TID distributions. To achieve this, we employ the PELT as the CD method. Operational teams (e.g., TMCs) document each incident's TID, which is then aggregated by month. Using monthly averaged TID data, the CD algorithm identifies whether the underlying TID distribution has changed significantly over time. Whenever a substantial shift is detected, the monitoring stage triggers an alert to the evaluating stage, signaling the need to reassess and potentially update the predictive models.

Upon receiving a change alert, the evaluating stage determines whether the BNs used for TID prediction require retraining. Both the monitoring and evaluating stages function as the "back-end" processes, ensuring that the prediction models remain up-to-date and properly aligned with the most current traffic incident characteristics. By doing so, the system maintains continuous readiness to adapt its predictive capabilities to new patterns of traffic disruptions.

The modeling stage maintains the BNs, which are continuously refined based on the inputs from the monitoring and evaluating stages. These updated BNs serve as the "front-end" prediction tools that traffic operators can rely on during real-time incident management. As mentioned in Sec.4.3.2, when a traffic incident happens, not all information is immediately available. For example, the severity or exact nature of the incident may only become clear once response teams arrive on the scene. The BNs can generate initial TID predictions using the limited, early-stage data and then update these predictions as more information emerges. This flexible approach enables TMCs and other traffic operation teams to proactively adjust their management strategies, implement timely interventions, and swiftly respond to evolving incident conditions, ultimately reducing congestion and improving overall traffic flow.

5. Numerical Results

In this section, we will discuss the numerical results from our proposed adaptive framework: detected change results from CD method and the predicted TID from the BNs.

5.1. *Experimental settings*

As aforementioned in the previous section, the CD method monitors the distribution of the TIDs. This paper extracts and aggregates the TIDs on a monthly basis and adopts the CD method to detect any changes in the TID distribution. The input variables to the BN model are listed in Table.1, including a total of 10 independent variables, which further split into 35 dummy variables. Moreover, the proposed methodology requires a pre-trained BN model to serve as the base model for the evaluation and comparison purposes. In this study, we obtained and adopted the data from the year of 2015 for pre-training purposes. To be specific, the dataset in the year of 2015 is used to pre-train and test the base model. The dataset is split into 80% and 20% for training and testing purposes. Starting from the first month of 2016, the CD method starts to monitor the distribution of TIDs and the pre-trained base model will be used to compare with updated model whenever a change is detected. The comparison metrics supplied to the update strategy include mean absolute percentage error (MAPE) and root mean square error (RMSE). The formulations of MAPE and RMSE are listed as following:

$$MAPE = \sqrt{\frac{1}{n}\sum_{n=1}^{n}|\frac{Y_t - \hat{Y}_t}{Y_t}|} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(Y_t - \hat{Y}_t)^2} \tag{11}$$

Where n is the number of observations, Y_t is the actual TID at timestamp t and $\hat{Y}_t$ is the predicted TID at timestamp t.

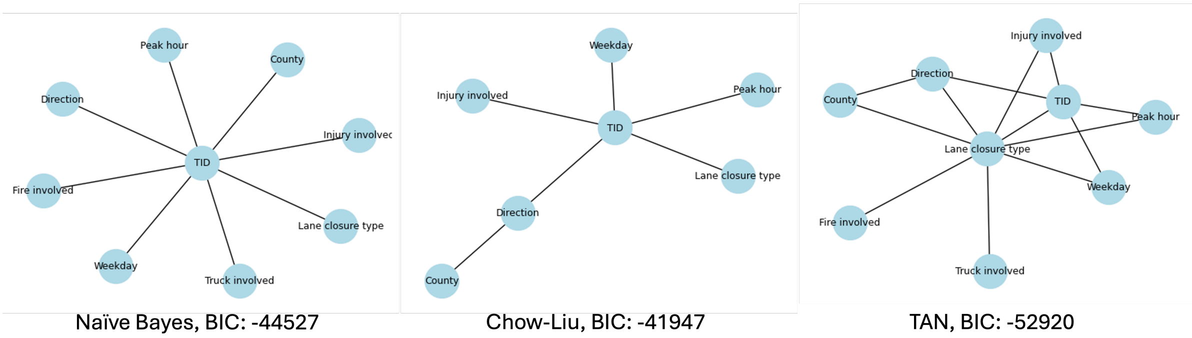


Figure 5. Comparison of BIC reported by the learned DAGs using Naive Bayes (left), Chow-Liu (mid), TAN (right).

5.2. *Feature selection analysis*

This subsection shows how the DAG learned from BNs can serve as feature selection analysis. The BN first evaluates three different structure learning algorithms, Naive Bayes, Chow–Liu, and TAN, and selects the algorithm that produces the highest BIC score. By using the pre-trained base model as an example (trained with data from 2015), we illustrate how the chosen DAG can highlight the most influential features for predicting TID.

Fig.5 shows the DAGs learned from each of the three algorithms. Among them, the Chow–Liu algorithm achieves the highest BIC score (-41947), indicating the best fit to the given dataset. In the resulting Chow–Liu DAG, four features, *Injury involved*, *Lane closure type*, *Peak hour*, and *Direction* are directly connected to TID, while *County* appears as a secondary factor linked through *Direction*. This structure implies that these four directly connected features are the most important predictors of TID, guiding our focus to severity (injury), traffic management (lane closure), temporal factors (peak hour), and spatial orientation (direction). In contrast to Naive Bayes and TAN, the Chow–Liu DAG does not include features related to heavy vehicle or fire involvement. Within the scope of the 2015 dataset, this absence suggests that these factors provide limited value for predicting TID.

In contrast, the other two algorithms include more variables in their DAGs. Naive Bayes produced a higher (less negative) BIC score than TAN (Naive Bayes: -44527 vs. TAN: -52920), connecting every feature directly to TID. On the other hand, TAN forms a hierarchical structure and also includes *Fire involved* and *Truck involved*, but its lower BIC score suggests that adding these extra features does not improve the overall predictive quality of the model.

Once the final algorithm is selected and its DAG is established, the subsequent parameter estimation step will further refine our understanding of each included feature's influence by quantifying their conditional probability distributions. It is important to emphasize that these results are based solely on the 2015 dataset and the pre-trained base model. As the training data is updated to reflect different time periods or additional records, the learned DAG may also change. Consequently, the set of selected features and their parameters will evolve to remain aligned with the most current data and the evolving characteristics of TID.

It is important to acknowledge that there are also other features, such as traffic flow, road types, weather conditions, and speed limits, that could potentially enhance the accuracy and explanatory power of the learned DAGs. For example, traffic flow and speed limit data can provide direct insights into congestion levels at the time of an incident, while road types and weather conditions could help capture the travel constraints faced by response teams. These factors can significantly influence the accessibility of response teams to the accident scene, ultimately affecting the clearance time of traffic incidents. However, due to the limit of data availability, these additional attributes were not included in this study.

5.3. *TID temporal analysis*

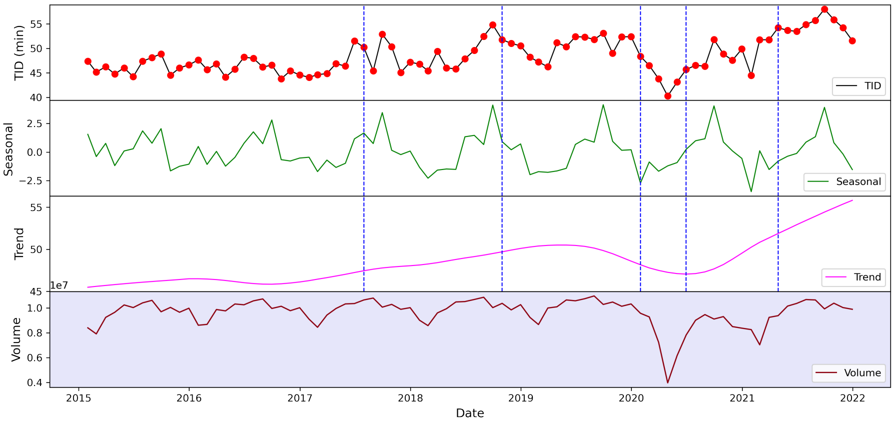
This subsection shows the detected TID results from CD method (PELT) for both crashes and disabled vehicles, respectively. As shown in Fig.6(a) and (b), the top figure shows the distribution of TIDs, the middle figure shows the extracted seasonal patterns using STL and the bottom figure shows the extracted trends via STL. A total of five change points were detected by the CD method in the distribution of TIDs for crash and disabled vehicle, respectively. The detected change points are shown as blue dashed lines in Fig.6.

For TIDs of crash, the first detected change is reported in July 2017 (the leftmost blue dashed line). This is further confirmed by the seasonal patterns. The year of 2015 and 2016 shared similar seasonal patterns, which shifted starting from the mid 2017. The trend pattern is also observed to increase starting from 2017. The years of 2017, 2018 and 2019 share the similar seasonal patterns while the trend shows a monotonous increase until its first drop in mid 2019. The CD method reported the second detected change in October 2018, at the late stage of the increase in trend. In 2020, the CD method reported two detected changes, one in February 2020 and the other in June 2020. The first detected change is a sharp drop within 5 months, caused by the occurrence of the COVID-19 pandemic and the associated Non-Pharmaceutical Interventions (NPIs) released by the government (Bian et al. 2021), while the second detected change happened during the period when public activities began to resume and the traffic started to recover, which can be confirmed by the rebound of the trend pattern.

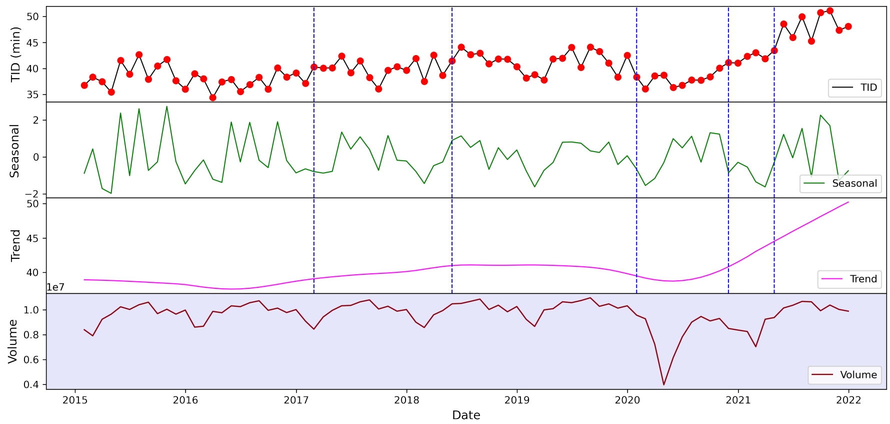
Similar findings are also observed in TIDs of disabled vehicles, as shown in Fig.6(b). The TIDs of disabled vehicles also share the similar seasonal patterns in the year of 2015 and 2016, and started to change from the beginning of 2017. The CD method detected this change and reported its first detection in February 2017. Analogous to the detection results in TIDs of crash, the CD method also reported two changes in the year of 2020, indicating the significant changes occurred due to the COVID-19 pandemic.

Illustrated at the bottom of Fig.6 (a) and (b), we compared the vehicular traffic volume with TID distribution and its decomposed patterns. The vehicular traffic volume data is obtained from New York State's Open Data Platform nys (n.d.). The traffic volume data consists of the hourly traffic volume of all vehicular traffic throughput each of the nine bridges and tunnels, both inbound and outbound directions to New York City. The traffic volume data is further aggregated on a monthly basis and in pairwise correlation with TID distribution. It is clearly shown that the seasonal patterns of TIDs follow the similar seasonal patterns of traffic volume, implying that the TID is positively correlated with multiple factors in the transportation system and therefore shares similar temporal characteristics.

Furthermore, during special circumstances such as the COVID-19 pandemic, the TIDs showed similar sharp drops to the traffic volume starting March 2020, a reason-



(a) Crash



(b) Disabled Vehicle

Figure 6. Change detection results for crash and disabled vehicle, blue dashed lines represent CD detected results.

able shift since the less congested roadway can help emergency response team to move and clear the incident from the scene. Moreover, an interesting finding shows that the TIDs increased immediately after the sudden drop while the traffic volume has not fully recovered to its normal level. A potential cause could be the shortage of labors in emergency response team (e.g., police, fire department and EMS). Between 2020 and 2021 (FOX 5 NY Staff 2021), a high ratio of staffing shortage was reported by NYPD as 21%, FDNY as 17% and EMS as 30% due to the medical leave caused by COVID-19. Such labor shortages may increase the response time of a traffic accident and therefore cause an increase in TID. This finding also implies our assumptions that the TID is affected by multiple factors in the transportation system and may evolve and change in different temporal dimensions.

Therefore, it is necessary for the TID prediction model to capture these dynamics and make corresponding adaptations, which will be discussed in the following sections.

5.4. *Model evaluation*

In this subsection, we conduct an ablation experiment to evaluate the effect of our proposed adaptive module in terms of the prediction accuracy. We compare the accuracy of BNs with and without the CD method and the update strategy. The BN without the update strategy is denoted as a 'One-time' model while we denote the proposed model as 'CPD-adapt'. The details of each model are as follows:

- **One-time**: the BN model that is trained and validated using the data in 2015, the model is only trained once and therefore the learned parameters will not change. The training and test dataset contains a total of 6217 traffic incidents.
- **CPD-adapt**: our proposed model, the model will adopt one-time as the base model and will re-train itself when there is a change detected by the CD method and adopt the update strategy to decide whether to update the parameters and use either push or replace strategy to update the parameters.

To be specific, we calculated the MAPE and RMSE using the data from 2017 to 2021. It is important to note that the 'One-time' model was trained and tested using data from 2015, which includes a total of 6,217 traffic incidents. In contrast, the training and testing datasets for the 'CPD-adapt' model varied, depending on the period between two detected changes and the specific update strategy implemented. For instance, the first update for 'CPD-adapt' occurred in July 2017. During this update, the model was re-trained using a dataset containing a total of 3,637 incidents. As illustrated in Table.2, the MAPE and RMSE are reported based on five different TID groups of crash and disabled vehicle, respectively.

Overall, we observe that the 'CPD-adapt' outperforms the 'One-time' model across 5 TID groups in both MAPE and RMSE. An improvement with over 10% decrease from 'One-time' model to 'CPD-adapt' model is observed in MAPE for the group of 0-30 min, for both crash and disabled vehicles, respectively. The second largest improvement is shown in the group of 30-60 min, where 'CPD-adapt' reports 20.9% for crash and 18.5% for disabled vehicles, while 'One-time' model reports 23.41% and 21.94%, respectively.

For 'CPD-adapt' itself, the best prediction accuracy in terms of MAPE is demonstrated across larger TID groups in 60-90, 90-120 and larger than 120 min, of which groups have higher tolerance in larger MAPE and RMSE, as compared to groups of small TIDs. The best MAPE is illustrated as 5.7% for the group of 120 and larger and the best RMSE is demonstrated as 8.3 min for the group of 0-30 min for disabled vehicle. However, the MAPE for the group of 0-30 min is reported as 50.1%, shown as deficient accuracy. The same large MAPE is also observed for the group of 0-30 min for crash TIDs. The main reason could be the scale of the duration since the MAPE mainly describes the relative bias to its absolute value. Moreover, small denominator in MAPE could enlarge the errors between the predicted and actual values. The RMSE values of group 0-30 min for both crash and disabled vehicle are 9.13 min and 8.3 min, which are acceptable since the granularity of the TIDs is relatively small.

5.5. *Adaptability analysis*

In addition to evaluating the effect of adaptation module in terms of the model prediction accuracy, it is important to evaluate the capability of adapting to the up-to-date situation without the loss of prediction accuracy. To be specific, we designed several performance measures to evaluate the adaptability of our proposed model: 1) the

Table 2. Comparison of prediction performance on different TID groups

TID group (min)	Crash				Disabled Vehicle			
	One-time		CPD-adapt		One-time		CPD-adapt	
	MAPE(%)	RMSE(min)	MAPE(%)	RMSE(min)	MAPE(%)	RMSE(min)	MAPE(%)	RMSE(min)
0-30	64.97	10.05	53.5	9.13	60.10	9.89	50.10	8.30
30-60	23.41	12.26	20.9	11.34	21.94	12.02	18.50	11.09
60-90	13.51	12.05	12.69	11.81	13.19	12.48	11.54	11.29
90-120	9.53	12.18	7.43	9.11	10.84	13.12	9.00	11.93
≥120	8.64	13.83	5.87	9.20	7.15	11.55	5.70	8.36

number of re-trainings of the model, 2) the number of updates of the model, and 3) the prediction accuracy. It is important to note that the number of model updates does not have to be identical with the number of re-trainings. When the model is re-trained, the prediction performances before and after the re-training are compared. If the prediction performance does not see an increase, then parameters are not updated and the model thereby remains unchanged as before. Ideally, a desired adaptive model should provide satisfactory prediction accuracy without requiring a large number of re-trains and updates the modeling parameters. To accomplish this, we evaluate our 'CPD-adapt' against two baseline BN models governed by different update rules: 'Fixed-year' model and 'Acc-adapt' model. (Demiroluk and Ozbay 2014):

- **Fixed-year**: after the initial pre-training phase, the model will receive manual updates on an annual basis. Specifically, at the end of each year, the BN model will be updated using data from the current year. Then the updated model will be used to predict TID for the forthcoming year.
- **Acc-adapt**: the BN model acquired from (Demiroluk and Ozbay 2014), this model adopts the update strategy that is driven by accuracy: it will re-train the model every month and compare the prediction accuracy with the current model, if the accuracy is better, then the learned parameters will be updated.

The training and testing datasets for the 'Acc-adapt' model are determined by the monthly number of traffic incidents, with the distribution detailed in Fig.2 (e). The 'Fixed-year' model's training and testing datasets are based on the annual number of traffic incidents. The size of these datasets can also be referenced from the aggregation of monthly numbers in Fig.2 (e).

We evaluate the performance of adaptability starting from the CD's first detected result. The comparison results across 'One-time', 'Fixed-year', 'Acc-adapt' and 'CPD-adapt' models are illustrated in Fig.7 and Table.3. As shown in Fig.7, the green curve represents the MAPE results by each month for the 'one-time' model. Setting the 'One-time' model as the base model, we compare the MAPE results of 'CPD-adapt', 'Acc-adapt' and 'Fixed-year' and calculate the difference in their MAPE results from the base model. The difference between the base model and the adaptive model is illustrated using the metric of %Decrease, which is the percentage of MAPE improvement as compared to 'one-time' model, showing in the dark red bar for 'Acc-adapt', light red bar for 'CPD-adapt', and the black bar for 'Fixed-year'. It can be noted from Fig.7 and Table.3 that the proposed 'CPD-adapt' model surpasses both the 'Fixed-year' model and the 'Acc-adapt' model in terms of MAPE throughout the entire study period and demonstrates the highest MAPE improvement over 'One-time' model in each month.

The proposed 'CPD-adapt' model required significantly fewer re-trainings and updates compared to the 'Acc-adapt' model. In Figure 7, the pink up-pointing triangle ($\triangle$), the blue down-pointing triangle ($\triangledown$), and the black star ($*$) denote the update timestamps for the 'CPD-adapt' model, the 'Acc-adapt' model, and the 'Fixed-year'

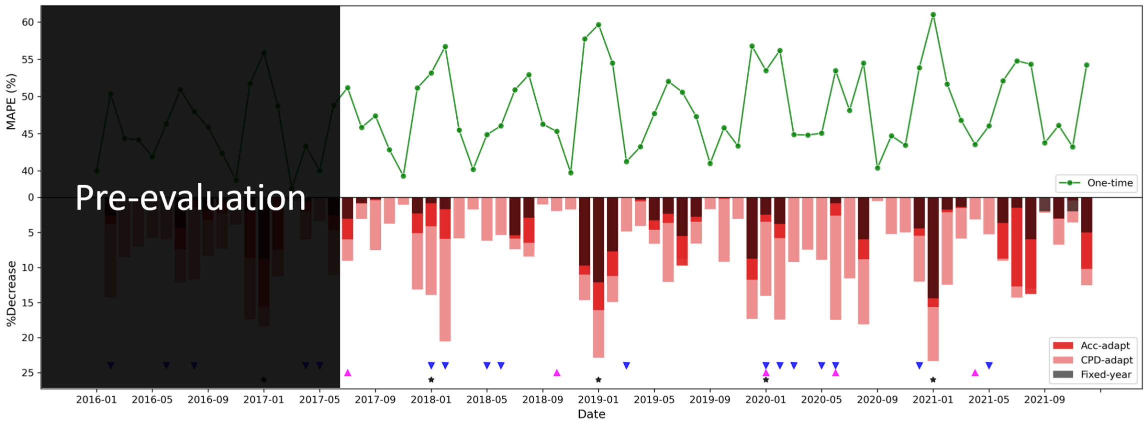


Figure 7. Adaptability comparison in MAPE for crash TIDs.

Table 3. Comparison of adaptability across models using crash data

Model	Interval	Date range	MAPE	%Decrease range	Avg. %Decrease	#Month	#Re-train	#Update
One-time	$[t_1,t_2)$	Jul.2017-Oct.2018	47.61	-	-	-	-	-
	$[t_2,t_3)$	Oct.2018-Jan.2020	48.4	-	-	-	-	-
	$[t_3,t_4)$	Jan.2020-Jun.2020	48.88	-	-	-	-	-
	$[t_4,t_5)$	Jun.2020-Apr.2021	49.8	-	-	-	-	-
	$[t_5,T_{max})$	Apr.2021-Dec.2021	48.67	-	-	-	-	-
	Overall	July.2017-Dec.2021	48.53	-	-	-	-	-
Fixed-year	$[t_1,t_2)$	Jul.2017-Oct.2018	47.5	0-5.41	1.16	15	1	1
	$[t_2,t_3)$	Oct.2018-Jan.2020	45.6	0-12.12	3.53	15	1	1
	$[t_3,t_4)$	Jan.2020-Jun.2020	49.91	0-3.79	1.25	5	1	1
	$[t_4,t_5)$	Jun.2020-Apr.2021	47.5	0-14.4	2.89	10	1	1
	$[t_5,T_{max})$	Apr.2021-Dec.2021	46.69	0-6.01	2.27	9	0	0
	Overall	July.2017-Dec.2021	47.53	0-12.12	2.33	54	4	4
Acc-adapt	$[t_1,t_2)$	Jul.2017-Oct.2018	47.49	0-6.49	2.3	15	15	4
	$[t_2,t_3)$	Oct.2018-Jan.2020	44.29	0-16.08	4.84	15	15	1
	$[t_3,t_4)$	Jan.2020-Jun.2020	49.3	0-5.82	1.86	5	5	4
	$[t_4,t_5)$	Jun.2020-Apr.2021	47.46	0-15.65	3.62	10	10	2
	$[t_5,T_{max})$	Apr.2021-Dec.2021	43.89	0-13.78	5.45	9	9	1
	Overall	July.2017-Dec.2021	46.16	0-16.08	3.6	54	54	12
CPD-adapt	$[t_1,t_2)$	Jul.2017-Oct.2018	40.39	1.4-20.5	7.21	15	1	1
	$[t_2,t_3)$	Oct.2018-Jan.2020	39.69	1.7-22.9	8.7	15	1	1
	$[t_3,t_4)$	Jan.2020-Jun.2020	37.45	7.5-17.5	11.43	5	1	1
	$[t_4,t_5)$	Jun.2020-Apr.2021	38.61	0.5-23.4	11.18	10	1	1
	$[t_5,T_{max})$	Apr.2021-Dec.2021	40.91	2.2-14.3	7.8	9	1	1
	Overall	July.2017-Dec.2021	39	0.5-23.4	8.84	54	5	5

model, respectively. Notably, the 'CPD-adapt' was updated only 5 times during the study period, which is substantially less than the 12 updates performed on the 'Acc-adapt' model.. The same results can be confirmed in Table.3, where it shows that 'Acc-adapt' model re-trained itself every month (54 times in total) and achieved the range of %Decrease in 0-16%. The 'CPD-model' re-trained itself at each detected change points (10 times in total), achieving both the best MAPE and %Decease. The 'Fixed-year' achieved the range of %Decrease in 0-12.12% and the average MAPE as 47.53%, lower than 'One-time' but higher than 'Acc-adapt'.

It is also possible to observe that both 'Acc-adapt' and 'CPD-adapt' model react intensively to the COVID-19 pandemic, shown in intervals of $[t_3, t_4)$ and $[t_4, t_5)$: 'Acc-adapt' updated its learned parameters 6 out of 12 times in order to adapt to the significant variants and sudden changes, while 'CPD-adapt' updated 2 out of 5 times to keep itself up-to-date. Both 'Acc-adapt' and 'CPD-adapt' model achieved the improvement in %Decrease of MAPE as compared to the 'One-time' and 'Fixed-year' model. However, the effort of accomplishing the adaptability varies.

Despite the fact that 'Acc-adapt' uses an exhaustive updating strategy and re-trains the model every month, the MAPE of each month is still higher than the 'CPD-adapt'. One potential reason could be due to the granularity and scale of the size of training data. The TIDs have noticeable seasonal patterns in New York City, as shown in Fig.6. Therefore, a model lacking a mechanism to adjust the temporal scale of training data in response to the change of TID, such as Acc-adapt which only uses the most recent data for re-training, may be insufficient to learn the complete causal relationship of the input variables and thus may not provide satisfactory prediction results. On the contrary, 'CPD-adapt' responds to the detected changes from CD method, of which we showed in Fig.6 and could capture the changes in both seasonality and trend, as well as sudden changes. Furthermore, the strategy of push and replace could ensure the model to learn the complete causal structure and avoid the bias caused by the partial-fitting problem.

Unlike 'Acc-adapt,' which employs an exhaustive updating strategy, the 'Fixed-year' model only updates its parameters annually at the end of each year. While the MAPE for each month in the 'Fixed-year' model is lower than that of the 'one-time' model, it is consistently higher than both 'Acc-adapt' and 'CPD-adapt'. This performance discrepancy indicates that an annual update strategy is insufficient for addressing the dynamic nature of Traffic Incident Durations (TIDs), which are affected by varying seasonal trends and disruptive events like the COVID-19 pandemic. This observation highlights the need for adopting an adaptive updating approach, or at least a more frequent updating strategy, rather than relying on annual updates. The ability to adapt more dynamically to changes ensures that the model remains effective in capturing the complexities and fluctuations of TID behavior over time.

In summary, we evaluate the adaptability of our proposed model in terms of its prediction accuracy, number of re-trainings and number of parameter updates. We then compared our model with the 'One-time' model and 'Acc-adapt' model. Using the MAPE distribution and timestamps of parameter updates, the results demonstrated the power of the adaptability of our model by showing that the model can achieve the best prediction accuracy without frequently re-training and updating its learned parameters as well as overall modeling structure.

5.6. *Implication of implementation in traffic incident management*

The effectiveness of modern traffic incident management systems hinges on their adaptability and resilience, particularly in terms of Traffic Incident Duration (TID) prediction methods. The rapid advancement of traffic sensing technologies has accelerated data acquisition processes, thereby providing an extensive amount of traffic incident information.

Our framework integrates change detection and Bayesian network models to enhance prediction capabilities amidst continuously evolving traffic incident information and environmental conditions. The design of our model, which is both adaptive and self-updating, is specifically aligned with the rapid advancements in traffic sensing technologies. These technologies are increasingly capable of providing a broad spectrum of traffic incident data, making adaptability essential for effectively utilizing comprehensive data from diverse platforms. This ensures that the model can respond appropriately to new information without the need for frequent manual re-calibrations.

The ability of our model to continuously adjust to new data ensures its effectiveness and resilience even in rapidly changing urban environments. This adaptability has the potential to lower operational costs and reduce the workload for traffic management personnel, resulting in long-term resource savings. As a result, traffic management can prioritize other critical areas of operation. Our model requires fewer updates compared to other benchmarks, proving particularly valuable in environments where traffic conditions can change swiftly, such as those observed post-COVID-19 pandemic.

Furthermore, several practical applications can be derived from our proposed model, enhancing its utility in real-world settings. The ability of the model to analyze how TID evolves over long-term periods enables a deeper understanding of traffic patterns and peak incident seasons. By identifying trends and peak periods of incidents, agencies can strategically deploy resources when and where they are most needed, improving response times and overall traffic flow efficiency.

Additionally, the integration of TID predictions with traffic forecasting models allows management agencies to anticipate traffic flow during incidents more accurately. This predictive insight is crucial for planning detours and managing traffic around incident hotspots effectively. It ensures that traffic can be redirected in a manner that minimizes congestion and reduces the impact of incidents on overall traffic conditions.

6. Conclusion

There are understudied aspects of predicting TIDs over long time periods without re-training. These temporal stability aspects of the TID prediction have been introduced for the first time by our study to provide insights for traffic researchers to re-evaluate the change in the predictive power of their models in the presence of temporal changes.

In this paper, we developed a hybrid framework which combines the CD method with BN models to ensure that the predictive results remain stable over time by adapting to ever changing conditions in a transportation network. A descriptive analysis was conducted to analyze whether the TIDs evolve and vary in a long-term dimension over seasons and years. The decomposed temporal patterns found significant seasonal and trend patterns and the transitions/changes were observed in these patterns in the aftermath of different seasons and years. Sudden interruptions to transportation systems, such as the COVID-19 pandemic, also caused significant disruptions to the TIDs. With the use of PELT as the main CD method, we demonstrated that the detected changes captured these shifts both terms of slower moving temporal patterns as well as sudden disruptions. To evaluate the adaptability of our proposed model, we designed several measures to illustrate its ability to adapt to the most current conditions and provide satisfactory prediction results by only re-training and updating the models when a change is detected, which saves significant efforts of frequent, repeated re-training of the parameters in an exhaustive manner (e.g., periodically re-training every month).

There are several limitations in this study, which could be improved as the next steps in the future: 1) due to the limitations of the data availability, the proposed model is estimated with a limited number of input variables which limited its prediction accuracy, especially for the small TID group (0-30 min); and 2) this study only focused on the long-term development of the TID without assessing the short-term effects due to detection, response and clearance stages for each individual incident. In the future, the research team will consider both the long-term and short-term adaptability in the aspects of prediction of TIDs need to be considered and the interactions between the short-term and long-term development of a traffic incident duration prediction model can be evaluated.

Acknowledgements

This work was supported by the C2SMARTER, a Tier 1 U.S. Department of Transportation (USDOT) funded University Transportation Center (UTC) led by New York University funded by USDOT. The contents of this paper only reflect the views of the authors who are responsible for the facts and do not represent any official views of any sponsoring organizations or agencies.

References

n.d. "New York's Open Data Portal, 2021 Hourly Traffic on Metropolitan Transportation Authority (MTA) Bridges and Tunnels: Beginning 2010." https://data.ny.gov/Transportation/Hourly-Traffic-on-Metropolitan-Transportation-Auth/qzve-kjga. Accessed: 2022-01-30.

Alkaabi, Abdulla Mohammed Saeed, Dilum Dissanayake, and Roger Bird. 2011. "Analyzing clearance time of urban traffic accidents in Abu Dhabi, United Arab Emirates, with hazard-based duration modeling method." *Transportation Research Record* 2229 (1): 46–54.

Amin-Naseri, Mostafa, Pranamesh Chakraborty, Anuj Sharma, Stephen B Gilbert, and Mingyi Hong. 2018. "Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze." *Transportation research record* 2672 (43): 34–43.

Auger, Ivan E, and Charles E Lawrence. 1989. "Algorithms for the optimal identification of segment neighborhoods." *Bulletin of mathematical biology* 51 (1): 39–54.

Baena-Garcıa, Manuel, José del Campo-Ávila, Raúl Fidalgo, Albert Bifet, R Gavalda, and Rafael Morales-Bueno. 2006. "Early drift detection method." In *Fourth international workshop on knowledge discovery from data streams*, Vol. 6, 77–86.

Bai, Jushan, and Pierre Perron. 2003. "Computation and analysis of multiple structural change models." *Journal of applied econometrics* 18 (1): 1–22.

Bian, Zilin, Fan Zuo, Jingqin Gao, Yanyan Chen, Sai Sarath Chandra Pavuluri Venkata, Suzana Duran Bernardes, Kaan Ozbay, Xuegang Jeff Ban, and Jingxing Wang. 2021. "Time lag effects of COVID-19 policies on transportation systems: A comparative study of New York City and Seattle." *Transportation Research Part A: Policy and Practice* 145: 269–283.

Chickering, Max, Dan Geiger, and David Heckerman. 1995. "Learning Bayesian networks: Search methods and experimental results." In *Proceedings of the fifth international workshop on artificial intelligence and statistics*, .

Chow, CKCN, and Cong Liu. 1968. "Approximating discrete probability distributions with dependence trees." *IEEE transactions on Information Theory* 14 (3): 462–467.

Chung, Younshik. 2010. "Development of an accident duration prediction model on the Korean Freeway Systems." *Accident Analysis & Prevention* 42 (1): 282–289.

Chung, Younshik, and Byoung-Jo Yoon. 2012. "Analytical method to estimate accident duration using archived speed profile and its statistical analysis." *KSCE Journal of Civil Engineering* 16 (6): 1064–1070.

Cleveland, Robert B, William S Cleveland, Jean E McRae, and Irma Terpenning. 1990. "STL: A seasonal-trend decomposition." *J. Off. Stat* 6 (1): 3–73.

Cooper, Gregory F, and Edward Herskovits. 1992. "A Bayesian method for the induction of probabilistic networks from data." *Machine learning* 9 (4): 309–347.

Demiroluk, Sami, and Kaan Ozbay. 2014. "Adaptive learning in bayesian networks for incident duration prediction." *Transportation Research Record* 2460 (1): 77–85.

Dimitriou, Loukas, and Eleni I Vlahogianni. 2015. "Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions." *Analytic Methods in Accident Research* 5: 59–71.

FOX 5 NY Staff, X. 2021. "Major staffing shortages at NYPD, FDNY due to COVID surge." *FOX 5 NY* Accessed 2021-12-30. https://www.fox5ny.com/news/highest-rate-of-sick-nypd-officers-hes-ever-seen-says-outgoing-commissioner-shea.

Garib, A, AE Radwan, and HJJoTE Al-Deek. 1997. "Estimating magnitude and duration of incident delays." *Journal of Transportation Engineering* 123 (6): 459–466.

Golob, Thomas F, Wilfred W Recker, and John D Leonard. 1987. "An analysis of the severity and incident duration of truck-involved freeway accidents." *Accident Analysis & Prevention* 19 (5): 375–395.

Harrou, Fouzi, Ying Sun, and Muddu Madakyaru. 2016. "Kullback-leibler distance-based enhanced detection of incipient anomalies." *Journal of Loss Prevention in the Process Industries* 44: 73–87.

Hassani, Marwan. 2019. "Concept Drift Detection Of Event Streams Using An Adaptive Window." In *ECMS*, 230–239.

He, Qing, Yiannis Kamarianakis, Klayut Jintanakul, and Laura Wynter. 2013. "Incident duration prediction with hybrid tree-based quantile regression." In *Advances in dynamic network modeling in complex transportation systems*, 287–305. Springer.

Keogh, Eamonn J, and Michael J Pazzani. 1999. "Learning augmented Bayesian classifiers: A comparison of distribution-based and classification-based approaches." In *AIStats*, .

Khattak, Asad J, Jun Liu, Behram Wali, Xiaobing Li, and ManWo Ng. 2016. "Modeling traffic incident duration using quantile regression." *Transportation Research Record* 2554 (1): 139–148.

Khattak, Asad J, Joseph L Schofer, and Mu-Han Wang. 1995. "A simple time sequential procedure for predicting freeway incident duration." *Journal of Intelligent Transportation Systems* 2 (2): 113–138.

Killick, Rebecca, Paul Fearnhead, and Idris A Eckley. 2012. "Optimal detection of changepoints with a linear computational cost." *Journal of the American Statistical Association* 107 (500): 1590–1598.

Knibbe, Willem Jan J, Tom P Alkim, Jon FW Otten, and Moses Y Aidoo. 2006. "Automated estimation of incident duration on Dutch highways." In *2006 IEEE Intelligent Transportation Systems Conference*, 870–874. IEEE.

Kuang, Li, Han Yan, Yujia Zhu, Shenmei Tu, and Xiaoliang Fan. 2019. "Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor." *Journal of Intelligent Transportation Systems* 23 (2): 161–174.

Li, Ruimin, Francisco C Pereira, and Moshe E Ben-Akiva. 2018. "Overview of traffic incident duration analysis and prediction." *European transport research review* 10 (2): 1–13.

Lin, Lei, Qian Wang, and Adel W Sadek. 2016. "A combined M5P tree and hazard-based duration model for predicting urban freeway traffic accident durations." *Accident Analysis & Prevention* 91: 114–126.

Mohanty, SR, AK Pradhan, and A Routray. 2007. "A cumulative sum-based fault detector for power system relaying application." *IEEE transactions on power delivery* 23 (1): 79–86.

NHTSA, X. 2022. "Newly Released Estimates Show Traffic Fatalities Reached a 16-Year High in 2021." *NHTSA* Accessed 2022-5-17. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813298.

Owens, Nicholas D, April H Armstrong, Carol Mitchell, Rebecca Brewster, et al. 2009. *Federal Highway Administration Focus States Initiative: traffic incident management performance measures final report.* Technical Report. United States. Federal Highway Administration.

Ozbay, Kaan, and Pushkin Kachroo. 1999. "Incident management in intelligent transportation systems." .

Ozbay, Kaan, and Nebahat Noyan. 2006. "Estimation of incident clearance times using

Bayesian Networks approach." *Accident Analysis & Prevention* 38 (3): 542–555.

Peeta, Srinivas, Jorge L Ramos, and Shyam Gedela. 2000. "Providing real-time traffic advisory and route guidance to manage Borman incidents on-line using the Hoosier helper program." .

Qi, Yi, and Hualiang Teng. 2008. "An information-based time sequential approach to online incident duration prediction." *Journal of Intelligent Transportation Systems* 12 (1): 1–12.

Reed, Trevor. 2019. "Inrix global traffic scorecard." .

Salem, Osman, Farid Naït-Abdesselam, and Ahmed Mehaoua. 2012. "Anomaly detection in network traffic using Jensen-Shannon divergence." In *2012 IEEE International Conference on Communications (ICC)*, 5200–5204. IEEE.

Schwartz, G. n.d. "Estimating the dimension of a model (1978) The Annals of Statistics, 6." *MR0468014* 461–464.

Scott, Andrew Jhon, and Martin Knott. 1974. "A cluster analysis method for grouping means in the analysis of variance." *Biometrics* 507–512.

Truong, Charles, Laurent Oudre, and Nicolas Vayatis. 2020. "Selective review of offline change point detection methods." *Signal Processing* 167: 107299.

Wei, Chien-Hung, and Ying Lee. 2007. "Sequential forecast of incident duration using artificial neural network models." *Accident Analysis & Prevention* 39 (5): 944–954.

Yu, B, YT Wangy, JB Yaoz, and JY Wang. 2016. "A COMPARISON OF THE PERFORMANCE OF ANN AND SVM FOR THE PREDICTION OF TRAFFIC ACCIDENT DURATION." *Neural Network World* 26 (3).

Zhan, Chengjun, Albert Gan, and Mohammed Hadi. 2011. "Prediction of lane clearance time of freeway incidents using the M5P tree algorithm." *IEEE Transactions on Intelligent Transportation Systems* 12 (4): 1549–1557.

Žliobaitė, Indrė. 2010. "Learning under concept drift: an overview." *arXiv preprint arXiv:1010.4784* .

Zuo, Fan, Jingqin Gao, Kaan Ozbay, Zilin Bian, and Daniel Zhang. 2023. "Urban Work Zone Detection and Sizing: A Data-Centric Training and Topology-Based Inference Approach." In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, 3235–3240. IEEE.