REVOLUTIONIZE TRANSPORTATION MANAGEMENT

MINDSET WITH DATA-DRIVEN ANALYTICS

by

Zhuo Chen

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Civil and Environmental Engineering

The University of Utah

August 2017

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Zhuo Chen

has been approved by the following supervisory committee members:

Xiaoyue Cathy Liu , Chair 06/09/2017
Date Approved

Richard J. Porter , Member 06/13/2017
Date Approved

Steven Burian , Member 06/02/2017
Date Approved

Juan Medina , Member 06/14/2017
Date Approved

Ran Wei , Member 06/12/2017
Date Approved

and by Michael E. Barber , Chair/Dean of

the Department/College/School of Civil and Environmental Engineering

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Data-driven analytics has been successfully utilized in many experience-oriented areas, such as education, business, and medicine. With the profusion of traffic-related data from Internet of Things and development of data mining techniques, data-driven analytics is becoming increasingly popular in the transportation industry. The objective of this research is to explore the application of data-driven analytics in transportation research to improve traffic management and operations. Three problems in the respective areas of transportation planning, traffic operation, and maintenance management have been addressed in this research: exploring the impact of dynamic ridesharing system in a multimodal network, quantifying nonrecurrent congestion impact on freeway corridors, and developing an infrastructure sampling method for efficient maintenance activities.

First, the impact of dynamic ridesharing in a multimodal network is studied with agent-based modeling. The competing mechanism between dynamic ridesharing system and public transit is analyzed. The model simulates the interaction between travelers and the environment and emulates travelers' decision making process with the presence of competing modes. The model is applicable to networks with varying demographics.

Second, a systematic approach is proposed to quantify Incident-Induced Delay on freeway corridors. There are two particular highlights in the study of nonrecurrent congestion quantification: secondary incident identification and K-Nearest Neighbor pattern matching. The proposed methodology is easily transferable to any traffic operation system that has access to sensor data at a corridor level.

Lastly, a high-dimensional clustering-based stratified sampling method is developed for infrastructure sampling. The stratification process consists of two components: current condition estimation and high-dimensional cluster analysis. High-dimensional cluster analysis employs Locality-Sensitive Hashing algorithm and spectral sampling. The proposed method is a potentially useful tool for agencies to effectively conduct infrastructure inspection and can be easily adopted for choosing samples containing multiple features.

These three examples showcase the application of data-driven analytics in transportation research, which can potentially transform the traffic management mindset into a model of data-driven, sensing, and smart urban systems. The analytical approach presented will inform evidence-based and data-driven decision making in transportation policy and investment choices.

To my family and friends

TABLE OF CONTENTS

Chapters

LIST OF TABLES

Tables

LIST OF FIGURES

Figures

ACRONYMS

MAP-21	The Moving Ahead for Progress in the 21 Century Act
DOT	Department of Transportation
IoT	Internet of Things
HOV	High Occupancy Vehicle
HOT	High Occupancy Toll
ITS	Intelligent Transportation System
TIM	Traffic Incident Management
IID	Incident-Induced Delay
AADT	Annual Average Daily Traffic
SRS	Simple Random Sampling
LOM	Level-of-Maintenance
LSH	Locality-Sensitive Hashing
ABM	Agent-Based Modeling
VHT	Vehicle Hour Traveled
DQT	Deterministic Queuing Theory
TOD	Time-of-Day
DOW	Day-of-Week
PeMS	Freeway Performance Measurement System
LMDP	Latent Markov Decision Process

KNN	K-Nearest Neighbor
M&R	Maintenance and Rehabilitation
MMQA	Maintenance Management Quality Assurance
RMSE	Root-Mean-Squared Error
HDCSS	High-Dimensional Clustering-based Stratified Sampling

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my dissertation committee chair and advisor, Dr. Xiaoyue Cathy Liu, for her patient and insightful guidance over the years. I could not have completed this work without her support and mentorship. My thanks also go to my dissertation committee members, Dr. Richard J. Porter, Dr. Juan C. Medina, Dr. Steven Burian, and Dr. Ran Wei, for their timely feedback and enlightening suggestions. I am truly grateful for the time spent with my fellow graduate students in the Utah Traffic Lab, who made my graduate study a very entertaining and unique experience. Finally, I would like to thank my parents, who taught me the value of hard work and independence. I would not be who I am today without their unconditional love and support over the years.

CHAPTER 1

INTRODUCTION

Data-driven analytics, which has been utilized in many experience-oriented areas, such as education, business, and medicine, is becoming increasingly popular in the transportation industry. Over the years, many state Department of Transportation (DOTs) and regional transportation agencies have been looking into adopting data-driven analytics into their business processes, especially in traffic operation and transportation infrastructure investments. In July 2012, the Moving Ahead for Progress in the 21th Century Act (MAP-21) was signed to formally embrace performance-based planning and data-driven decision-making as a national policy. It encourages transportation agencies to conduct decision-making based on data-driven analytics in order to increase the accountability of federal highway programs *(1)*.

Another factor that triggered the blast of data-driven analytics is the intensive influx of transportation data resulting from technological advancement. Particularly, the profusion of data from Internet of Things (IoT) presents unprecedented opportunities for creating a cohesive and seamless integration of urban transportation and technology. Massive data collected from various mobile sources and advanced sensors provide transportation researchers endless possibilities for making interconnected knowledge discovery Together with the explosion of transportation data, data mining techniques, such as clustering analysis, machine learning, and dimensionality reduction, have been improved and become accessible to transportation practitioners.

With data mining techniques, researchers and engineers make decisions based on the data-driven analytics rather than intuition *(2)*. Data-driven analytics is applicable to almost all areas of interest in transportation, especially in planning, operation, and maintenance, which traditionally rely heavily on the crew's experience. It shows great potential to improve the program accountability and reliability in such areas. However, a challenge presents itself for data mining technique implementation. One critical issue lies in high quality data acquisition. Tufte *(3)* summarized several factors limiting the application of data-driven analytics that demand pressing attention, including the isolation across databases and data resources, and the lack of informative description about data collection and data manipulation. Since an efficient and accurate data-driven analytics depends heavily on reliable data, these factors undermine its adaptability to similar programs.

Another challenge in the application of data-driven analytics is the complexity of the methodology itself. Compared with traditional methods used in decision-making, data mining techniques provide new insights to transportation problems at costs of sophisticated data processing and analytics. The philosophy of the data mining technique is to turn massive data into actionable information *(3)*. The process is thus generally divided into five phases: 1. organizing for success, 2. building assessment literacy, 3. identifying data sources, 4. aligning data systems, and 5. altering instruction *(4)*. Put into perspective of transportation, the application of data-driven analytics has two major issues. One is the analysis of massive data, namely performance measurement, including: 1) extracting effective information from the data collected by devices or personnel, 2) developing quantitative performance metrics based on the derived information, and 3) validating the reliability and variability of the developed metrics. The other issue is action determination, including: 1) deciding actions that efficiently improve the current conditions, and 2) matching the actionable strategies with quantified performance. Data mining techniques vary with the type of decisions and the applied areas, yet performance measurement and action determination should always be addressed constantly. Figure 1 illustrates the flowchart of implementing data-driven analytics in transportation.

1.1 Problem Statement

Data-driven analytics encompasses a series of data mining techniques to quantify the existing performance and solve respective transportation issues that are beyond the capability of traditional decision making methods. This study focuses on applying data-driven analytics to solve several critical issues in transportation planning, operation, and infrastructure maintenance management.

1.1.1 Dynamic Ridesharing

Dynamic ridesharing has been considered as an emerging solution to traffic congestion due to the growing ubiquity of the Internet of Things. It refers to a transportation mode that facilitates the one-time match of drivers and passengers with similar travel itineraries. Compared with mass transit and taxi, dynamic ridesharing provides carpoolers considerable flexibility to make one-time, on-the-fly trip offers/requests, that can be just minutes before their desired departure times. Currently, dynamic ridesharing oftentimes gains popularity within the region with wide deployment of High Occupancy Vehicle (HOV) or High Occupancy Toll (HOT) lanes, where dynamic ridesharing offerors can meet the requirements of HOV2+ or HOV3+ to utilize such lanes. Under such circumstance, the program provides both parties a faster and more reliable travel experience. Besides travel cost and/or time-saving for individual travelers, dynamic ridesharing also has societal and environmental benefits. For example, by combining travelers in Single Occupancy Vehicles (SOVs), dynamic ridesharing system increases vehicle utilization and reduces the total number of automobiles on the road. Energy consumption, emission, traffic congestion, and parking infrastructure demand are reduced correspondingly *(5)*.

There are several critical factors that determine the success of dynamic ridesharing systems, including driver-passenger matching algorithm (resource allocation), incentives, business model, identity verification, competing travel modes, just to name a few. Dynamic ridesharing system has been proved successful in reducing traffic congestion and providing travelers more reliable travel time in an idealized network, with the presence of only SOV and HOV travelers. The effects of dynamics ridesharing with the presence of other competing modes are yet unknown. In an actual traffic network, people are exposed to various travel alternatives, e.g., public transit, SOV, HOV, etc. The co-existence of different options can influence people's decision-making process, and correspondingly impact the market penetration of dynamic ridesharing. Rather than sharing a private vehicle with strangers and having safety concerns, some travelers may prefer taking public transit or paying tolls.

1.1.2 Nonrecurrent Congestion

The burgeoning development of Intelligent Transportation System (ITS) over the past decades inspires the smart and efficient management of current roadway networks. One concern of freeway performance management is congestion, which can be attributable to recurring and nonrecurring causes. According to the 2012 Urban Mobility Report, urban congestion cost about $12.1 billion dollars and a total of 5.52 billion hours delay in 2011*(6)*. Congestion has surely been growing over the past years. Transportation agencies are therefore actively seeking ways to better monitor the traffic, identify bottlenecks, and respond efficiently and effectively to incidents. From an operations perspective, using a set of meaningful performance measures to obtain comprehensive assessment of the roadway system is one of the most effective solutions for congestion management. It is also critical to decision making. The Moving Ahead for Progress in the 21st Century Act (MAP-21) establishes a performance-based transportation program to guide the transportation capital investment and development. It thus enables the need to carry out a performance-based approach in evaluating the transportation system. Freeway networks play a very critical role in providing accessibility to a multitude of resources and serves as the backbone of a region's economic vitality.

There are seven potential sources that contribute to the travel unreliability on freeway network identified by the FHWA SHRP 2 program. They are traffic incidents, weather, work zones, demand fluctuations, special events, traffic control devices, and inadequate base capacity. As one of the most critical contributors to traffic congestion, incidents account for approximately 50-60% delay on U.S. highways *(7)*. In order to mitigate the impacts of incidents, it would be crucial for the incident management program to develop strategies that can effectively estimate the incident impact range and respond appropriately. The Traffic Incident Management (TIM) is a planned and coordinated process to detect, respond to, and remove traffic incidents and restore capacity as safely and quickly as possible. Accurate estimation of Incident-Induced Delay

(IID) would assist with a better understanding of incident related congestion, and thus provide insights for effective TIM. Transportation agencies use information regarding IID for transportation planning purposes at different levels. Lately, the successful incorporation of reliability analysis into the planning and programming processes also demonstrates the importance of incident effects modeling *(8)*. The estimation and prediction of IID can further be applied to traffic simulation calibration and validation. Accurate estimation of such delay can help identify appropriate decisions regarding incident response so that limited monetary and labor resources can be efficiently allocated. The IID is also essential for the development of active traffic management and integrated corridor management strategies. One critical step for the IID estimation is to determine the impact range of incidents in both spatial and temporal domains, which also makes it feasible to identify secondary incidents due to the congestion caused by a previous incident. According to FHWA, secondary incidents account for 20% of all incidents. They include not only crashes, but also engine stalls, overheating and running out of fuel scenarios where vehicles experience unexpected delay due to the primary incidents. Secondary incident is another criterion for evaluating the effectiveness of TIM. According to Karlaftis et al. *(9)*, the likelihood of a secondary crash increases by 2.8% for every minute that the primary incident continues to be a hazard.

A variety of incident management programs have been launched in recent years to monitor and respond to incidents in an effort to effectively minimize this negative impact *(10–12)*. An accurate and efficient estimation of IID can help facilitate corridor reliability assessment. It can assist with bottleneck identification (e.g., roadway geometric design deficiencies). Benefits can also accrue when corresponding strategies are implemented to enhance safety and smooth traffic, such as ramp metering, variable speed limit, etc.*(13, 14)*. IID can vary significantly in different settings, depending on ambient traffic, roadway configurations, incident severity, lane blockage, etc.

1.1.3 Infrastructure Inspection Sampling

Infrastructure management, often referred to as the decision-making process to allocate resources for infrastructure preservation *(15)*, includes three major components: inspection, maintenance, and rehabilitation. Infrastructure management agencies assess the current conditions of infrastructures, e.g., road shoulder, signage, and pavement marking, via inspection. With the infrastructure inspection results, decisions with regard to which prescribed maintenance and rehabilitation activities are conducted and how the transportation investments are prioritized can be made. Inspection is thus critical as its result is directly tied to the planning of maintenance and rehabilitation activities. Any inaccuracy in inspection will impair the reliability of maintenance and rehabilitation decisions. Yet collecting condition information of infrastructures is very demanding in terms of labor and time, and oftentimes agencies inspect only a portion of the infrastructures, a.k.a. samples, rather than the entire infrastructure inventory to estimate the overall condition. As Mishalani and Gong *(16)* mentioned, there are four factors associated with the accuracy of inspection results: inspection frequency, inspection technologies and data processing methods, sample size, and correlation between observations. Three of the aforementioned factors (inspection frequency, sample size, and correlation between observations) are relevant to the selection of sampling method. Improperly selected sampling method can be a major source of error, while if chosen appropriately, it can be a useful tool for accurate condition estimation. Compared with other options for improving inspection accuracy, e.g., adopting advanced inspection technologies, the usage of proper sampling methods improves the quality of inspection results with marginal investment.

Besides the accuracy of inspection results and initial investment costs, another concern about infrastructure inspection is the recurrent inspection costs. The most common sampling method used by the DOTs is simple random sampling (SRS), which selects the samples based on a random draw. The method is unbiased and able to generate samples that represent all types of infrastructures simultaneously. However, it always requires a large sampling rate to justify the representativeness of samples. Another widely used sampling method is stratified random sampling, which divides the population into strata and selects a sample from each stratum. It applies relatively small sample rates, but the selected sample is only confined to a single type of infrastructures. In infrastructure management, DOTs usually use a highway segment as the sampling unit, where more than one type of infrastructures exist for inspection. It is thus time consuming and operationally inefficient for field personnel if the samples of different infrastructures are widely distributed across all segments. The ideal sampling method is to select the group of highway segments for inspection, in which the sampled infrastructures are representative to reflect their respective Level-of-Maintenance (LOMs) within the entire network. Such a sampling method, allowing conducting once-for-all inspection instead of once-for-each-infrastructure-type, will significantly reduce the inspection costs.

Different from quality control sampling or acceptance sampling where the previous condition of individual sample is unknown, infrastructure management agencies have full access to the historical records of infrastructure conditions on the sampled segments, i.e., location, maintenance log, inspection results, and even latent risk estimation. It facilitates the design of an information-based sampling method that can extract useful historical information to select the most representative samples. Moreover, when the background information includes updated inspection results and maintenance records, effective sampling methods can always dynamically adjust the sample selection.

1.2 Objectives and Scope

The ultimate goal of this study is to explore the implementation of data-driven analytics in transportation research. The study also brings new insights to solving the long-existing issues and improving current methods in transportation planning, traffic operation, and transportation infrastructure maintenance. The specific objectives and scope of this study are as follows:

1.2.1 Dynamic Ridesharing

In the study of dynamic ridesharing, we are interested in two intriguing issues with dynamic ridesharing that have not been thoroughly addressed by the existing studies: 1. The competing mechanism between dynamic ridesharing and public transit; 2. people’s decision-making process under the presence of competing modes. To solve these two issues will help traffic planners, public transit authorities, and ridesharing service providers analyze the market, improve the market penetrations, and plan or deploy the dynamic ridesharing program.

To address these issues, an agent-based model is designed to simulate dynamic ridesharing system in a multimodal network with the presence of HOV lanes and public transit. The model considers travelers’ mode choice preference and simulates their decision-making process for mode selection. By adjusting parameters representing travel mode preferences, the model is applicable to any traffic networks with diverse socioeconomic attributes, e.g., a network with a large number of private vehicle ownership, a network with expensive parking costs, or a network with high public transit demand. The modeling framework developed in this study can be an effective tool for traffic operation agencies to assess the benefits of dynamic ridesharing across different cities and make corresponding marketing strategies.

1.2.2 Nonrecurrent Congestion

In previous studies, IID modeling has not been thoroughly conducted at the individual incident level that can provide an accurate and efficient estimation, owing to analysis methods that either had theoretically stringent assumptions or looked at only one-dimensional changes in traffic data. So the objective of this study is to dynamically identify IID at the individual incident level for performance assessment and modeling purposes. The algorithm should not only capture the dynamic evolution of an incident, but also disentangle the convoluted impact of nonrecurrent vs. recurrent congestions.

To accomplish the objective, a spatiotemporal method to extract information from roadway sensors for IID estimation is presented. The algorithm can be trained by the data itself, leveraging the relationship between historical recurrent data and new information incurred by the dynamic evolution of an incident. This method is data-driven and spatiotemporal in nature to fully uncover the impact and causal mechanism of incident occurrence. IID quantification at individual incident level will enable further analysis on delay-based behavior modeling and inspire follow-up research exploring relationships between the incident itself and its associated features (e.g., severity, lane blockage, or traffic conditions).

1.2.3 Maintenance Infrastructure Sampling

The study of maintenance infrastructure sampling is focused on selecting samples that can accurately reflect LOMs of all infrastructures throughout the network, so DOTs can save enormous resources and time for infrastructure inspection. The sampling method should be capable of choosing proper segments where the conditions of sampled infrastructures can represent the LOMs of the full inventory within the network. It should also allow transportation agencies (e.g., DOTs) to customize the parameters such as sample size, inspection frequency, and infrastructures of interest.

High-dimensional clustering-based stratified sampling (HDCSS) method for infrastructure inspection is presented in this study. The proposed method integrates infrastructure deterioration prediction, high-dimensional cluster analysis, and Locality-Sensitive Hashing. It can incorporate different features, such as infrastructure condition, geographic information, traffic condition, and geometric design, as the information based on which sample is selected. The sampling process is constantly updated with previous inspection results and maintenance records.

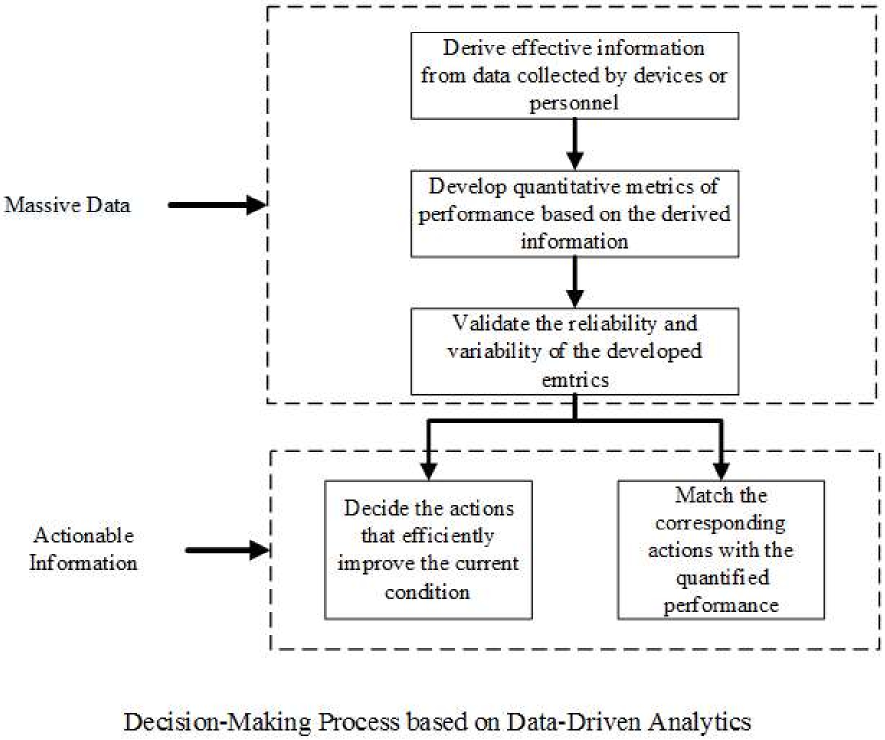

Derive effective information from data collected by devices or personnel
Develop quantitative metrics of performance based on the derived information
Validate the reliability and variability of the developed emtrics
Massive Data
Decide the actions that efficiently improve the current condition
Match the corresponding actions with the quantified performance
Actionable Information
Decision-Making Process based on Data-Driven Analytics

Figure 1 Flowchart of Implementing Data-driven Analytics

CHAPTER 2

DYNAMIC RIDESHARING

An agent-based approach to identify the effects of dynamic ridesharing system in a multimodal network is presented in this chapter. It incorporates the travelers' decision-making process and agent-based modeling of traffic assignment. The approach is implemented in an artificial network for further analysis of dynamic ridesharing demand. This chapter is organized as follows: The first section summarized previous studies on dynamic ridesharing. The agent-based approach is described in the second section. The third section presents an application of the approach to the Sioux Falls test network, followed by the modeling results in the fourth section. The fifth section concludes this study with direction for future research.

2.1 Literature Review

Most of the dynamic ridesharing studies focus on matching algorithm optimization and service design. Furuhata et al. *(17)* summarized a list of such challenges in building a successful dynamic ridesharing system, including traveler matching algorithms, pricing, and institutional design. Among the studies of dynamic ridesharing system design, the driver-passenger matching algorithm has been attracting the most attention *(18–21)*. Agatz et al. *(19)* proposed an optimization-based matching algorithm aiming at minimizing the system-wide vehicle miles and individual traveler's costs. They found that in a multicenter network, there are sustainable ridesharing populations even with low market penetration of dynamic ridesharing. Aissat and Oulamara *(22)* proposed a flexible ridesharing strategy, allowing the driver and the passenger meeting at an intermediate location, to reduce both the driver's detour and total travel costs. Nourinejad and Roorda *(23)* proposed different optimized matching algorithms based on the assumption that each vehicle carries multiple passengers. To maximize the short term revenue of dynamic ridesharing service agency, the commission rate can be as high as 50% of the travel cost. Some researchers tested dynamic ridesharing systems based on nonprivate vehicles. For example, Hosni et al. *(24)* and Santos and Xavier *(25, 26)* both considered shared taxis in a dynamic ridesharing system. Fagnant and Kockelman *(27)* proposed a dynamic ridesharing system using autonomous vehicles.

Another trending area in dynamic ridesharing studies is how to encourage travelers to utilize the systems. Deakin et al. *(28)* analyzed the potential dynamic ridesharing market based on the data collected from downtown and a university campus. They found that high parking charges and limited parking supply are the major boosts to dynamic ridesharing increase. Galland et al. *(29)* used traveler profiles and social media to initiate the agent communication model, and also included a negotiation process between agents. Stiglic et al. *(30)* explored how the flexibility of travelers changes the matching rate in ridesharing system. They found that any increased flexibility in desired departure time or maximum detour time will lead to a significant increment in matching rate. Shaheen et al. *(31)* analyzed the motivations of people using dynamic ridesharing with survey data. Based on their study, the top three motivations are convenience, time savings, and monetary savings. Mote and Whitestone *(32)* studied the influence of mass transportation policies and urban culture on dynamics ridesharing practice based on the discussion of specific cases. Liu and Li *(33)* proposed a compensation scheme based on the congestion evolution over time, to maintain the ridesharing ridership.

There is a very limited number of works accomplished on ridesharing in a multimodal network. Kramers *(34)* discussed integrating dynamic ridesharing system into a multimodal network conceptually. Chavis and Gayah *(35)* developed a mode choice model between fixed-route transit, ridesharing, and individual transit system based on a stated preference survey. In this study, we used an agent-based approach to explore the effects of dynamic ridesharing system in a multimodal network. Agent-based modeling (ABM) is a classical tool to study driver's behavior and the interaction between driver and traffic, which has been widely applied in dynamic ridesharing studies *(23, 27, 29, 36–39)*. Cho et al. *(37)* listed the steps of modeling ridesharing procedure with agents, including creating travel motive, communication and negotiation with other agents, execution of the agreed ridesharing plan, and providing feedback to the network. Bellemans et al. *(40)* applied an agent-based approach to simulate the traffic from city to large manufacturing plants. Sanchez et al. *(39)* addressed privacy concerns and trust issues between travelers in the dynamic ridesharing system by introducing a decentralized reputation management protocol in agent-based modeling

2.2 Methodology

An agent-based study has been conducted to model travelers' decision-making process between driving alone, ridesharing, and mass transit in a multimodal network. The purpose of the study is to find out how travelers switch their commuting mode in a network where ridesharing, HOV lane, and mass transit co-exists, and understand the competing mechanisms between these traffic modes. Based on the modeling results, we can make game-theoretic strategies in the operation of these modes in order to optimize the network performance and costs. The agent-based modeling framework of ridesharing follows the steps listed in *(37)*, including creating travel motive, agent matching, execution of travel plan, and network update. In this section, we will provide a detailed description of each step.

2.2.1 Creating Travel Motive

To create travel motive, a multimodal network with travelers is initialized. The multimodal network is modeled as a directed graph with nodes, general purpose lane arcs, and HOV2+ lane arcs. Nodes in network serve multiple functions, including origin and destination, bus stop, and ridesharing pick-up and drop-off location. We considered travel demands in peak hours in this study, since it is the period during which the traffic network is most likely to be under severe congestion. The mass transit is represented by bus, the most flexible mass transit mode. Traveler agents can only board or alight when the bus agents are at nodes. Since the study focuses on the competing mechanism rather than the ridesharing matching algorithm, we only match travelers traveling from and to the exact same locations (nodes). For individual travelers, there is no mode change in the modeling, neither transfer between buses, nor switch between driving and bus. Traveler agents waiting for ridesharing are assumed to wait at nodes. All travelers' trips start from their origin nodes once the trips are granted. Two types of path that vehicles drive along with: general purpose lane and HOV2+. To simplify the model, we assume that there are only three types of vehicles in the network: single-occupancy vehicle (SOV), HOV, and bus. HOVs and buses travel on HOV2+ lanes, and SOVs travel on general purpose lanes.

There are two types of agents in the model, bus agent and traveler agent. The modeling process is fulfilled by the interaction between agents. Each bus agent carries the features of bus route, bus location, and on-board traveler information, which is represented as:

$$B(route, location, passenger)$$

To justify the cultural and economic difference between different city networks, traveler agent is classified into several categories. Each category has its identical decision-making and interaction rules. They are HOV traveler, SOV traveler, bus traveler, bus and ridesharing traveler, bus and SOV traveler, SOV and ridesharing traveler, and all-mode traveler. By adjusting the percentage of each agent category, the model can simulate the multimodal networks in different cities. Each traveler agent is represented as:

$$T(origin, destination, category, depart\ time)$$

The traveler agent's decision-making process is introduced as follows:

HOV traveler: HOV travelers are the travelers who originally travel in groups, qualifying the requirements of HOV2+. These travelers depart immediately after the traveler agent is created. These travelers travel in HOV lanes.

SOV traveler: SOV travelers are the travelers who only consider traveling in SOV due to concern of security, convenience, or other reasons. During peak hours, it is quite probable that they are driving under more severe congested traffic conditions. The same as HOV travelers, SOV travelers depart immediately, but travel in SOV lanes.

Bus traveler: Bus travelers consider the bus as the only traffic mode. Since we do not consider transfer, this type of travelers only exists between nodes where a bus is available. Travelers who do not own any vehicle, do not want to drive, or are concerned about entering a stranger's vehicle fall into this category. Once the travel motive is triggered, a bus traveler agent will be sent to the waiting list for the next bus. Travelers in the waiting list will be boarding on the "first-come, first-serve" basis. Notice that due to the limited capacity of buses, some travelers may wait for the bus after the next one.

The above three categories of traveler agents are basic agents in the model, with the simplest decision-making process without uncertainty, which only relies on the features of traveler him/herself. The decision-making process for the other four categories of agents will consider the uncertainty of the network. After imposing the network condition to the decision-making process, the four categories of traveler agents can be downgraded to basic agents.

Bus and ridesharing traveler: Travelers who accept both taking bus and sharing a ride with strangers are identified as bus and ridesharing traveler. Casual carpooling passengers fall into this category. Travelers will wait in both bus waiting and ridesharing waiting lines and take the mode which arrives at the destination first. Several criteria must be met before the traveler agent is sent to the waiting lists: 1. A bus is available between the traveler's origin and destination; 2. the number of people on board and waiting for the next bus is less than the bus capacity. Notice that the number of people on board and waiting for the next bus is always higher than the actual people on board when the bus arrives since passengers may get off the bus at previous stops. This criterion tends to encourage travelers who feel comfortable with both bus and ridesharing to utilize a ridesharing system. It reduces the number of people in the bus waiting list, so the travelers who only take the bus would have less probability of waiting for an unreasonably long time.

Travelers in the ridesharing waiting list heading to the same destination are matched as a new HOV traveler agent. To simplify the matching process, we did not identify the driver or passenger. Travelers who have waited for a long time are given the priority in the matching process.

Bus and SOV traveler: Bus and SOV travelers are the travelers who are flexible between taking a bus and driving alone. Traveling by bus always has a more reliable travel time compared to driving in general purpose lanes, since buses use HOV lanes. However, bus trips are only available between certain matched origins and destinations, and have a fixed time schedule. In our model, a bus and SOV traveler is fully aware of the network traffic condition, bus waiting list, and passenger volume on board. Therefore, a bus and SOV traveler will estimate the arrival times by the two modes, and choose the one with the earlier arrival time. If the next bus is close to its capacity or has a later arrival time, the agent will be downgraded to an SOV traveler agent, and bus traveler agent, otherwise.

SOV and ridesharing traveler: SOV and ridesharing travelers would like to give strangers rides if it can significantly reduce their travel time, or they will drive alone. These travelers will estimate the shortest travel time of driving SOV and HOV. There is a trade-off time between driving alone and traveling with a stranger. The travelers will not consider ridesharing unless the sum of HOV lane travel time and the trade-off time is still lower than general purpose lane travel time. In the case where using an HOV lane can significantly reduce the travel time, the travelers will join the ridesharing waiting list. But unlike bus and ridesharing travelers, SOV and ridesharing travelers have a maximum waiting time. Once they have waited for more than the threshold, an SOV and ridesharing traveler agent switches to an SOV traveler agent and departs immediately.

All-mode traveler: All-mode travelers are the most flexible travelers in the model, which accept modes of driving SOV, sharing rides, and taking buses. Once the travel motive of an all-mode traveler is triggered, the traveler will estimate the arrival time of all three modes. Since both bus and HOV travel on HOV lanes, ridesharing always outperforms the bus, which has a fixed schedule and route, in terms of arrival time. If the mode of driving SOV has the earliest arrival time (which is very unlikely) or at least not much more travel time than the other two modes, the all-mode traveler agent is downgraded to an SOV traveler agent. If the model of ridesharing has a significantly early arrival time, the all-mode traveler will wait in the ridesharing waiting list. However, if the bus or the maximum waiting time comes before the traveler gets a match, the all-mode traveler will change his/her mind on taking the bus or driving SOV, respectively.

2.2.2 Agent Matching

By conducting the decision-making process, all traveler agents find their best route and travel mode in terms of arrival time. Then they are released to the network and start to interact with other agents and the travel environment. Agent matching is accomplished by the interactions between agents, including the interaction between traveler agents, and the interaction between traveler and bus agent.

The interactions between traveler agents mainly happen at nodes when traveler agents look for other traveler agents to share rides. The matching process is based on principle of first-come, first-serve. Traveler agents who have been waiting in the line for too long will take alternative modes, i.e., bus or SOV. The matching process is illustrated in Figure 2. There are 7 agents waiting in the line for ridesharing matching in the example. Each agent contains the information of the destination and alternative mode. For example, Agent 1 heads to node 1 and its alternative travel mode is driving alone. In the waiting list, Agent 3 and 4 are both heading to node 3, and therefore, they would drive share the ride. The same process would happen to Agent 5 and 7 as well. The matched agents will depart for their destination immediately. Agent 1 has been waiting for too long but has still not yet found a matched traveler. So Agent 1 will drive to his/her destination (node 1) alone rather than spending time on waiting for a match. The alternative travel mode for Agent 2 is taking the bus. The bus comes when Agent 2 is waiting for a match, so Agent 2 would take the bus.

The bus agents pick up traveler agents from the bus waiting line at each node on the bus route, and drop traveler agents at their destination nodes. The trip of the traveler is done after it is dropped by a bus agent.

2.2.3 Travel Plan Execution

The travel plan is executed by the interaction between agents and the environment. Buses and HOVs travel in HOV lanes, and SOVs travel in general purpose lanes. Due to the limited number of buses on the network, the impact of buses on the traffic network is negligible. Once a traveler agent is released into the network, the travel route is determined as the shortest path in terms of travel time based on the instantaneous traffic condition. The travel time between two neighboring nodes is calculated by BPR function:

$$t(v_h) = t_0[1 + A\left(\frac{v_h}{C}\right)^B] \tag{1}$$

where t is the average travel time, v_h is the traffic volume, C is the road traffic capacity, t_0 is the free-flow travel time, and A and B are calibrated parameters.

When a traveler or bus agent passes a node and starts to travel on another link, the travel time for the coming link is recalculated based on the current traffic condition on the link. The travel time update procedure is illustrated in Figure 3. The traveler agent in Figure 3 travels along the route of node 1-2-3-4. The travel time of each link is estimated based on current traffic condition and used as the potential travel time for entering traveler agents. Notice that when the traveler agent enters the link 1-2, the travel time of link 1-2 is 3 minutes, and the travel time of link 2-3 is 4 minutes. But when the agent enters the link 2-3, the travel time of link 2-3 becomes 5 minutes. So the travel time for the agent on link 2-3 is 5 minutes.

2.2.4 Network Update

The network volume is constantly changed as new agents joining the links and exiting the traffic at nodes. The traffic volume on each link is calculated as:

$$\mathrm{v} = \Sigma_i^n \frac{1}{t_i} \tag{2}$$

where n is the number of agents traveling on the link, and t_i is the travel time for agent i.

2.3 Case Study

We test the agent-based model in the classic Sioux Falls network, which consists of 24 nodes and 76 directed arcs. The spatial configuration of the network is illustrated in Figure 4. The network was originally proposed by *(41)*, based on a simplified road network of Sioux Falls. The network is widely used in numerical experiments of simulating traffic congestion, public transit, and dynamic traffic assignment. In this study, we downgraded the link capacity and estimated the HOV lane capacity based on the original capacity of each link, shown in Table 1.

The travel demand during peak hour is around 336,000 veh/hour. The peak hour spans for 2 hours, including three 40-minute periods. Traffic demands during each period are 30%, 40%, and 30% of the total travel demand, respectively. The original percentage of travelers $\mathrm{P_{HOV}}$ in HOV is assumed as 15% *(42)*. The probability of travelers belonging to each traveler type is determined by the market penetrations of ridesharing $\mathrm{P_{RS}}$ and public transit $\mathrm{P_{PT}}$. The probability of a traveler belonging to each type is calculated as:

HOV traveler: $\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})$

SOV traveler: $\left(1-\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})\right)*(1-\mathrm{P_{PT}})*(1-\mathrm{P_{RS}})$

Bus traveler: $\left(1-\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})\right)*(1-\mathrm{P_{PT}})*\mathrm{P_{RS}}/2$

Bus and ridesharing traveler: $\mathrm{P_{PT}}*\mathrm{P_{RS}}*\left(1-\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})\right)$

SOV and ridesharing traveler: $\left(1-\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})\right)*(1-\mathrm{P_{PT}})^2*\mathrm{P_{RS}}$

Bus and SOV traveler: $\left(1-\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})\right)*\mathrm{P_{PT}}*(1-\mathrm{P_{RS}})$

All-mode traveler: $\left(1-\mathrm{P_{HOV}}/2(1-\mathrm{P_{HOV}})\right)*(1-\mathrm{P_{PT}})*\mathrm{P_{RS}}/2$

The public transit network from *(43)* has been slightly modified and used in this study. The itineraries of five bus routes are defined and shown in Table 2. The bus headways range between 10 to 20 minutes. Considering the high traffic demand in the modeling, we use a bus fleet with the capacity of 600 passenger/fleet to serve the function of public transit.

2.4 Result Analysis

As mentioned in the literature review, many studies focused on the efficiency of the matching algorithm. Stiglic et al. *(30)* found that many more shared trips can be matched if the waiting time constraint is slightly extended in the matching process.

Figure 5 shows the sensitivity analysis of trip matching constraints. The agent-based model has been applied to scenarios with high public transit market penetration (80%) and low public transit market penetration (5%), respectively. The market penetration of dynamic ridesharing is set at 40%. Different with Stiglic's findings, the numbers of shared trips change very slightly as the maximum waiting time (constraint) changes. Especially when the market penetration of public transit is high, the number of shared trips slightly decreases as the maximum waiting time increases. That might be attributable to the setup of the Sioux Falls network. One important assumption in the network is that nodes serve as locations for trip departure and arrival, shared trip matching, and bus stops. Therefore, trips are matched in a very short period of time. The waiting list for trip matching reaches equilibrium within 1 minute, so increasing the maximum waiting time does not lead to more shared trips. When the market penetration of public transit is high, longer maximum waiting time increases the probability of travelers utilizing public transit, which causes a slight decrease in number of dynamic ridesharing trips.

Vehicle-Hour-Traveled (VHT) is one of the important criteria to measure the performance and efficiency of a traffic network. With unchanged traffic demand, a network with less VHT usually means enhanced vehicle occupancy and less congestion. The impact of a ridesharing system on traffic network has been studied by many researchers. In this paper, we will discuss the VHT reduction induced by the ridesharing system. Figure 6 shows the network VHT with different ridesharing market penetrations when the bus market penetration $P_{PT} = 5\%$. As shown in Figure 6, there is a significant drop in VHT when the ridesharing market penetration increases from 0 to 20%. As the ridesharing market penetration increases, the total VHT of the network keeps decreasing. The plot of ridesharing market penetration with low public transit market penetration can also be the potential VHT reduction prediction after a new ridesharing system is first introduced to an existing multimodal network. The VHT with no ridesharing market penetration represents the traffic condition during peak hours before the ridesharing system is introduced. When the ridesharing system is first introduced, SOV travelers looking for reliable travel experience will start to utilize the system. These travelers are combined and drive on HOV lanes, which were with quite low occupancy previously. Since the total ridership decreases and many vehicles switch from the general purpose lanes to HOV lanes, the traffic congestion in general purpose lanes is largely relieved. With publicity and marketing strategies applied, the ridesharing system would attract more attention, and more travelers would consider utilizing the ridesharing system in terms of travel mode selection. So the network VHT keeps decreasing significantly until the market penetration of ridesharing is 40%. However, if the market penetration of ridesharing keeps increasing, the HOV lanes become saturated. The ridesharing benefits in terms of travel time become marginal. In such a case, even if the market penetration of ridesharing keeps increasing, the corresponding VHT reduction becomes mild.

Figure 7 shows the VHT with different ridesharing market penetrations when the public transit market penetration is high (60%). Similar to Figure 6, the network total VHT decreases with high ridesharing market penetration compared with the multimodal network without ridesharing system. However, there is an obvious VHT increase when the market penetration of ridesharing system is low. In a network with high public transit market penetration, travelers using public transit can be classified into two types in terms of private vehicle ownership. The travelers of the first type do not own any private vehicles, so public transit is the only available travel mode before the ridesharing system appears. The second type of travelers using public transit own private vehicles but do not use them for a daily commute due to the unreliable travel time, expensive parking costs, or other reasons. When the ridesharing system becomes available in the network, many second type travelers would utilize their private vehicles for their daily commute at affordable costs by providing rides to other people via the ridesharing system. Travelers of the first type also would like to use the ridesharing system due to the convenience and travel reliability of traveling in private vehicles. As a result, the application of a ridesharing system will encourage travelers to utilize private vehicles for the daily commute. In a short period after the ridesharing system is launched, the network will have more VHT and suffer more severe congestion than before. The travel experience for each individual traveler who uses the ridesharing system becomes more comfortable and convenient, but the travel time becomes longer. In such a case, it is questionable if the ridesharing system can attract more users.

This is an identical Braess-like paradox that the network congestion actually increases by adding a congestion relief traffic mode to the network. Usually, in the studies of traffic operation, the most efficient method to avoid the Braess paradox is to remove the roads that most probably lead to the paradox. But in our case, if the ridesharing market penetration keeps increasing, the congestion brought on by the increased private vehicle ridership will be neutralized by the increased occupancy of HOV lanes. As shown in Figure 7, when the ridesharing market penetration is higher than 20%, the network VHT starts to decrease. So the challenge in boosting a ridesharing system in such a case is how to still attract users when the ridesharing system shows negative impacts to the network.

Another traffic network performance measurement is the total travel cost. Travel time, fuel consumption, and even greenhouse gas emissions can all be converted to monetized values to assess the impact of dynamic ridesharing to the existing network. Evaluating the performance of dynamic ridesharing based on travel costs together with VHT might potentially reveal the travelers' hidden motivation of using that service. Therefore, the proper computation of travel costs as a performance indicator would be an interesting potential research area.

In a network with low market penetration of public transit, a ridesharing system can significantly reduce the network-wise VHT as well as providing a good travel experience. An individual traveler using the ridesharing system can obtain reliable travel time and convenient travel experience. The ridesharing system in such a network is self-advertised and can easily attract users subject to traffic operation. It has no conflicts with the existing network infrastructure. However, in a network with high market penetration of public transit, the companies or agencies should consider more than marketing strategies. For example, they need to make sure that the government provides sufficient infrastructures to accommodate the extra congestion induced by the ridesharing system. To avoid that current users suffer too much extra travel time, the ridesharing system can provide incentives to travelers who would like to bring their private vehicles, or even only open to such travelers before the network VHT starts to decrease.

2.5 Summary

In the study of dynamic ridesharing, an agent-based approach is proposed to model the interaction between traffic and environment with the existence of dynamic ridesharing in a multimodal network. It integrates the decision-making process of travelers under uncertainty with agent-based modeling. Traveler mode choice is greatly influenced by the existence of dynamic ridesharing system. There are two major findings from this study: the first finding is that the matching rate of the dynamic ridesharing system is quite insensitive with the matching constraints when competing with public transit. Public transit attracts travelers who are waiting for the matched trip by providing a reliable alternative, which neutralizes the increased number of shared trips caused by loose matching constraints. Another finding is that the impact of dynamic ridesharing system on a multimodal network in terms of congestion relief varies significantly with the market penetration of public transit. When very few travelers utilize public transit, introducing the dynamic ridesharing system would lead a VHT reduction to the network, which means the less congested traffic condition. But when the network originally has high market penetration of public transit, introducing the dynamic ridesharing system would initially increase the network-wise congestion. As the market penetration of dynamic ridesharing keeps increasing, the network-wise congestion would decrease.

Due to the very different impacts of dynamic ridesharing system on a multimodal network, different marketing strategies should be applied. Especially when public transit is a preferred travel mode for most travelers in the network, it is the dynamic ridesharing service provider's and the government's responsibility to make sure that the existing infrastructure is sufficient to accommodate the extra congestion induced by the ridesharing system. Ridesharing service providers would like to deploy incentive strategies to accelerate the market penetration increasing and shorten the duration ofintroducing extra congestion.

Due to the complexity of ridesharing matching algorithms, in this study, ridesharing match was based on the principle of first-come, first-serve, which can be greatly optimized. In future efforts, the approach will be applied to actual multimodal networks with different demographics in order to study the real impacts of the dynamic ridesharing system. In the modeling of such a complicated network with high travel demand, optimized matching algorithms will be applied. This will also be beneficial in making marketing and operational strategies for the ridesharing system. Another intriguing topic is to develop a performance measurement index for the multimodal network with dynamic ridesharing. In the case study, VHT has been used as the performance measurement of the network, which quantifies the summarized volume throughout the network during peak period. Since VHT cannot represent the uneven spatial and temporal distribution of traffic congestion, it is necessary to develop a new spatial and temporal varied performance measurement index.

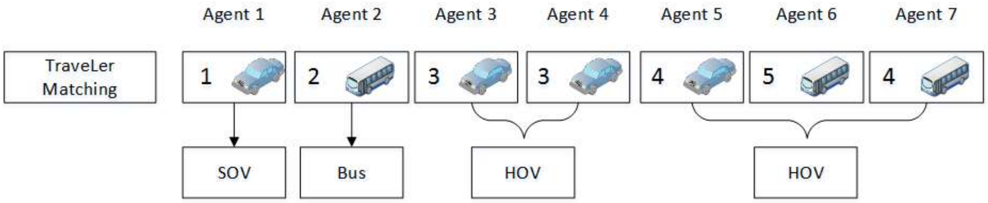


Figure 2 Illustration of Traveler Matching for Ridesharing

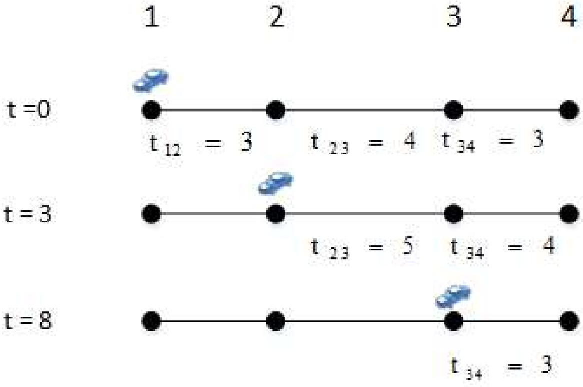


Figure 3 Illustration of Route Travel Time Recalculation

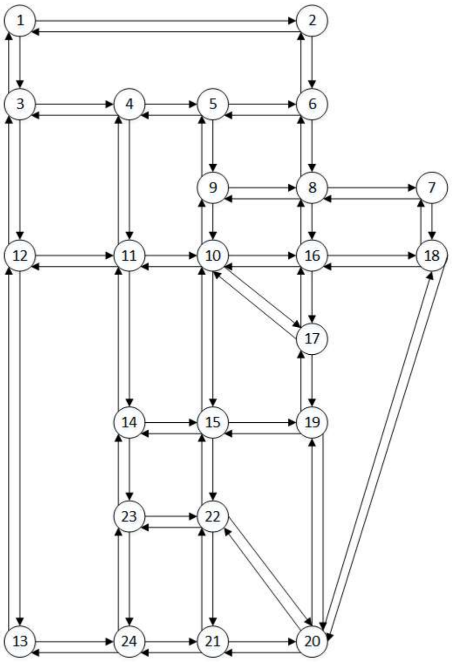


Figure 4 Illustration of Sioux Falls Network

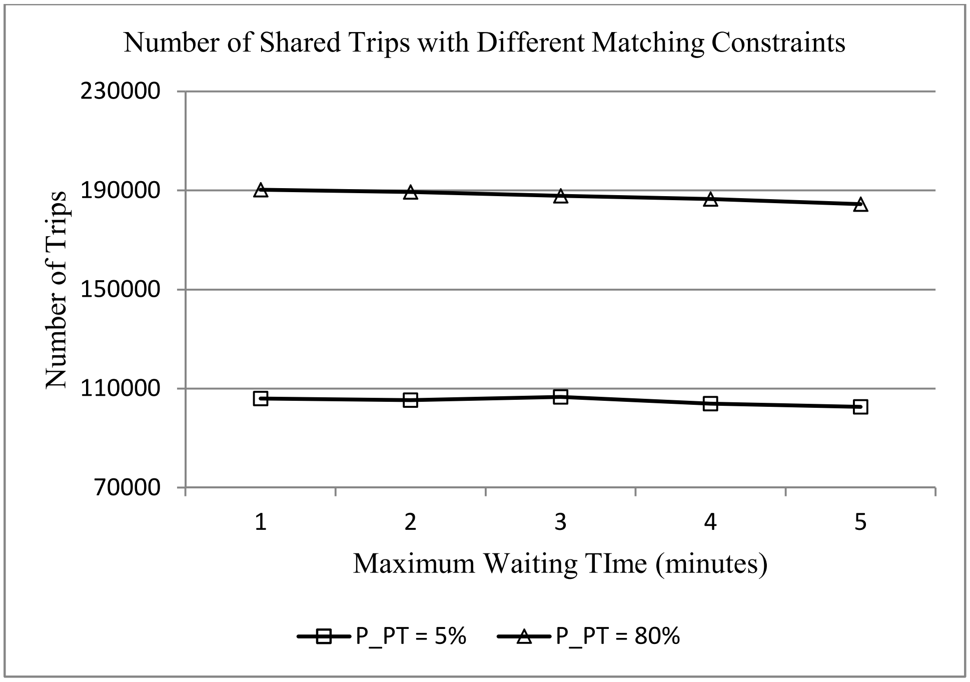


Figure 5 Number of Shared Trips with Different Maximum Waiting Time for Matching

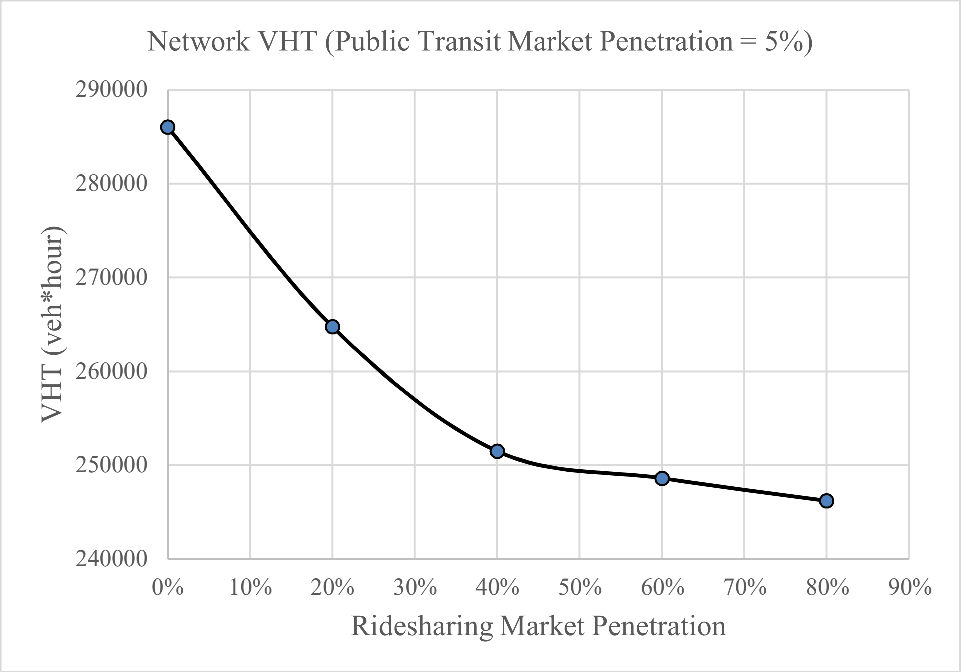


Figure 6 VHT with Different Ridesharing Market Penetrations When Bus Market Penetration is 5%

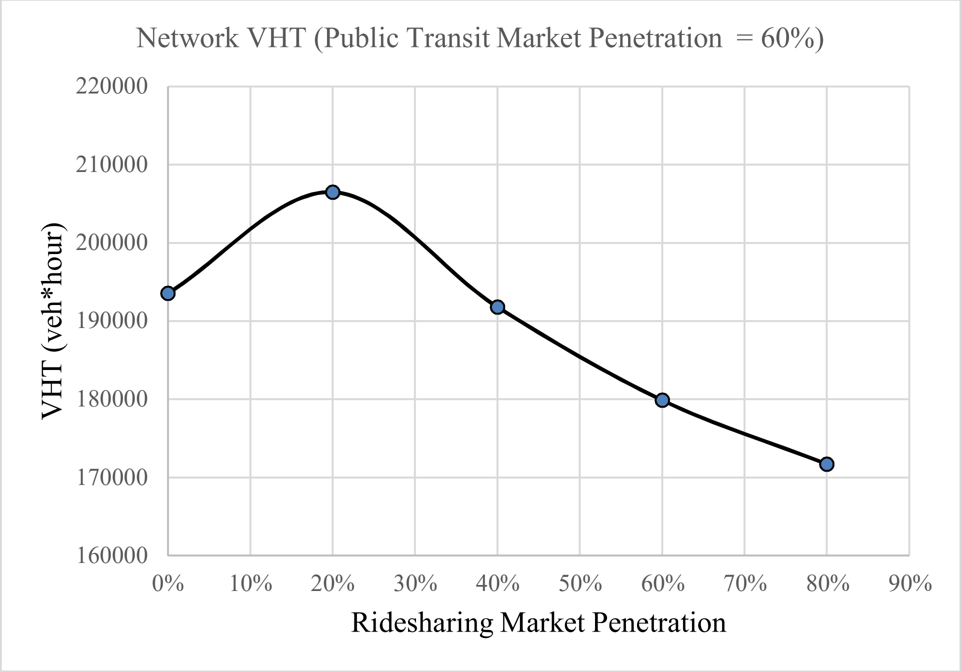


Figure 7 VHT with Different Ridesharing Market Penetrations When the Bus Market Penetration is 60%

Table 1 Criteria for Estimating the Number of HOV Lanes

Original Capacity (veh/hour)	General Purpose lane capacity (veh/hour)	Estimated HOV lane capacity (veh/hour)
Capacity >=15000	13,200	5,700
15000> Capacity >=7500	8,800	3,800
Capacity <7500	4,400	1,900

Table 2 Bus Route Itineraries

Route	Mean headway	Itinerary of route
1	10	1 3 4 5 9 10 15 22 21
2	20	2 6 5 9 10 15 22 21
3	10	12 11 10 16 17 19 20
4	10	13 24 23 14 15
5	10	7 8 16 18

CHAPTER 3

NONRECURRENT CONGESTION

This chapter demonstrates the implementation of big data analytics in quantifying IID at the individual level and the utilization of quantified IID in terms of traffic operation. The first section describes the previous studies on IID quantification. It is followed by a detailed explanation of the proposed methodology. The third section in this chapter describes a case study with the proposed methodology. The fourth and last section summarizes the results of the case study and conclusions drawn from the results.

3.1 Literature Review

IID is defined as the extra travel delay resulting from incidents on top of the recurrent congestion *(44)*. Previous studies on IID were based on the mechanism of delay quantification. In terms of methodology, Deterministic Queueing Theory (DQT) and the Shockwave-based algorithm are most commonly adopted *(45, 46)*. For DQT, delay is determined as the difference between the curve of the original traffic condition and the curve of the queuing process after incidents. The model is implemented with assumed capacity reduction and empirically determined incident duration functions *(47)*. The results of DQT highly depend on assumed functions, impairing the robustness of the method.

Shockwave-based algorithms are developed on the basis of macroscopic traffic flow theory, treating incidents as flow perturbations. Once an incident (perturbation) occurs, shockwaves are generated and spread backward in the traffic flow. Features of perturbation (i.e., the variation of incident-induced perturbation, clearance time, and maximum queue length) can be calculated *(45)*. Similar to DQT, an analytical solution is not available unless simplified traffic conditions apply. Otherwise, a numerical solution is needed *(48)*. Rakha and Zhang *(49)* found that the two aforementioned methods yield consistent results. However, at highway bottlenecks, DQT provides a more accurate estimation of incident delays. Besides the mechanism-based approach, the statistical method has been applied to study the general distribution of incident delay. Skabardonis et al. *(50)* estimated the average and probability distribution of delay using loop detector data and showed that nonrecurrent congestion accounts for only 13% to 30% of total delays, depending on the extent of recurrent congestion. A statistical method provides data-based estimation of IID but fails to link IID with location-specific and incident-associated characteristics. These studies offer general insights for IID estimation. Yet neither mechanism-based methods nor statistical methods are capable of quantifying IID at the disaggregate level, where important features associated with individual incidents remain unknown.

Early research on incidents' impact at disaggregate level used the static method, assuming that the maximum impact area of incidents has a fixed boundary *(9, 51)*. Karlaftis et al. *(9)* believed that the impact area of an incident to induce secondary incident is 1.5 km and 15 minutes. Moore et al. *(51)* defined the impact area as 2 miles/2 hours. Due to the fact that the predefined boundary may not be suitable to all the incidents, researchers turned to a dynamic method defining the dynamic influence area of incident on the basis of analytical/empirical approach with traffic data. Efforts to uncover the dynamic impact of incidents have led to a wide-scale use of spatiotemporal analysis *(52–55)*. Spatiotemporal analysis can be used to determine the additional delay within the spatial and temporal extent under the impact of incidents. It focuses on the intrinsic variations of individual incidents, and thus avoids the bias induced by the assumed relationship between delay and surrounding conditions. There are two challenges in such analysis. The first challenge is the incident's impact range in spatiotemporal domain. Under ideal situations, the spatiotemporal extent is a contiguous region originating from the moment an incident occurs. Yet in reality, disturbances exist within the region due to traffic fluctuation. The second challenge is the identification of recurrent congestion. A method is needed to disentangle the compounded impact of nonrecurrent and recurrent congestions.

Common practice is to choose an empirical threshold based on historical traffic conditions and use speed or travel time as delay indicators to distinguish the two types of congestion. Spatiotemporal region with indicator value below this threshold is considered experiencing only recurrent congestion. Chung *(53)* applied the spatiotemporal concept to freeway incident delay quantification. If speed falls below the threshold $\bar{\mathrm{s}} - \alpha\, \sigma_{\mathrm{s}}$ (where $\bar{\mathrm{s}}$ is the average speed, σ_{s} is the standard deviation of speed, and α is the scaling factor), the spatiotemporal cell is considered congested. The single average speed $\bar{s}$ is used as recurrent condition indicator, which does not consider the traffic variation under an incident-free scenario. Also note that α is determined empirically. Thus, any bias may result in an over- or under-estimation of the nonrecurrent congestion. Chung and Recker

(54) then improved upon this method using an optimization model to minimize the probability of errors, including the speed threshold being falsely higher than the speed at uncongested cells or falsely lower than the speed at congested cells, such that the optimum value of α can be calculated and applied to the spatiotemporal extent determination. Note that both empirical and optimization methods are built on the assumption that recurrent congestion can be estimated analytically by$\bar{s} - \alpha\sigma_s$. Snelder et al. *(55)* expanded the spatiotemporal method from corridor to freeway network. They assumed that the boundary of spatiotemporal extent is a parallelogram with a slope of 70 km/h (shockwave speed). To determine the recurrent congestion, they constructed a Vehicle Loss Hour (VLH) series with weighted VLH from weeks before and after an incident, and used the median as the referencing case. This empirical method is simple to implement but lacks validation. Anbaroglu et al. *(52)* applied the spatiotemporal clustering analysis to freeway network using links as unit. The threshold for recurrent congestion was also determined via optimization.

Another stream of applications for spatiotemporal analysis is secondary incident identification *(56–58)*. Secondary incidents are considered stochastic events induced by traffic congestion originating from the primary incident. The key focus has been determining the primary incident's spatiotemporal impact boundary. According to Yang *(57, 58)*, when incidents are identified as secondary, not only the subsequent incident itself but also all the spatiotemporal cells between the previous and subsequent incidents should be within the spatiotemporal extent. However, this would falsely exclude secondary incidents when the spreading of impacts is dominant in one direction (either spatial or temporal). Chung *(56)* applied extra criteria based on the shape of impact extent in his optimization model, such as the uninterrupted progression of shockwave, upstream-directional progression, and position of dot-shaded area.

The spatiotemporal analysis also applies to other nonrecurrent factors besides incident, e.g., adverse weather *(59)*, rubbernecking *(55, 60)*, and work zone *(53)*, just to name a few.

An emerging popular approach for recurrent congestion estimation under incidents' influence is the data mining method. Habtemichael and Cetin *(61)* applied clustering analysis to identify a similar traffic condition pattern for incidents, and then predicted the recurrent congestion based on that. Their results showed that travel time outperforms traffic volume as a pattern recognition indicator, and the K-Nearest Neighbor (KNN) method yields the best prediction results. Compared to empirical methods, the data mining approach adopts unsupervised learning techniques and can be customized to different datasets. Park and Haghani *(62)* used neural network models to predict the likelihood of secondary incidents. To better explain neural networks, a pedagogical rule extraction approach was developed and applied to extract comprehensible rules.

Spatiotemporal analysis and data mining techniques offer greater insight into the convoluted causes of delay and unveil the true impact of nonrecurrent congestion to explore unsettling reasons for safety enhancement that have escaped notice. This study complements existing literature by combining analytical approach with data mining techniques to dynamically determine the spatiotemporal extent of individual incidents. The IID quantification methodology excludes the impact of secondary incidents for the first time and includes shockwave theory in spatiotemporal analysis. The information construction process can be further used to uncover a variety of features that are associated with particular incidents for an optimal freeway management.

3.2 Methodology

IID quantification at individual incident level will enable further analysis on delay-based behavior modeling and inspire follow-up research on exploring relationships between the incident itself and associated features (e.g., severity, lane blockage, or traffic conditions). The proposed algorithm in this study starts by ruling out the influence of secondary incidents, as the subsequent events occurring in spatiotemporal domain can result in an overestimation of the primary incident impact. This is achieved by mapping cascading incidents onto the spatiotemporal extents of the potential primary incidents. The total delay induced by each individual incident is then dynamically calculated using a spatiotemporal clustering approach. Recurrent congestion can eventually be determined through heuristically searching in the historical database for pattern recognition. The methodology is data-driven in nature and the algorithm is easily transferable to any traffic operation system that has access to the sensor data at corridor level.

The algorithm for IID estimation follows a three-component scheme: secondary incident identification, spatiotemporal extent determination (total delay), and recurrent congestion identification. The detailed explanation for each component is presented in this section:

3.2.1 Secondary Incident Identification

Due to the cascading effect of secondary incidents, delays can be elongated substantially. To separate the delay induced by primary and secondary incidents, a method considering the spatiotemporal effects of primary incidents is required. As mentioned in Chung (2013), the secondary incident identification should be fulfilled by defining the primary incident impact area. Delay induced by an incident, defined as the excess Vehicle Hours Traveled (VHT) with a reference speed of 60 mph as an example, can be visualized in a spatiotemporal contour map as shown in Figure 1. In this figure, the spatiotemporal impact extent is established on the basis of three criteria: IID detection, shockwave front location, and contiguity of impact region. Any cascading incidents occurring within the spatiotemporal extent are identified as secondary incidents. Specific explanation of the criteria follows.

3.2.1.1 IID Detection

Let $D_{Sec\_tot}(i,j)$ be the representative of total delay at location i with Time-of-Day Day-of-Week (TOD DOW) j induced by an incident, $D_{Sec\_rec}(i,j)$ be the representative of corresponding recurring delay, and $d_{Sec}(i,j,k)$ be the historical delay under incident-free scenario at the same location i with TOD DOW j, but at different week k. The incident-free scenario is defined as no incident occurring within 5 hours prior to the time stamp and within 10 miles upstream of the location. The recurring delay $\mathrm{D}_{Sec\_rec}(\mathrm{i,j})$ is estimated with $d_{Sec}(i,j,k_1)$, $d_{Sec}(i,j,k_2),\ldots,d_{Sec}(i,j,k_n)$, where $k_1,k_2,\ldots,k_n$ are the weeks under incident-free scenario. The spatiotemporal extent based on the difference between total and recurrent delays, within which any new incident occurred, offers a sense of existence of secondary incidents. In case a secondary incident appears, its impact extent would be connected with one of the primary incidents, expanding upon the original spatiotemporal range. As a result, a secondary incident would never appear at the boundary of a spatiotemporal impact region. Using fixed percentiles of historical delay to represent the recurring congestion (e.g., 80th percentile), a binary contour map to detect the existence of IID can be generated by subtracting $D_{Sec\_rec}(i,j)$ from $D_{Sec\_tot}(i,j)$:

$$I_{Sec}(i,j) = \begin{cases} 1, if\ D_{Sec\_tot}(i,j) - D_{Sec\_rec}(i,j) > 0 \\ 0, if\ D_{Sec\_tot}(i,j) - D_{Sec\_rec}(i,j) \leq 0 \end{cases} \qquad (3)$$

where I_{Sec} is the indicator of IID existence. $I_{Sec}(i,j) = 1$ suggests that there is IID at spatiotemporal location (i,j), otherwise $I_{Sec}\ (i,j) = 0$.

3.2.1.2 Shockwave Front

Considering random factors that may influence the delay after an incident occurs (e.g., adverse weather, work zone), $I_{Sec}(i,j) = 1$ does not necessarily mean that the delay is purely incident-induced. To rule out such possibilities, the shockwave front location method is used to filter out other nonrecurrent delays. As soon as an incident occurs, a shockwave is triggered. The shockwave is originated from the incident and spread spatially backward and temporally forward. Thus, an incident impact region should coincide with the spatiotemporal area behind the front of shockwave. The spatiotemporal contour map can be broken down into two parts:

$$S_{Sec}(i,j) = \begin{cases} 0, if\ (i,j)\ is\ ahead\ of\ shockwave\ front \\ 1, if\ (i,j)\ is\ on\ or\ behind\ shockwave\ front \end{cases} \qquad (4)$$

where S_Sec is the indicator for determining whether the location (i,j) is behind the shockwave front.

The shockwave effect is a complicated process and varies based on traffic volume, density, and even severity of incidents. The shockwave front location is defined with a dynamic threshold as developed in *(63)*: the sensor station whose traffic density is greater than twice the density at the upstream station and the sensor station whose average speed is greater than twice the speed at the downstream station. Therefore, if a delay is detected ahead of the shockwave front in the spatiotemporal context, it is not considered to be induced by the incident.

3.2.1.3 Contiguity of Impact Region

The propagation of congestion is unidirectional in both spatial and temporal domains. Thus, the IID at spatiotemporal location (i,j) (if any) must be inherited from a prior location that is spatially forward or temporally backward. The contiguity of impact region suggests that if IID exists at (i,j), it must also exist in either $(i-1,j)$ or $(i,j-1)$, or both. Mathematically, this can be expressed as:

$$C_{Sec}(i,j) = \begin{cases} 1 & if\ i = 0\ and\ j = 0 \\ \min\{1, I_{Sec}(i-1,j) * S_{Sec}(i-1,j) * C_{Sec}(i-1,j) + I_{Sec}(i,j-1) * S_{Sec}(i,j-1) * C_{Sec}(i,j-1)\}, & else \end{cases} \tag{5}$$

where C_{Sec} is the indicator for contiguity. $C_{Sec}\,(i,j) = 1$ suggests the criterion of contiguity is met, otherwise $C_{Sec}(i,j) = 0$.

Based on the aforementioned criteria, every spatiotemporal cell within the impact range of an incident must satisfy:

$$I_{Sec}(i,j) * S_{Sec}(i,j) * C_{Sec}(i,j) = 1 \tag{6}$$

Therefore, any incident that falls within the impact range of a prior incident would be considered secondary (1-zone in Figure 8(d)). Figure 8(a)-(c) demonstrates the results of applying the IID detection, shockwave front, and contiguity of impact region criteria to the spatiotemporal profile of delay after incident. Cells marked as 1 represent the spatiotemporal units that meet the criterion in each plot. Figure 8(d) is the conjunction plot based on the three criteria. Note that the spatiotemporal impact extent due to a cascading incident would be much greater than those of independent incidents. If a delay is calculated based on such overlapping effects, it would significantly overestimate IID, especially for locations with high secondary incident frequency. With all the primary incidents and secondary incidents identified, attention is directed to total delay and recurrent delay quantification, which are spatiotemporal-sensitive.

3.2.2 Total Delay of Independent Incident

The total delay of an incident refers to the accumulated delay augmented within its spatiotemporal impact extent. Compared to secondary incident identification, total delay quantification is more sensitive to the spatiotemporal range. The same spatiotemporal clustering analysis applies here except the fixed percentile threshold for defining normal condition. Instead, a statistical model is utilized to provide a more reasonable threshold and can be trained with empirical data.

To implement the threshold estimation, we randomly chose 1,000 TOD DOW and locations, and constructed histograms of the delay occurring during those periods. Two typical patterns of delay frequency emerge as shown in Figure 9. Nonparametric estimation determines that the incident-free delay follows Weibull distribution, whose probability density function is expressed as:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}}, x \geq 0 \\ 0, x < 0 \end{cases} \tag{7}$$

where k is the shape parameter. When k=1 or k=2, the distribution becomes Exponential Distribution or Rayleigh Distribution. The Cumulative Distribution Functions (CDF) are:

$$\begin{aligned} k = 1, F(x;\lambda) = 1 - e^{-\lambda x}\ (x \geq 0)\ \text{(Exponential Distribution)} \\ k = 2, F(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}\ (x \geq 0)\ \text{(Rayleigh Distribution)} \end{aligned} \tag{8}$$

The parameters can be estimated as:

$$\hat{\lambda} = \frac{n}{\Sigma_{\mathrm{k}=1}^{n} d(i,j,k)} \tag{9}$$

$$\hat{\sigma} = \sqrt{\frac{1}{2n}\Sigma_{i=1}^{n} x_i^2}$$

Let $\mathrm{d(i,j,k)}$ refer to the historical delay under incident-free scenario at location i, TOD DOW j, and week k. With distribution parameters known, the Pth percentile of delay can be estimated as:

$$\widehat{D}_{exp}(i,j) = \frac{\ln\left(\frac{1}{1-P}\right)}{n}\Sigma_{k=1}^{n} d(i,j,k) \quad (10)$$

$$\widehat{D}_{Ray}(i,j) = \sqrt{\frac{\ln(\frac{1}{1-P})}{n}\Sigma_{k=1}^{n} d(i,j,k)_k^2} \quad (11)$$

where $\widehat{D}_{Exp}$ and $\widehat{D}_{Ray}$ are the estimated threshold when delay follows Exponential and Rayleigh Distribution, respectively. The distribution of historical delay varies by TOD DOW, so instead of exploring the distributions for any TOD DOW, we used the minimum of $\widehat{D}_{exp}$ and $\widehat{D}_{Ray}$ as the threshold.

Let $D_{Ins}(i,j)$ be the representative of instantaneous delay at location i and TOD DOW j after an incident and D_{Tot} denote the total delay of an incident. The spatiotemporal extent of an incident's impact is defined as:

$$I(i,j) = \begin{cases} 1, if\ D_{Ins}(i,j) - \min\{\widehat{D}_{exp}(i,j), \widehat{D}_{Ray}(i,j)\} > 0 \\ 0, if\ D_{Ins}(i,j) - \min\{\widehat{D}_{exp}(i,j), \widehat{D}_{Ray}(i,j)\} \le 0 \end{cases} \quad (12)$$

$$D_{Tot} = \Sigma_i \Sigma_j I(i,j) * D_{Ins}(i,j) \quad (13)$$

where I is the congestion indicator,$I(i,j) = 1$ indicates that it is congested at location (i,j), otherwise $I(i,j) = 0$.

It is important to reiterate that the congestion threshold estimation is performed in both Secondary Incident Identification and Total Delay Determination. Compared to the fixed percentile method in secondary incident identification, applying a statistical distribution model can avoid bias due to limited sample size and outliers. Yet the selection of thresholds can be risky. A lower threshold may incorporate any possible delay into the total delay, but also significantly expand the spatiotemporal impact range, compromising the accuracy of the method. Though rarely observed, for extreme incident cases where the spatiotemporal extent is unreasonably long (e.g., more than 5 hours), a fixed spatiotemporal extent should apply.

3.2.3 Recurrent Delay Determination by Pattern Recognition

Generally, recurrent delay is defined as congestion caused by routine traffic operations in a typical setting. Yet traffic conditions vary on a daily basis even for recurrent congestion. Thus, when predicting the recurrent delay for an incident scenario, the "background congestion" requires special attention to trace from historical record. The "typical recurrent congestion" determined from statistical models in previous studies oftentimes is not applicable to every incident scenario. We remedy this through a pattern recognition process, where recurrent delay is considered as a function of location, TOD DOW, traffic condition, and other miscellaneous factors that can be expressed as:

$$d_{Rec^*} = F(i, j, T, \dots) \tag{14}$$

where d_{Rec^*} is the accumulated delay within the incident's impact extent if there was no incident, i is the location, j is the TOD DOW, and T is background traffic condition.

Other variables are considered to have marginal effects and were thus ignored in the equation. When considering incident scenario, it is impossible to infer what the recurrent congestion would be if the incident did not occur, but recurrent delay can still be deduced through matching the traffic conditions from historical database. For any historical traffic scenario T_{his}, if there exists $|T - T_{his}| < \epsilon$, where ϵ is a threshold for the difference of traffic condition. It is reasonable to assume that:

$$|d_{Rec^*} - d_{his}| < \epsilon' \qquad (15)$$

where d_{his} is the recurrent delay of the matching historical scenario, and ϵ' is threshold for the difference of recurrent delay.

The sensitivity analysis of thresholds ϵ and ϵ' will be investigated in future work. Previous research compared three pattern recognition techniques (DOW, cluster, KNN) with different weighting methods *(61)*. Yet without knowing the relationship between delay, location, time, and traffic condition, any weighting attempt is susceptible to questioning due to lack of validation. In this study, we performed pattern recognition based on TOD-DOW. Quantifying the recurrent congestion becomes equivalent to identifying the best-matched historical traffic scenario at the same location and TOD-DOW. The performance measure for pattern recognition is VHT, which can best describe speed and volume and is easy to obtain from traffic sensors. It is critical that the historical matching scenarios be incident-free. Therefore filtration should be applied to the database (no incident within a 5-hour span at the same location and TOD-DOW). Statistical performance indicator Root-Mean-Square-Error (RMSE) is used for choosing the matching scenario:

$$RMSE = \sqrt{\frac{\Sigma_{t=1}^{2}\left(\widehat{V_t} - V_t\right)^2}{n}} \qquad (16)$$

where $\widehat{V_t}$ is VHT for historical incident-free scenario, V_t is VHT for traffic scenario prior to the incident, and n is the number of observations.

The pattern recognition process is conducted on traffic conditions within a 30-

minute time frame prior to the incident. The number of observations is determined by both the interval selection and aggregation level of sensor data. The pattern recognition is essentially a heuristic search on historical database until the matching traffic scenario with the least RMSE is found. The recurrent delay within the incident's impact extent can thus be estimated as the accumulated delay from the matching scenario at the same location i and TOD DOW j but a different week K, expressed as:

$$D_{REC} = \Sigma_{j=1}^{J}\Sigma_{i=s_{m,j}}^{s_p} d_{his}(i,j,K) \tag{17}$$

Pattern recognition based on single VHT for the same TOD DOW at the same location may be subject to inaccuracy in providing a holistic view of traffic conditions. To compensate for this, we applied the KNN method in the pattern recognition process to determine the closest incident-free scenarios that can be used to describe recurrent congestion. KNN is a classification method that offers a nonparametric procedure for assigning a class label to the input pattern based on the K-closest neighbors of the vector *(64)*. In this study, we used similarity (RMSE) of the K-closest neighbors (historical scenarios) as the means of classification. The delay at matching scenario is calculated as:

$$d_{his}(i,j)_{KNN} = \frac{1}{K}\Sigma_{k=1}^{K} d_{his}(i,j,W_k) \tag{18}$$

where $d_{his}(i,j)_{KNN}$ is the mean of KNN recurrent delay, and W_k is the week when the KNN traffic scenario occurred. The robustness of VHT as measurement and value of K in KNN method are discussed in the next section.

The entire algorithm, deconstructed into three major components as described above, is depicted in Figure 10. Note that the congestion threshold estimation used in both Secondary Incident Identification and Total Delay Determination might bear two types of errors for incident spatiotemporal extent determination. First, when the actual recurrent delay is higher than the threshold, the incident spatiotemporal extent and the total delay would both be over-estimated. However, over-estimated spatiotemporal extent would also cause recurrent delay being over-estimated. The overall effect is canceled out when estimating IID. Second, when the actual recurrent delay is less than the threshold, the spatiotemporal extent is under-estimated. But in the region near the boundary of spatiotemporal extent, the impact of the incident is almost dismissed. Therefore, the delay in such a region is negligible.

3.3 Case Study

The proposed algorithm is applied onto the I-15 Northbound corridor between 15600 S and 1000 N in the Salt Lake Metropolitan area. This 25-mile long segment includes a total of 62 loop detector stations. The 2013 traffic data from loop detectors were retrieved and aggregated at 5-minute intervals, including speed, volume, delay, occupancy, and VHT. Incident records in 2013 were also retrieved from incident databases maintained at the Utah Department of Transportation (UDOT) Traffic Operation Center (TOC). Traffic data have been automatically collected and archived every 30 seconds, hosted in the PeMS (Freeway Performance Measurement System) database by the UDOT. PeMS aggregates the loop detector data at 5-minute interval with imputation parameters calculated offline. Regression is applied during imputation based on data from good loops that are spatiotemporally adjacent to the bad ones. When regression is not possible, cluster median is used to fill in the missing samples. The postprocessed traffic data set used in this study has good completeness and consistency. The incident dataset offers details regarding incident ID, time, location (milepost), duration, and brief description. It also provides incident characteristics, such as incident type (crash, debris, vehicle on fire, signal problem, etc.), severity (fatality and property damage), priority (lane blockage), and impact (incident clearance time estimated by emergency personnel). In 2013 there were 1,377 incidents that occurred in the selected segment. For the purpose of spatiotemporal analysis, traffic data up to 5 miles upstream from the starting point of the segment were also obtained. Delay is quantified as excessive VHT with a threshold of 60 mph.

To identify potential secondary incidents, spatiotemporal analysis was performed as mentioned earlier. A spatiotemporal boundary of 5 hours and 10 miles was preselected to accommodate the largest possible impact region of an incident. According to Khattak et al. *(65)*, the spatiotemporal boundary of secondary incident is usually within 2 hours and 2 miles. It is thus reasonable to assume that our spatiotemporal boundary is sufficient to capture all the associated incident impact. Note that the spatiotemporal boundary is set to ensure the computational efficiency of our algorithm. The resolution for spatiotemporal mapping was carefully chosen as $0.02\ miles \times 1\ minute$, which satisfies the accuracy of data without significantly overloading on computation. Imprialou et al. *(66)* pointed out that it is not necessary to use uniform temporal intervals in spatiotemporal mapping, yet a 1-minute interval was chosen in this study for simplicity. Traffic conditions between stations were estimated via interpolation of loop detector data. When conducting secondary incident identification, the 80th percentile of historical delay was used as the threshold to determine the prevailing congestion condition. This percentile was chosen based on random testing of delay distribution using the 2013 dataset and the result from previous studies related to secondary incident identification on freeways *(67)*. All the cascading incidents are mapped on the spatiotemporal extent, with secondary incidents further identified. The congestion threshold for total delay quantification was determined as described in Equations (8) and (9). The total delay was accumulated within the generated spatiotemporal extent. Recurrent delay was further estimated through the pattern recognition process by choosing the best matching KNN traffic pattern 30 minutes prior to the incident. Results and discussion of this implementation follow.

To validate the robustness of VHT as measurement index for traffic scenarios in pattern recognition, we compared the effectiveness of different measures, including VHT, speed, and volume, in predicting recurrent delay. To accomplish pattern recognition for each incident, we built a dataset of traffic pattern spans that are incident-free at the same TOD DOW as the incident scenarios for each incident. We randomly chose one span from database as the span whose traffic pattern was to be predicted. The rest of the spans were used as candidates for matching. A dataset with 800 incidents was used in this validation. Namely, 800 spans were chosen for prediction. VHT, speed, and volume were used as determination variables separately for different K-values ($\mathrm{K} < 10$). The RMSEs of delays were calculated to measure the robustness of different indicators. Table 3 shows the sum of RMSEs with different K values for KNN. It shows that KNN is more reliable than single value since the sum of RMSEs decreases as K increases. At lower K-value, speed outperforms the other two measures. With higher K-value, VHT is slightly better than speed, and both outperform volume. Overall, using relatively high K-value KNN and VHT as the determination variable can best predict recurrent delay. Therefore, we used VHT as determination variable and KNN method with $K = 9$ when processing pattern recognition.

Using the 2013 incident database (1,377 incidents in total) for the study corridor, a total of 109 primary incidents were identified with 270 secondary incidents. A total of 778 incidents were independent incidents, and 220 incidents (16%) were censored by the spatiotemporal boundary. These 220 incidents' spatiotemporal extents were beyond the 5-hour 10-mile maximum boundary set forth by the algorithm.

On average, the primary and secondary incidents were 3.2 miles and 70 minutes apart. Note that multiple consecutive secondary incidents are all considered to be traced from the original primary incident thereby resulting in an elongated time span. Figure 11 illustrates a secondary incident (23:40, MP 304, marked as 1) that occurred 3 miles upstream of the primary incident (23:19, MP 307, marked as P). Notice that another incident (23:32, MP 305, marked as 2) also appears in the vicinity. However, according to the spatiotemporal analysis, causality is not inferred.

Figure 12 shows the heat map and profiles of primary and secondary incidents along the study corridor. The profiles exhibit very similar trends with few exceptions, and the distribution of secondary incident is upstream skewed due to the hysteresis nature. Lag between the two ranges from 1 to 4 miles, which is consistent with the average distance reported. A reverse pattern appears in the segment between MP 298 and 300, where denser secondary incidents are induced by fewer primary incidents. This is to be expected due to the presence of a freeway junction between I-15 and I-215 that triggers more intensive weaving with an AADT of 77,000 vehicle/day. This can be contrasted with another junction between I-15 and I-80 with an AADT of 54,000 vehicle/day, which had an aligned incident occurrence pattern. The average IID is 43 vehicle-hours. IID distribution is right-skewed, indicating a small portion of incidents with extremely high

IID. For freeway management purposes, IID should be jointly studied to trace the reason behind their occurrence and for effective incident mitigation strategies. To this end, hot spot analysis is utilized to observe incident frequency along the corridor. Note that incident occurrence is usually not an isolated event. For example, Ord and Getis *(68)* proposed spatial statistics based on the weighted spatial autocorrelation between incidents. We applied a similar concept in the hot spot analysis: we considered not only the number of incidents at the spot but also the number of incidents associated to the spot. At each location, the occurrence of an incident is weighted by its independence. For example, an independent incident has the lowest weight since the occurrences of independent incidents are very random. A secondary incident has higher weight since it carries on the influence from primary incidents. A primary incident has the highest weight since it tends to induce more congestion and damage. Thus, each incident is weighted by the number of incidents it is associated with (including itself). For comparison purposes, the top 5 locations from each method were identified as hot spots. Figure 13 (a) and (b) show the hot spots identified by incident frequency with or without considering the weighting. Interestingly, they yield quite similar results with hot spots identified at two freeway junctions and between MP 295 and 297. However, when illustrating the spatial profile of IID as shown in Figure 13 (c), hot spots are clustered between MP 285 and MP 287, which is distant from the freeway junctions. One-way ANOVA with Tukey HSD test is conducted to evaluate the difference between the IID distributions across the entire corridor. Table 4 shows the result of the ANOVA test. It shows that the difference between the IID distributions at each milepost is significant. Table 5 shows part of the Tukey HSD test result. By conducting Tukey HSD test, we find that the difference between IID distributions at different MPs is insignificant, except for MP 285.

Several factors might contribute to this phenomenon. First, the existence of a bottleneck might exacerbate the impact of the incident, which might be one of the contributing factors for the extremely high IID. MP 285 is at the onramp of I-15 from a major arterial (Timpanogos Highway) where severe congestion is observed frequently. Evidence from a closer scrutiny of the spot validates the assumption. Most of the incidents happened during peak periods, which greatly impeded the queue clearance. Another reason might be the way IID is calculated as it solely considers the delay induced by independent incidents for accurately identifying their spatiotemporal extent. This might downplay the delay effect of cascading incidents. Therefore, the two hotspots analyses methods in this study complement each other and can be jointly used for decision making on incident mitigation. Note that the segment between MP 295 and 297 is identified as a hot spot in both methods. This may be due to the convoluted effects of multiple causes. This segment has an AADT as high as 100,000 vehicles/day and is located upstream of the spaghetti junction where triggered secondary incidents introduce great disturbances in traffic. Also the segment between MP 295 and 297 has the shortest distance between curvatures along the corridor. There are two curvatures that are less than 3 miles apart, which may cause instability in the traffic flow. This aligns with Zhang and Khattak's *(69)* finding that short segments are prone to secondary incidents.

Based on the analysis, we further conclude that locations with higher IID are prone to be bottlenecks that have severe recurrent congestion. When incidents occur at freeway junctions under heavy traffic volume, a significant increase in IID with induced secondary incidents upstream may occur. Freeway management strategies might be especially ripe for assessment based on this result. For example, when an incident occurs at a bottleneck, speed harmonization, such as variable speed limit, can be implemented upstream to accelerate bottleneck clearance and create a uniform speed upstream.

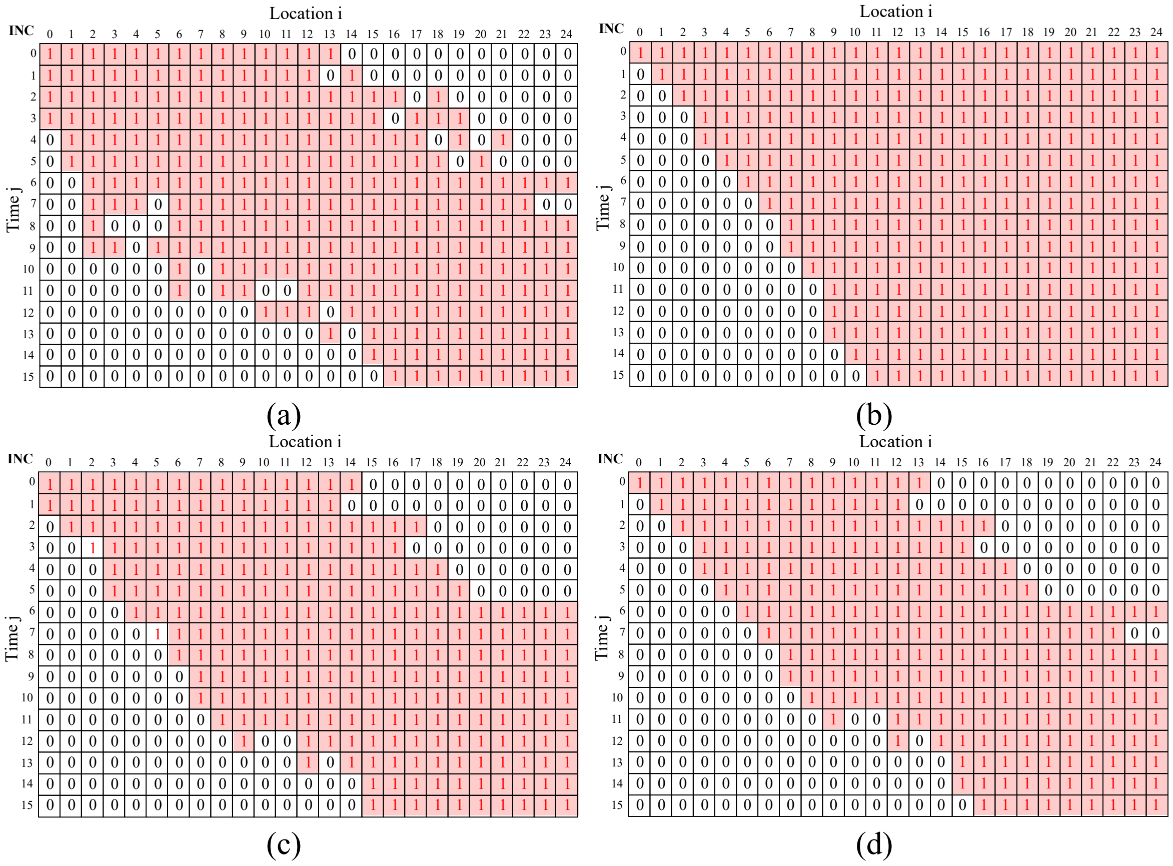
3.4 Summary

A systematic approach to quantify IID is proposed. The algorithm presents a three-component scheme: secondary incident identification, spatiotemporal extent determination, and recurrent congestion identification. This method is data-driven and spatiotemporal in nature to fully uncover the impact and causal mechanism of incident occurrence. IID quantification at the individual incident level will enable further analysis on delay-based behavior modeling and inspire follow-up research exploring relationships between the incident itself and its associated features (e.g., severity, lane blockage, or traffic conditions). Spatiotemporal analysis offers greater insight on the convoluted causes of delay and unveil the true impact of nonrecurrent congestions to explore unsettling reasons for safety enhancement that have escaped notice. This study complements the existing literature by combining analytical approach with data mining techniques to dynamically determine the spatiotemporal extent of individual incidents. The IID quantification methodology excludes the impact of secondary incidents for the first time and includes shockwave theory in spatiotemporal analysis. The information construction process can be further used to uncover a variety of features that are associated with particular incidents for an optimal freeway management.

This study contributes to the literature with two major highlights. The secondary incident identification, as a preprocessing for IID estimation, eliminates the mingled influences of subsequent incidents. Previous IID modeling ignored this critical step and oftentimes results in an overestimation of the impact of individual incidents. Our proposed method uses KNN pattern recognition, which essentially is a heuristic search process to separate the delay solely induced by incidents from the recurrent congestion. The algorithm is implemented based on data collected on the I-15 freeway corridor in Salt Lake City, Utah. A total of 109 primary incidents was identified with 270 secondary incidents. On average, the primary and secondary incidents were 3.2 miles and 70 minutes apart. The average IID of incidents was 43 vehicle-hours with the entire distribution right-skewed.

Hot spots analysis was conducted based on algorithm output. Two methods are demonstrated in such analysis: incident frequency-based with/without spatial correlation and IID based. The two hotspots analytics in this study complement each other and can be jointly used for incident mitigation and to inform investment decisions. Freeway management strategies might be especially ripe for assessment based on this result. For example, when an incident occurs at a bottleneck, speed harmonization, such as variable speed limit, can be implemented at the upstream to accelerate bottleneck clearance and create a uniform speed. The proposed framework is data-driven in nature for performance assessment of nonrecurrent congestion. It is self-adaptive to any data set and can be used to further uncover the relationship between the incident and associated features.

Based on this study, three intriguing topics emerge. First, as a follow-up research on the result, it is necessary to quantitatively model the identified features that are associated with IID. Second, it might be interesting to further explore the separate effect of primary and secondary incidents via simulation approach. Third, it is appealing to further explore the disaggregated impact of the primary incidents on the induced secondary incidents, particularly through likelihood estimation. All topics will drive more efficient strategic planning and project prioritization.



Location i

Time j

INC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0
4	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0
5	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0
6	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
8	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

(a)

Location i

Time j

INC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(b)

Location i

Time j

INC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
3	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
4	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
5	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
6	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

(c)

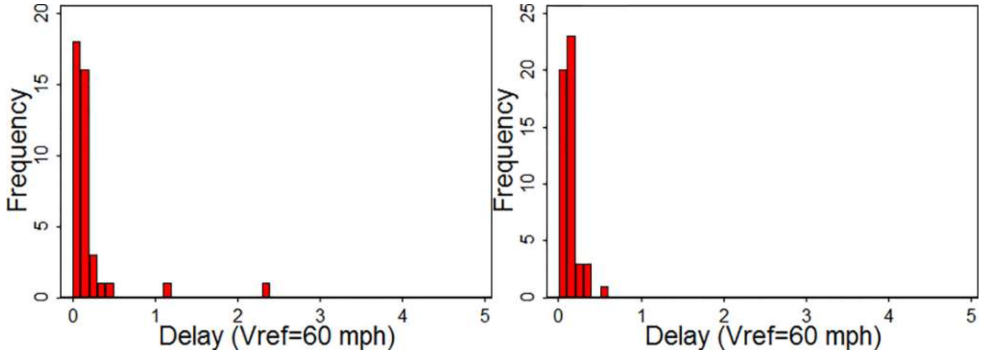
Location i

Time j

INC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
3	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
4	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
5	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
6	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
8	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

(d)

Figure 8 Illustration of Secondary Incident Identification Process: Spatiotemporal Profile of (a) Function I.; (b) Function S; (c) Function C; and (d) Function I*S*C



(a)

(b)

Figure 9 Typical Patterns of Delay Distributions: (a) Exponential Distribution (b) Rayleigh Distribution.

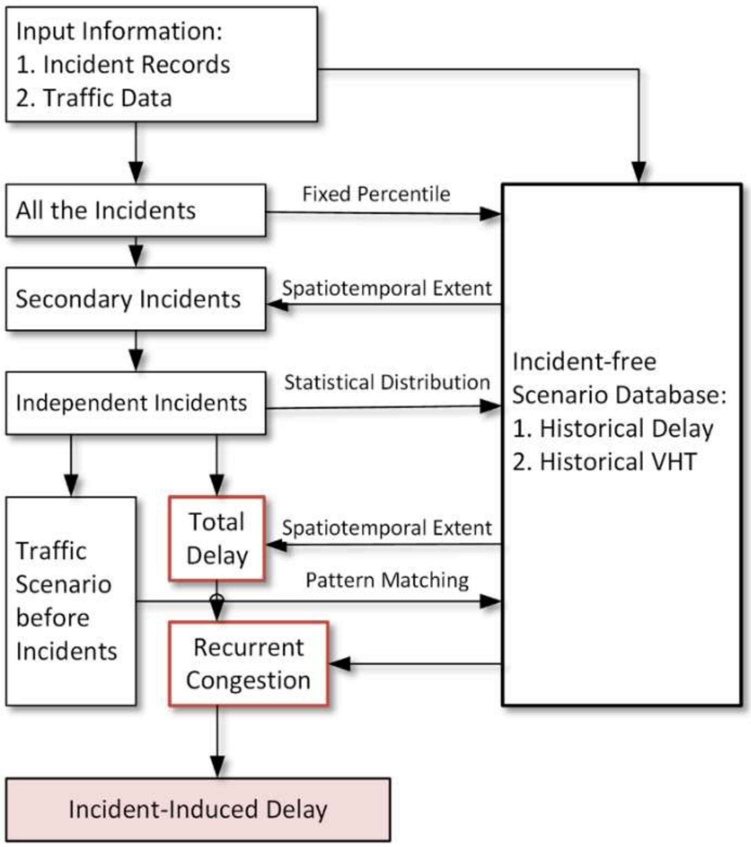


Figure 10 Illustration of Proposed IID Quantification Framework

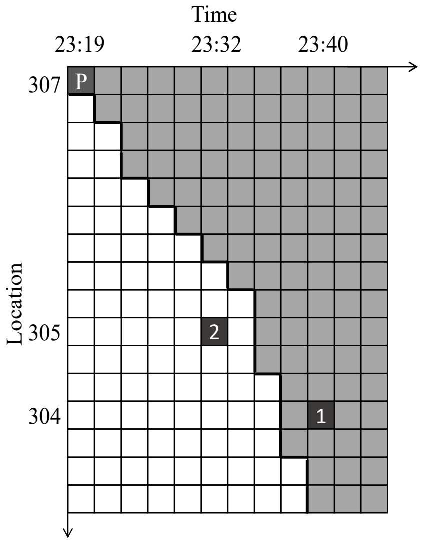
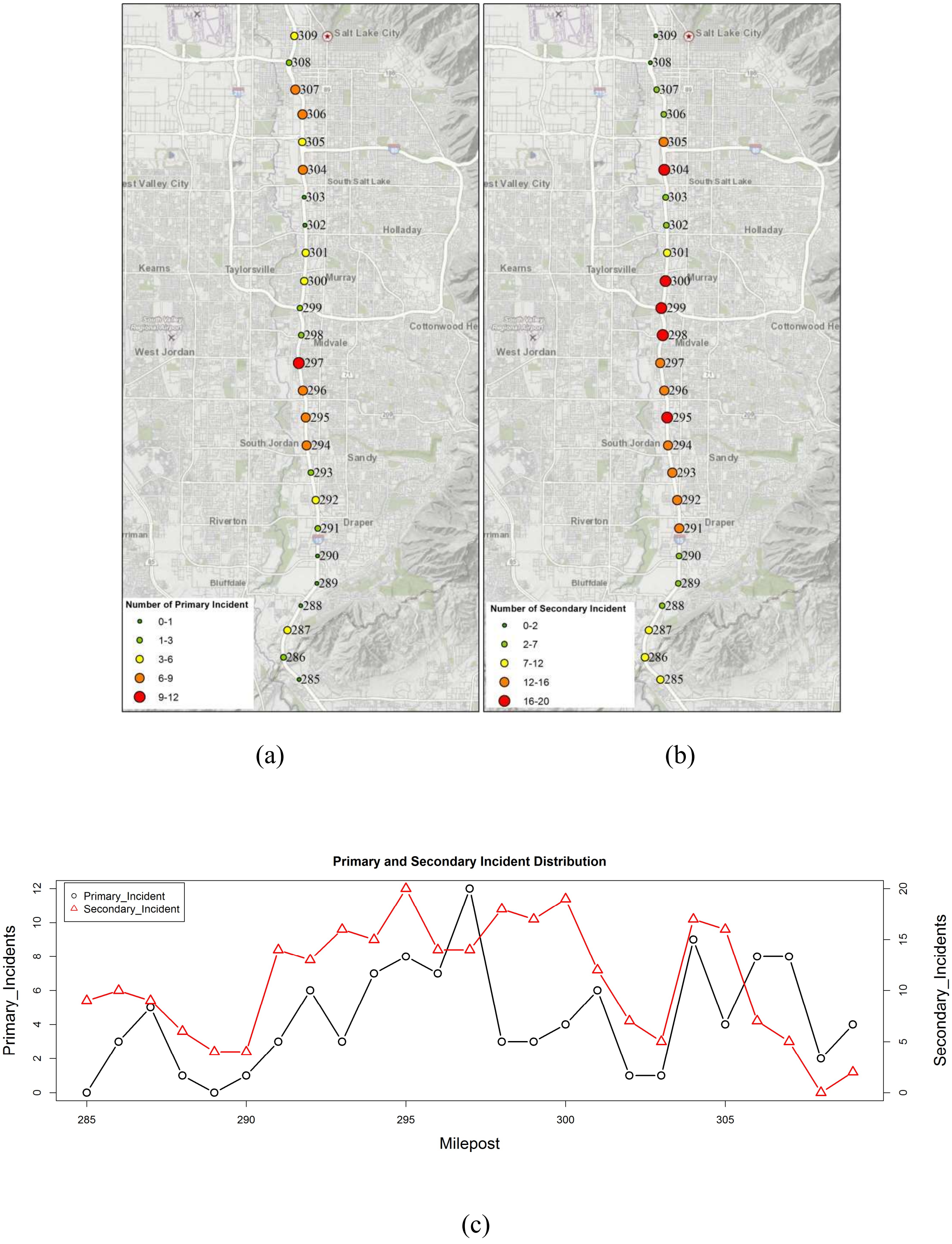


Figure 11 Example of Secondary Incident Identification (P: Primary Incident, 1: Secondary Incident, 2: Independent Incident, Grey: Spatiotemporal Extent)

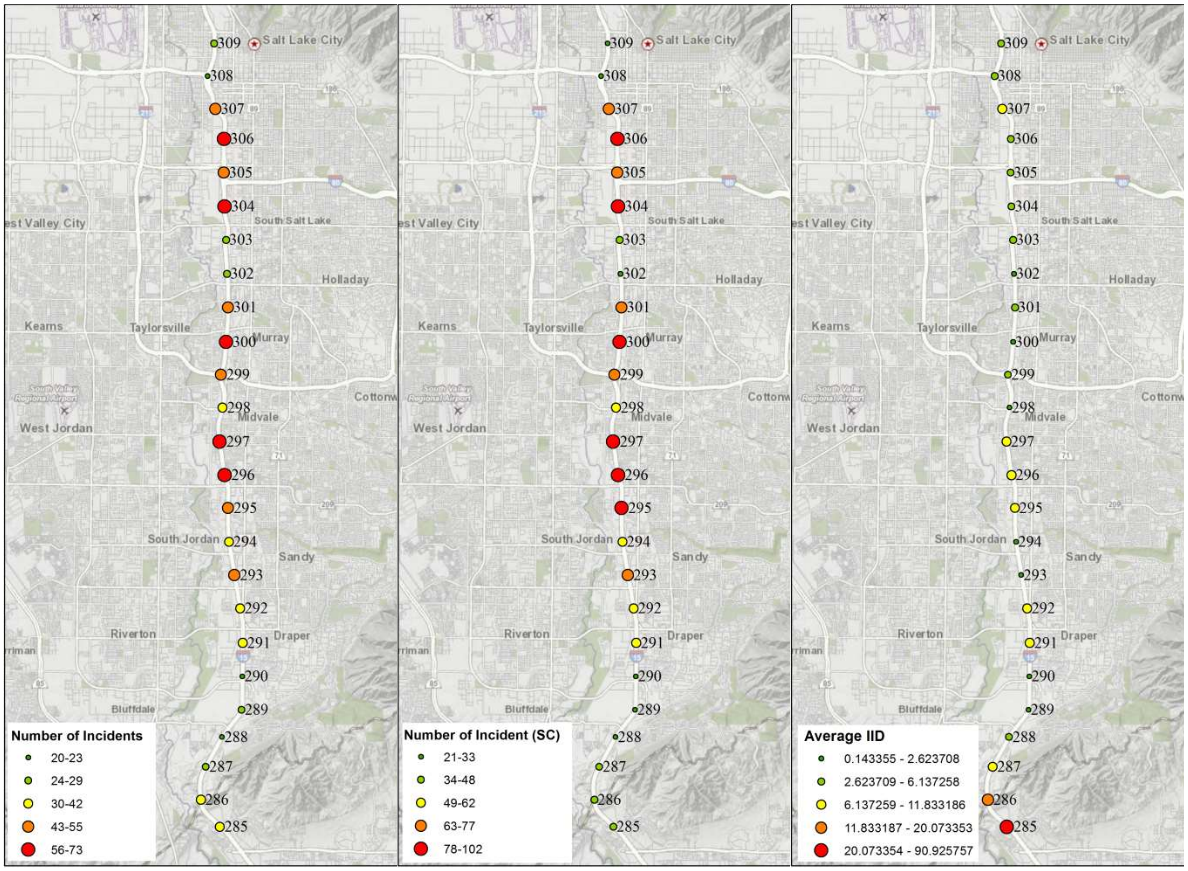


(a)

(b)

(c)

Figure 12 Heat Map (a) (b) and Profile of Primary and Secondary Incidents (c) Along the I-15 Corridor



(a) (b) (c)

Figure 13 Hot Spot Identification Analysis with (a) Incident Frequency Method Without Spatial-Correlation; (b) Incident Frequency Method with Spatial-Correlation; and (c) Average IID Method

Table 3 Sum of RMSE of Delay with Volume, Speed, VHT as Determination Variable When $K = 1,2,\ldots,9$

K	1	2	3	4	5	6	7	8	9
Volume	413.1	242.8	228.4	217.6	196.0	161.5	146.5	132.0	123.4
Speed	245.6	238.3	202.4	161.0	145.3	131.0	119.4	109.5	98.8
VHT	378.9	287.0	222.5	165.8	136.2	119.3	112.8	102.3	97.4

Table 4 ANOVA Test Result of IID Distributions Between MPs

	Degree of Freedom	Sum of Squares	Mean Square	F-Value	P-Value	Significantly Different, Yes/No
Between Features	1	1445	1444.5	31.86	2.48e-08	Yes
Within Features	625	29564	45.3			

Table 5 Tukey HSD Test Results (partial)

Group 1	Group 2	Meandiff	Lower	Upper	reject
285	286	335.806	482.117	189.496	TRUE
285	287	387.073	550.835	223.311	TRUE
285	288	408.295	575.724	240.865	TRUE
285	289	427.484	573.795	281.174	TRUE
285	290	420.854	575.72	265.988	TRUE
285	291	391.451	541.671	241.23	TRUE
285	292	391.813	536.39	247.235	TRUE
285	293	418.631	551.985	285.278	TRUE

CHAPTER 4

MAINTENANCE INFRASTRUCTURE SAMPLING

In this chapter, a high-dimensional clustering-based stratified sampling (HDCSS) method for infrastructure inspection is presented. The proposed method integrates infrastructure deterioration prediction, high-dimensional cluster analysis, and Locality-Sensitive Hashing. This chapter is organized as follows： The first section summarizes previous studies on maintenance infrastructure sampling methods and high-dimensional clustering. The sampling method is described in the second section. The third section presents a numerical test of the sampling method with data collected on freeway network from Utah. The fourth section concludes this study with direction for future research.

4.1 Literature Review

Previous studies on infrastructure sampling and high-dimensional clustering will be discussed in the following section. The proposed method and its mechanism are explained later and followed by an application of highway segment sampling with infrastructure data collected by the UDOT. Results and implications are discussed at the end. In the stream of research on infrastructure management, inspection is usually jointly studied with maintenance, within the scope of optimal infrastructure management *(70–73)*. One of the most widely used infrastructure maintenance optimization algorithms is Latent Markov Decision Process (LMDP). In LMDP, infrastructure conditions are represented with a set of discrete states, and the deterioration process is encoded as Markovian transition matrix with probabilities. It considers uncertainty introduced by infrastructure performance prediction and measurement *(74)*. Output of the method is the optimal inspection and maintenance policies. LMDP aims to minimize total managing cost or maximize infrastructure performance over finite/infinite planning horizon. A lot of efforts have been made along the line to refine LMDP method since it was first proposed *(16, 70, 75)*. Mishalani and Gong *(70)* extended LMDP model to a network-level problem by including network-level constraints, such as allowed fraction of infrastructures in best or worst condition and yearly expenditures within a specified budget. Medina et al. *(75)* improved LMDP method with adaptive control formulations. Instead of using one Markov Decision Process (MDP) to represent the infrastructure performance and transition, their optimization model used finite mixtures of MDPs. Also in their method, infrastructure deterioration process was constantly updated with feed of new condition measurements. Guillaumot et al. *(16, 71)* incorporated uncertainties from sampling (i.e., sample size, spatial sampling) and inspection process, and used them as decision variables in LMDP models. Maintenance activity on each segment (repair, inspection, or do nothing) was optimized based on potential infrastructure conditions if such maintenance activities were conducted. Notice that all these aforementioned approaches discussed uncertainty from sampling in terms of sample size rather than sampling method. Stratified sampling is a classic sampling method in transportation maintenance management since it balances tradeoffs between inspection costs and sampling accuracy *(76–78)*. Steinbach et al. *(78)* proposed a stratified sampling method for road maintenance evaluation. The stratification criteria include geographical location, weather variation, urban and rural setting, and traffic volume. Garza et al. *(76)* evaluated the effectiveness of a stratified random sampling method for transportation infrastructure. They pointed out that by employing stratified sampling techniques rather than the SRS method, agencies can reduce sample size or greatly improve precision. Bellman *(77)* proposed a sampling protocol which stratifies the population with functional classification, AADT range, and infrastructure category. In most previous studies, segments were stratified with features of road segments rather than the features of the infrastructures. In such cases, stratification may produce large bias if the infrastructures are unevenly distributed across segments.

In our proposed method, stratification is implemented via high-dimensional cluster analysis. Since each highway segment often contains multiple infrastructures, we consider a segment as a high-dimensional vector and each type of infrastructure as one dimension of that vector. By applying high-dimensional cluster analysis, we divide all segments into several clusters based on their infrastructures' conditions. The challenge in dealing with high-dimensional data lies in the Curse of Dimensionality. The concept is originally defined by *(79)*, referring to the difficulty of optimizing a multivariable function within the multidimensional context. In cluster analysis, as dimensionality increases, the number of data points within each dimension becomes increasingly "sparse" *(80)*. As illustrated in Figure 14, a dataset with 10 points is randomly distributed from 0 to 1 in one-dimensional space. The points are in close vicinity of each other. There are four points within the range [0, 0.5]. But when the dataset is expanded to two dimensions, if we still use 0.5 as the discretization unit in each dimension, there are then only 3 points in the range of [0, 0.5] in each dimension. When we further expand the dataset to three dimensions, there are only 2 points within the same unit. So for high-dimensional data, distance may no longer be effective to distinguish points and most cluster techniques applicable to low-dimension data (e.g. centroid-based clustering, density-based clustering) are rendered meaningless.

During the past decades, much effort has been devoted to avoiding the Curse of Dimensionality. One approach to high-dimensional clustering is to develop new measurements for distance or similarity across clusters, including grid *(81)*, sum of similarities along dimensions *(82)*, and approximate similarity *(83)*. Charikar *(83)* also suggested a practical similarity measurement called Locality-Sensitive Hashing (LSH). LSH is a widely-used algorithm to search similarity between high-dimensional data for fast indexing and database searching. LSH maps high-dimensional data points to a low-dimensional space by applying hash functions. As mentioned in *(84)*, a hash function family $H = \{h_1, h_2, h_3, \ldots, h_i, \ldots\}$ is called (d_1, d_2, p_1, p_2) sensitive for any two high-dimensional vectors q and v:

- if $\mathrm{D}(q, v) \leq \mathrm{d}_1$, then $P_H[\mathrm{h_i}(q) = \mathrm{h_i}(v)] \geq \mathrm{p}_1$
- if $\mathrm{D}(q, v) > \mathrm{d}_2$, then $P_H[h_i(q) = h_i(v)] \leq p_2$

where d_1 and d_2 are the critical distances to determine if q and v are similar, p_1 and p_2 are the critical probabilities, and D is the distance measurement in the low-dimensional space. If the distance between the mapped values is less than d_1, then the probability that q and v are similar is greater than p_1. On the contrary, if the distance between mapped values is greater than d_2, then the probability that q and v are similar is less than p_2. Based on such a definition, researchers proposed different function schemes and validated their reliability in capturing the underlying similarity, including inner product *(85)*, learned Mahalanobis distance *(86)*, and normalized kernel function *(87)*.

4.2 Methodology

The stratification in the proposed sampling method, as illustrated in Figure 15, consists of two major components: current condition estimation and high-dimensional cluster analysis. Current condition estimation "predicts" the infrastructure condition (e.g., in the form of LOM) based on historical records. This is to ensure that for the next round of inspection, sampling is conducted based on previous inspection results and deterioration rate of the infrastructure. High-dimensional cluster analysis then divides segments into clusters and selects representative segments as samples. Segments within each cluster share similar pattern with regard to infrastructure conditions. Thus by selecting segments across clusters, we select representative samples across all patterns. The sample size is a fixed percentage of segments in the network, constrained by labor or budget limits. Segments within each cluster are chosen randomly. Once the sampled segments are inspected, maintenance and rehabilitation (M&R) activities can be further conducted accordingly on those segments whose performance is below a certain threshold. The M&R records and inspection results will be applied to the next round of sampling process for inspection.

4.2.1 Current Condition Estimation

As the sampling unit for maintenance activities, segment possesses multiple features: infrastructure facilities (shoulder work, litter, weed, sweeping, etc.), geometric characteristics (number of lanes, segment length, etc.), and traffic information (AADT, peak hour volume, etc.), just to name a few. Each segment can therefore be described as a high-dimensional vector:

$$S_n = \{a_{shoulder\ work}, a_{litter\ pickup}, a_{weed}, \dots ;\ g_{leng} \quad , g_{lane\_num}, \dots ; t_{AADT}, t_{Peak\_Vol}, \dots\}$$

where S_n refers to the segment n, $1 \leq n \leq N$, N is the number of segments in the network, and a, g, and t refer to the features associated with infrastructure, geometric, and traffic, separately. In this paper, we will only consider infrastructure type and condition as segment features as they are the focus for sample selection.

The current condition estimation starts with translating the deterioration process of infrastructure on the segments into a deterioration matrix. The infrastructure conditions are described using 15 letter scores from A+ to F. A+ represents the best condition and F the worst. In previous studies, infrastructure deterioration has been considered as a linear *(88)* or nonlinear *(89)* process. We assume that infrastructure deterioration is a linear process, yet the rate may vary across different types of infrastructures or different segments. For example, on Segment 1, the time during which segment's shoulder condition deteriorates from A to A is the same as the time from A to B+. Yet in the meantime, the condition of littering might deteriorate from A to C. And on Segment 2, while the shoulder deteriorates from A to A on Segment 1, the shoulder condition might deteriorate from A to B+.

The deterioration matrix is constructed based on the paired consecutive inspection records without any intervention (e.g., M&R) in between. In this study, we filtered out all the consecutive inspection records whose latter result was better than the former one. Yet exceptions might occur when the asset might still deteriorate to a worse condition even after repair or maintenance, in which case this method might underestimate deterioration rates of the infrastructures.

To simplify calculations, the 15 letter grades of infrastructures from A+ of F are converted to numerical scores of 15 to 1. The deterioration process thus can be considered as the score is decreasing as time goes by. The deterioration rate of a segment is calculated as the score difference divided by time duration between two inspections.

The historical records in our study span years during which segments may be inspected multiple times. Therefore, some segments might have more than one pair of consecutive records, producing different historical deterioration rates. In such cases, average of historical deterioration rates is employed as the deteriorate rate of the segment. For segments without such prior records, deterioration rate is replaced with the network averaged value. For example, if no consecutive record for shoulder work is available on one segment, the deterioration rate of that segment is replaced with average shoulder work deterioration rate of all segments. Deterioration matrix is constructed as:

$$D = \begin{bmatrix} d_{seg1\_ShoulderWork}, & d_{seg1\_LitterPickup,} & d_{seg1\_IceSnow}, & \cdots \\ d_{seg2\_ShoulderWork}, & d_{seg2\_LitterPickup}, & d_{seg2\_IceSnow}, & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ d_{segN\_ShoulderWork}, & d_{segN\_LitrerPickup}, & d_{segN\_IceSnow}, & \cdots \end{bmatrix} \quad (19)$$

D can always be updated with the latest inspection results and maintenance activities.

With the deterioration matrix constructed, we can estimate current conditions of infrastructures on each segment. Previous condition of infrastructure network is expressed as:

$$M_{Previous} = (S_1, S_2, \dots, S_N)^T \quad (20)$$

where M represents the previous network infrastructure inspection conditions, S represents the infrastructure conditions within the segment, with:

$$S_i = (s_{i\_ShoulderWork}, s_{i\_LitterPickup}, s_{i\_IceSnow}, \dots) \quad (21)$$

The current conditions of infrastructure are estimated by considering previous conditions and inspection frequency, which is expressed as:

$$M_{Current} = M_{Previous} + tD \quad (22)$$

where $M_{Current}$ is the estimated current network infrastructure conditions, and t is time duration between previous and current inspections.

4.2.2 High-dimensional Cluster Analysis

The key of high-dimensional cluster analysis is to jointly analyze all the infrastructure conditions on a segment rather than examining them individually. With the current infrastructure conditions estimated, LSH is implemented to define the similarity between segments. All segments are then divided into clusters based on the similarity matrix via spectral clustering. A fixed percentage of the segments can then be randomly chosen from each cluster.

The input to high-dimensional cluster analysis is the estimated current infrastructure conditions, including:

$$M_{current} = (S_1^*, S_2^*, \dots, S_N^*)^T \tag{23}$$

$$\mathrm{S}_{\mathrm{i}}^* = \{s_{i\_ShoulderWork}^*, s_{i\_LitterPickup}^*, s_{i\_IceSnow}^*, \dots\} \tag{24}$$

where S_i^* represents the estimated current infrastructure conditions on Segment i.

The first step in LSH is to define hash functions. In this study, we use inner product hash functions proposed by Kulis and Grauman *(85)*. Hash function transforms a k-dimensional segment into a binary string. For example, in Figure 16, each segment is transformed into 8digit binary strings. To determine the first digit of a binary string, we pick a k-dimensional vector $\boldsymbol{r} = (r_1, r_2, r_3, \dots, r_k)$. Each dimension (r_1 , r_2 ,…) in vector $\boldsymbol{r}$ is randomly generated following Gaussian distribution. Then we calculate the inner product between the segment and r as:

$$h = \boldsymbol{r} \cdot S^* = r_1 s_{ShoulderWork}^* + r_2 s_{LitterPickup}^* + r_3 s_{IceSnow}^* + \cdots \tag{25}$$

where h is the inner product. When h is greater than or equal to 0, the first digit of the binary string is 1, and 0 otherwise. By repeating the process eight times, an 8-digit binary string is constructed. The binary strings are the hash keys of a hash function family. The same process applies to all segments, with each segment assigned a hash key. Note that some segments may have the same hash keys.

Since the hash keys are binary strings, we use Hamming distance as the difference measurement to compare them *(90)*. For two strings with equal length, Hamming distance is defined as the number of digits at which the corresponding symbols are different. As illustrated in Figure 17, Hamming distance between the two strings is 1.

When the difference between two segments' hash keys is less than a certain threshold, it is referred to as "collision" between the two segments and they are considered similar. As illustrated in Figure 16, the hash key of Segment 1 is 01100111, and the hash key of Segment N is 11100111. Hamming distance between the two hash keys is 1. If the threshold to define collision is set at 2, the difference between Segment 1 and N's hash keys fulfills the requirement and thus the two segments are deemed similar.

Until this step, the LSH algorithm is fully implemented. Yet the algorithm can only determine whether two segments are similar or not rather than quantifying such similarity. Considering that hash function utilizes randomly generated vectors, using different vectors would lead to different hash keys. In Figure 16's example, Segment 1 and Segment N are considered similar. But if we generate another eight vectors, there is a probability that Segment 1 and Segment N are no longer similar. To remedy this, we perform LSH algorithm multiple times (i.e., 300 runs), and define similarity as the probability that two segments are similar across all the runs. For example, if Segment 1 and Segment N are identified as similar for 240 times out of 300 times, the similarity between them is $240/300 = 0.8$. With the similarity between each pair of segments in the network quantified, a matrix of similarity $\mathbb{S} = [\mathrm{Sim}_{\mathrm{ij}}]$is constructed, where $\mathrm{Sim}_{\mathrm{ij}}$ represents the similarity between segments i and j.

Then we apply spectral clustering, which is one of the most popular clustering algorithms due to its simplicity and efficiency *(91)*. It originates from partitioning clustering, which gives weights to links between data points and divides clusters by removing the least weighted links between clusters. Spectral clustering combines partitioning clustering with graph Laplacian matrices. The calculation is based on the spectrum of similarity matrix. The detailed computation is available in the Appendix.

4.3 Case Study

We implemented the proposed method on highway infrastructure inspection record provided by the Utah Maintenance Management Quality Assurance (MMQA) Program. Previously, MMQA performed full inventory inspections for infrastructure maintenance. The maintenance personnel recorded total numbers of infrastructures to be maintained and deficient infrastructures on each segment. Then inspection records were entered into the MMQA+ software to calculate the LOM (letter grade). One motivation to develop an infrastructure sampling method is to reduce costs of infrastructure inspection by estimating the overall network LOM on a sample basis. For the state of Utah, the entire highway network is divided into 489 segments. Inspection was performed semiannually from September 2014 to March 2016, with several segments inspected multiple times within one inspection period. The inspection record achieves overall infrastructure condition, as well as segment id, infrastructure type, inspection date, and deficiency locations. With more than 7,000 records in the database, 14 types of infrastructures are used in our study, including Shoulder Work (SW), Curb & Gutter (CG), Litter Pickup (LP), Weed Control (WC), Grade & Clean Ditches (GCD), Maintain Inlets (MI), Erosion Repair (ER), Pavement Markings (PM), Repair & Replace Signs (RRS), Repair & Replace Delineation (RRD), Guardrail Maintenance (GM), Sweeping (SP), Vegetation Control (VC), and Fence Maintenance (FM). Table 6 shows the average deterioration rate for each infrastructure. Note that the table only shows the aggregated (averaged) deterioration rates of all infrastructures. For example, the average deterioration rate of CG is 0.0996. It means that the conditions of CG deteriorate by 1 level (from A+ to A, or from A to A) in approximately10 months on average. Yet on individual segments, the rate can be different. For example, the deterioration rate of CG on some segments can be 0.2, indicating that it takes 5 months for CG to deteriorate from A+ to A.

In high-dimensional cluster analysis, we use a 14-digit binary string as the hash key. The “collision” threshold is set at 2, indicating that when the Hamming distance between the hash keys of the two segments is less than 2, those two segments are similar. To avoid too many or few segments in each cluster, all segments were divided into 10 clusters. For comparison purposes, SRS is also conducted. In the following section, both methods have been performed 50 times for sensitivity analysis.

The purpose of infrastructure inspection is to assess infrastructure conditions and report LOMs of overall highway network for investment decisions. Ideally, infrastructures’ grade (LOM) distribution, measured from samples, can reflect both overall condition and condition variation. To assess the effectiveness of our sampling method, the difference between condition estimated from samples and full inventory is computed with RMSE. For any infrastructure, the letter grade distribution is expressed as $(X_{A+}, X_A, X_{A-}, \dots, X_{F-})$, where X_i is the actual percentage of grade i in the full inventory (all segments). The grade distribution estimated from the sample is expressed as $(x_{A+}, x_A, x_{A-}, \dots, x_{F-})$, where x_i is the estimated percentage of grade i among all the sampled segments. The RMSE between estimated (from sample) and ground-truth grade distributions is then calculated as:

$$\text{RMSE} = \sqrt{\frac{\Sigma(x_i - X_i)^2}{15}} \tag{26}$$

RMSE reflects the error induced during the sampling process. As the value increases, the estimated condition deviates from the ground truth. To compare the performance of our proposed sampling method and the SRS method, we conducted experiments of estimation accuracy between the two methods with the full inventory data collected by UDOT. We chose the most recent inspection records of each segment as the infrastructure conditions to be sampled and inspected, and the second most recent inspection records as the historical record based on which to estimate the current infrastructure conditions. To validate the robustness of the sampling method, particularly, its sensitivity to different data dimensionalities, a series of sensitivity tests were performed. Using the same highway network, sampling was conducted with 6, 8, 10 and 14 different types of infrastructures, separately. The types of infrastructures were randomly selected when the number of types was less than 14. Figure 18 shows the average RMSEs when sampling is conducted based on sample rates ranging from 5% to

30% of the entire segment inventory. The sampling rate for the proposed sampling method refers to the percentage of samples chosen from each stratum. Since the number of samples in each stratum is rounded to the nearest integer, the number of samples is always less than the same rate of the entire population. For example, when the sample rate is 10%, there are about 49 segments chosen as samples from the network with 489 segments. But in reality, the total number of segments chosen is less than 49. To make sure that the sampling methods are compared based on the same sample rate, the sample size of SRS method is the same as the number of segments chosen by the stratified sampling method. It is noted that for low-dimensional (less than 10 types of infrastructures) data, the average RMSEs show no significant difference when dimensionality changes. But when the dataset becomes high-dimensional (more than 10 types of infrastructures), the average RMSEs start to demonstrate improvements. It further validates the effectiveness and suitability of our proposed sampling method for high-dimensional clustering analysis. The LSH algorithm is designed for data from high-dimensional space where the Euclidean distance is no longer valid as a similarity measurement. As dimensionality increases, the proposed method tends to provide a more accurate LOM estimation of the overall infrastructure condition.

Figure 19(a) shows sensitivity analysis of sample size. Note that the RMSE is averaged out both across grades and across infrastructures. As shown in Figure 19, there is a tradeoff between accuracy and sampling rate. Both average RMSE and standard deviations of RMSE decrease as the sampling rate increases. We observe a clear cutoff point at around 20% sampling rate, where the RMSE drops significantly as the sampling rate increases to 20%. After that, the trend becomes mild. The HDCSS method constantly outperforms SRS by providing lower average RMSE. Figure 19(b-d) show the RMSE distributions with sample size of 6%, 8%, and 10%. When the sample size is less than 10%, there is a distinct difference between the performances of two sampling methods.

One highlight of the proposed method is that the selected sample segments can accurately reflect the LOMs of all the infrastructures throughout the network. Figure 20 provides a detailed look on the sampling accuracy for each asset using SRS and our proposed method, where RMSE (mean and standard deviation) is shown for all 14 assets with a 20% sampling rate.

As seen in Figure 20, the RMSEs are similar between two methods but vary significantly across infrastructures. For most types of infrastructures, SRS has higher RMSE than the proposed method, indicating the superiority of our proposed method. However, also note that for certain infrastructures (WC, RRD, SP, and VC), SRS yields lower RMSE. To further explore the reasons, we compared the LOM distributions between each type of these infrastructures.

Figure 21(a) shows the ground truth grade distributions of WC, RRD, SP, and VC. For these infrastructures, more than 80% of segments are of A+ grade. Under such circumstances, since the difference between individual samples is insignificant, it is highly likely that choosing different samples would not influence the result much. An extreme case in such a situation is that if all the segments are of grade A+, then samples selected by any method would yield the same result. For infrastructures with such skewed grade distribution, both methods would estimate the overall conditions with low errors. Yet one unique aspect of high-dimensional cluster analysis is that when one dimension in the high-dimensional vectors lacks variation, clustering relies more on other dimensions, and the importance (weight) of that dimension thus diminishes. Correspondingly, the overall condition of that infrastructure (with little variation) is less represented by the sample selected by HDCSS than randomly picked. That explains the underlying reason for the low RMSEs for the four types of infrastructures and the outperformance of SRS for them.

The relation between the LOM distribution and RMSE is reflected in the ranking of RMSE values of these four infrastructures. As shown in Figure 21(a), the four infrastructures, ranked by the percentage of grade A+ for each type in descending order, are SP, VC, WC, and RRD. This sequence is exactly the same as the RMSE ranking using both methods.

Another interesting phenomenon observed from Figure 20 is that the average RMSEs of two infrastructures, FM and RRD yield same result with our proposed method. However, the values are quite different with SRS. In Figure 20(a), it is shown that for FM, almost 80% segments are either of grade A+ or grade F. Thus, the entire grade distribution is quite dispersed due to the occurrence of two dominant grades. In such case, the HDCSS method selects samples from both dominant grades yet such a scenario is not guaranteed with SRS. As the distribution shifted to a single peak instead of two (see Figure 8(b) for the comparison), the RMSE of SRS increases from around 1.5% to 1.9%. As shown in Figure 8 (b), with the other grades remaining at a very low percentage, two dominant grades have significantly higher percentages and those two percentages are comparable.

In infrastructure inspection sampling, one prominent concern is to reduce the sample rate without too much compromise in accuracy, since the sampling rate is directly tied to costs and budget allocation. According to *(92)*, lead states assume that the infrastructure conditions follow normal distributions, so the sampling rate can be estimated with given confidence interval and accuracy. For example, North Carolina DOT performs sampling based on 90 to 95% confidence interval and 6% accuracy. Virginia DOT requires the confidence interval of sampling be at 95% with an accuracy of 4%. However, there is a lack of evidence to justify the assumption. For comparison purposes, we define an "accuracy rate" for each method, representing the probability of a sample being considered as accurate within certain error threshold. The sampling result is considered "accurate" if and only if the errors between the estimated conditions of all assets and ground truth are within an acceptable range. The error is still quantified via RMSE.

Figure 22 shows the sensitivity analysis of accuracy rate when different sample sizes apply. It is noted that under the same error threshold, when the sampling rate is less than 20%, our proposed method always yields a higher accuracy rate than SRS. For an accuracy rate of 90% with an error threshold of 0.06, the required sampling rate is around 8% for our method as opposed to 10% for SRS. To achieve an accuracy rate of 95% with an error threshold of 0.4, by using the HDCSS method, the sample size can be reduced from 20% to 16%. Such a decreased sample size can bring a significant reduction in inspection costs for infrastructure management, especially for large scale highway network.

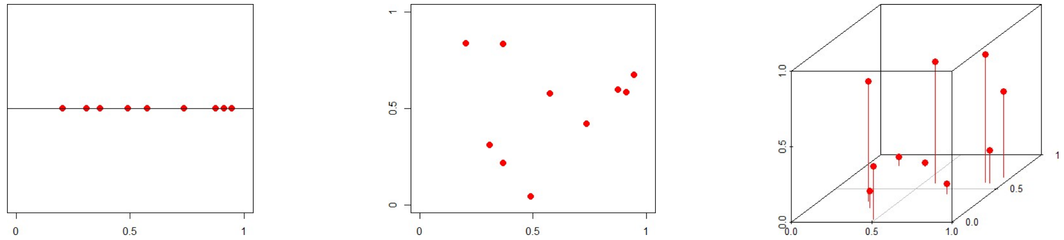
To further explore the sample rate reduction quantitatively, we performed a one-way ANOVA test to analyze the difference of errors between the two sampling methods with different sample rates. The results of ANOVA tests show how the sample size changes with different sampling methods when there is no significant difference between the sampling results. Table 7 shows the result of the ANOVA test between errors of samples selected by HDCSS method with sample rate of 16% and SRS method with a sample rate of 18%.

The results of ANOVA conclude that there is no significant difference between the errors of estimated infrastructure conditions by using the proposed method with sample rate of 16% and SRS with sample rate of 18%. That is to say, for any ongoing sampling scheme using the SRS method with 18% of the population as the sample rate, our method can effectively reduce the sampling rate to 16%. Similar sample rate reduction results have been observed under other precision requirements, as shown in Table 8. It is observed that when the sample rate of SRS is below 15%, our proposed method can reduce the sample rate by 1%. When the sample rate SRS is above 15%, it can reduce the sample rate by 2%. And most notably, when the sampling method is applied to large highway networks, these reductions in sample rates can significantly reduce the inspection costs.

4.3 Summary

A HDCSS method is proposed. The sampling segments selected by this method can accurately represent the overall conditions of the full infrastructure inventory. Our proposed method generally outperforms SRS method, which is widely used by DOTs. The method consists of two components: current condition estimation and high-dimensional cluster analysis. The current condition estimation aims at providing predicated infrastructure condition for cluster analysis based on historical inspection records. In high-dimensional cluster analysis, segments with multiple types of infrastructures are considered as high-dimensional vectors. By applying the Locality-Sensitive Hashing algorithm and spectral clustering, the similarity of the segments is measured and the segments are assigned to clusters. Using the inspection records from the State of Utah, our proposed method outperforms SRS for most types of infrastructures, especially under the circumstances where LOM varies greatly within infrastructures. For the infrastructures when most of the segments are of similar conditions, both the information-based sampling method and SRS yield low errors. The method can effectively reduce the sample rate without compromise in accuracy compared with the SRS method, leading to significant decrease in inspection costs, especially for large scale networks.

By using the proposed sampling method, DOTs can save resources and time for infrastructure inspection, due to the fact that inspection is carried out on the segment basis and the similarity identification introduced through the LSH algorithm. The method can be further applied in any high-dimensional sampling process of selecting corridor segments, intersections, or traffic infrastructures where multiple types of features, e.g., traffic conditions, geometric design, infrastructures, need to be considered. Based on this study, two intriguing topics emerge. First, as an important component of the method, deterioration matrix construction can significantly influence the accuracy of the sampling method. It is necessary to apply a more rigorous data analysis tool to enhance the estimation of the deterioration process. Second, it might be interesting to involve other more efficient high-dimensional cluster analysis methods in the sampling process which can potentially improve the accuracy of the sampling results.



(a) (b) (c)

Figure 14 Illustration of Sparsely Distributed Data Points due to Curse of Dimensionality

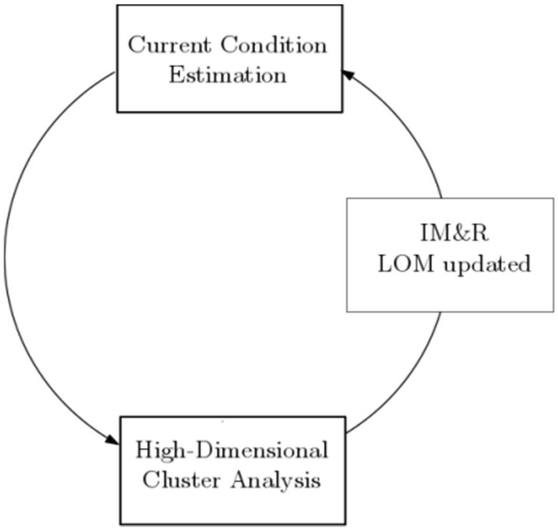

Current Condition Estimation
IM&R LOM updated
High-Dimensional Cluster Analysis

Figure 15 Illustration of Stratification in the Proposed Method

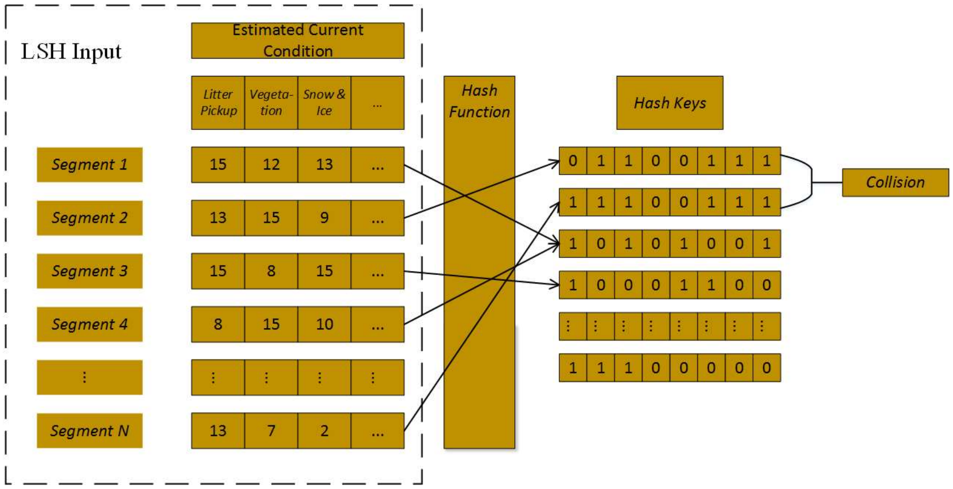


Figure 16 Illustration of LSH Process

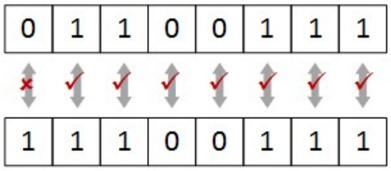


Figure 17 Illustration of Hamming Distance

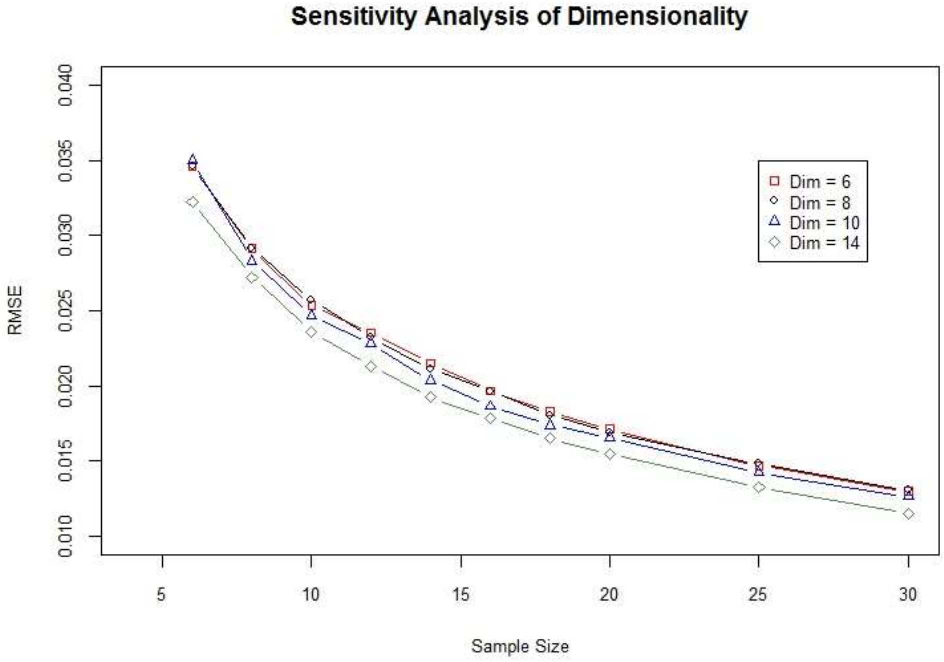
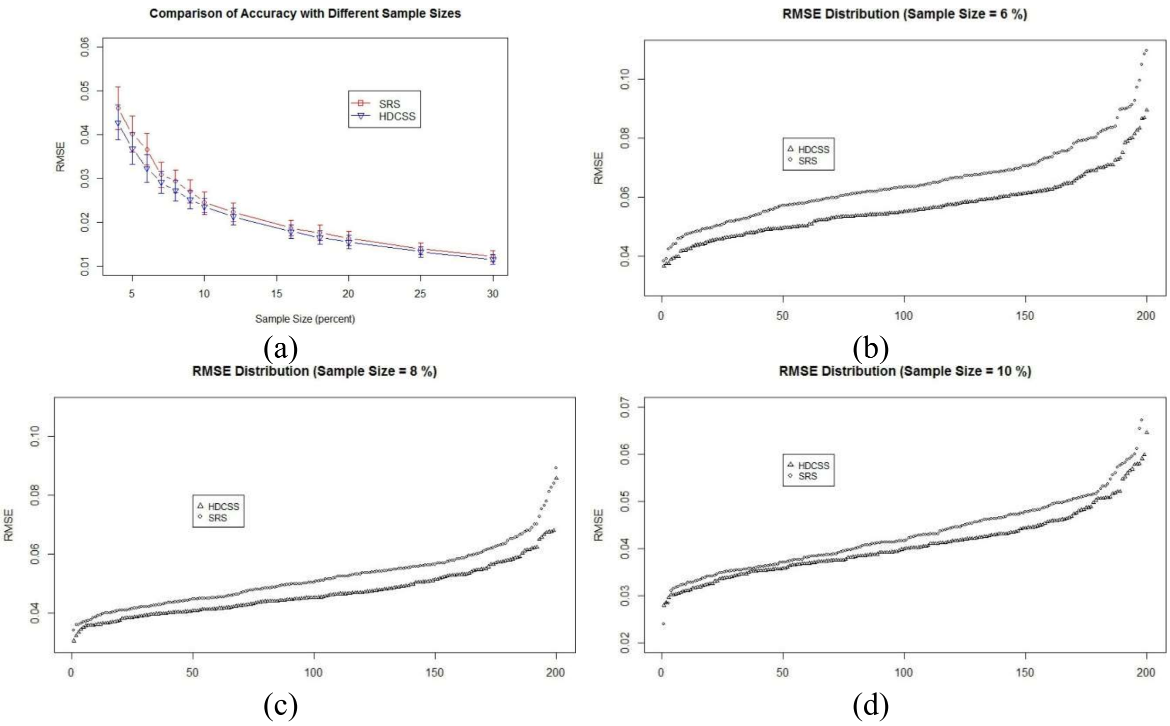


Figure 18 Sensitivity Analysis of Dimensionality (Types of Infrastructures) with Different Sample Sizes



(a)

(b)

(c)

(d)

Figure 19 Sensitivity Analysis of Sample Sizes Between SRS and HDCSS Methods

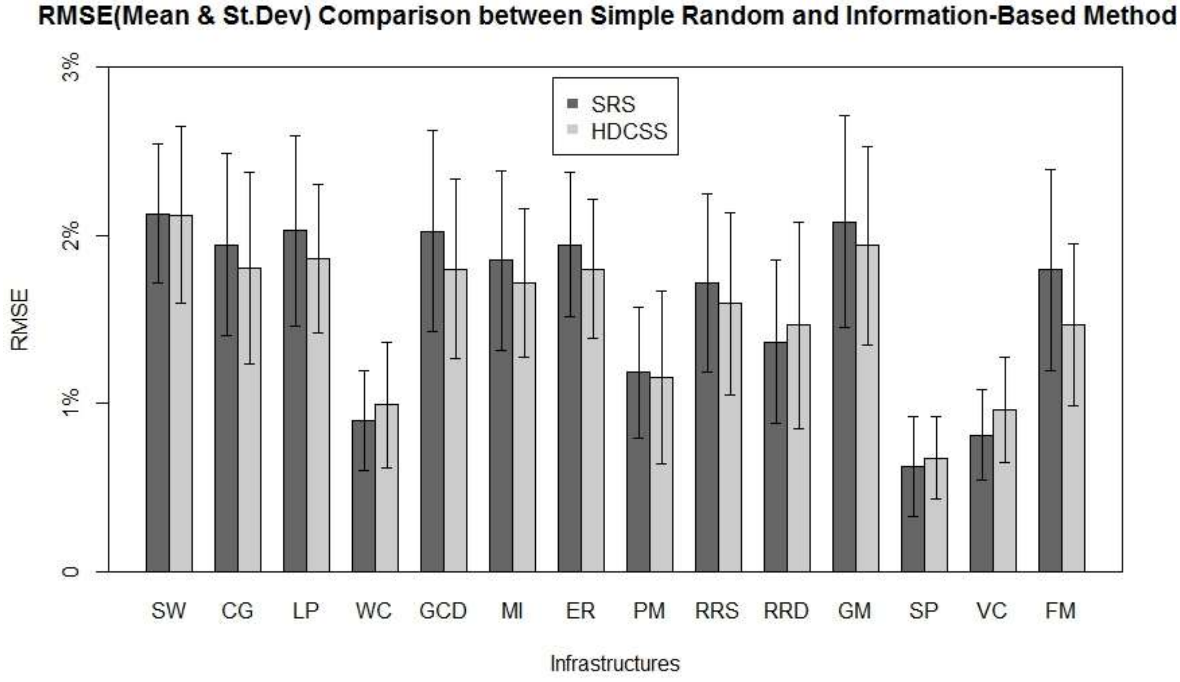
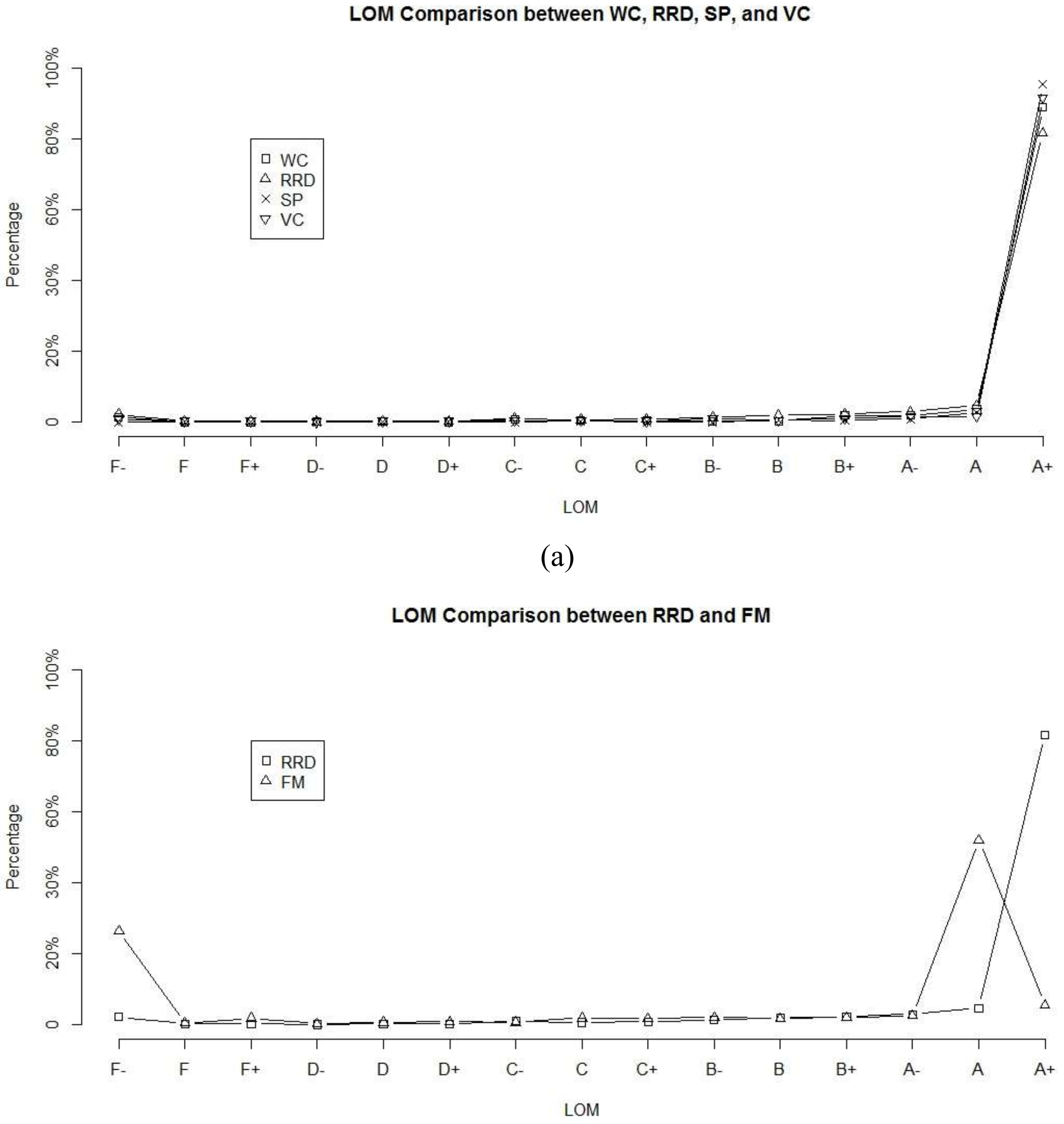


Figure 20 Comparison of RMSE (Mean and Standard Deviation) Between SRS Method and HDCSS Method



(a)

(b)

Figure 21 Grade Distribution Comparison Between Infrastructures: (a) WC, PM, RRD, SP, and VC; (b) FM and RRD

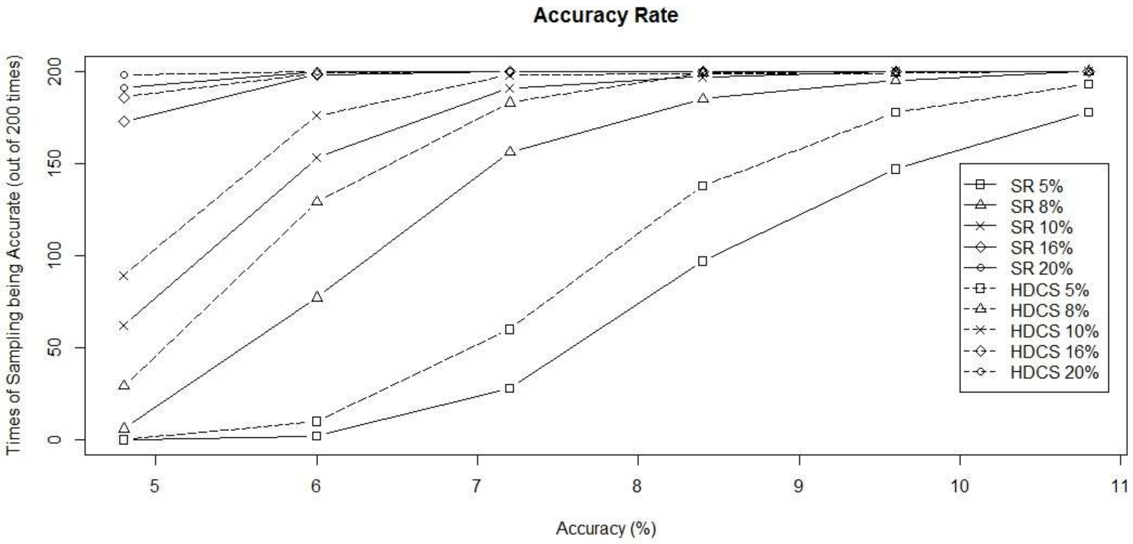


Figure 22 Sensitivity Analysis of Accuracy Rates Under Different Error Thresholds and with Different Sample Sizes

Table 6 Average Deterioration Rates of Infrastructures (per Month)

Infras	SW	CG	LP	WC	GCD	MI	ER
Det_rate	0.0756	0.0996	0.0821	0.0151	0.0394	0.0698	0.0813
Infras	PM	RRS	RRD	GM	SP	VC	FM
Det_rate	0.0181	0.0988	0.0244	0.0752	0.0022	0.0207	0.0825

Table 7 Results of ANOVA for Errors of Samples Selected by HDCSS (16%) and SRS (18%) Methods

	Degree of Freedom	Sum of Squares	Mean Square	FValue	PValue	Significantly Different, Yes/No
Between Features	1	0.0000033	3.314e06	1.28	0.259	No
Within Features	398	0.0010306	2.589e06			

Table 8 Results of ANOVA Tests

HDCSS Sample Rate (%)	SRS Sample Rate (%)	PValue
8	9	0.51
10	11	0.73
11	12	0.806
12	13	0.814
16	18	0.119
18	20	0.259
19	21	0.298

CHAPTER 5

CONCLUSION AND RECOMMENDATION

This chapter presents the summary of the research findings for each of the three studies addressed in this research, and describes the major research contributions and limitations for big data analytics in dynamic ridesharing, nonrecurrent congestion, and maintenance infrastructure sampling.

5.1 Dynamic Ridesharing

Dynamic ridesharing has been considered as a promising tool to mitigate traffic congestion in freeway network. But its effects on congestion relief are unknown when competing with other travel alternatives. This study on agent-based modeling of a dynamic ridesharing system investigates the impacts of dynamic ridesharing on multimodal network and the competing mechanism between a dynamic ridesharing system and public transit. The model considers traveler decision making process under the presence of the competing modes. It is important for the traffic planner to better analyze the market, improve the market penetration, and plan or deploy the dynamic ridesharing program. Travelers are classified into seven categories based on their travel mode preference. The number of each type of travelers are estimated with the travel mode parameters. These parameters include the percentage of group travelers and the market penetrations of dynamic ridesharing and public transit. By adjusting the parameters of travel mode preference, the model is applicable to any traffic network with diverse socioeconomic attributes. The modeling results are used in assessing the benefits and identifying the challenges of implementing dynamic ridesharing across different cities. Dynamic ridesharing service providers can also utilize such information to make corresponding marketing strategies.

The competing mechanism between dynamic ridesharing and public transit, as one objective in this study, has been summarized from the modeling results. When the public transit has low market penetration, namely, very few travelers utilize the public transit, there are very limited effects of public transit on the network. By adding the dynamic ridesharing system in the multimodal network, the occupancy on HOV lanes significantly increases, so the network-wise congestion keeps decreasing as the market penetration of dynamic ridesharing increases. However, when there are high public transit demands on the network, initially dynamic ridesharing system turns many public transit users to private vehicle users and encourages more ridership of private vehicles. Despite the fact that the congestion decreases eventually as the market penetration of dynamic ridesharing is high enough, it increases the congestion initially. The existence of public transit also influences the matching rate of ridesharing system. The dynamic ridesharing systems on the freeway network without public transit are usually quite sensitive to trip matching constraints, so any flexibility in the matching constraints would increase the number of matched trips. For the dynamic ridesharing system in a multimodal network competing with public transit, a loose matching constraint (longer waiting time) only brings a quite limited number of extra matched trips. Marketing strategies can be made based on such information by government or commercial dynamic ridesharing service providers.

The limitations of the dynamic ridesharing study lie in the assumptions designed to simplify the modeling process. One assumption in the model is that ridesharing travelers only search for other travelers who depart from the same origin node and to the same destination node. However, in reality, a shared trip can be granted as long as the passenger's destination is on the path which the driver drives along. This assumption excludes the paired travelers who depart for different modes, but also can share rides. To simplify the decision making process, the agent-based model does not distinguish the shared ride offerors and offerees, assuming that all the travelers waiting for a shared ride are able to provide a private vehicle. This may overestimate the matching rate for dynamic ridesharing by granting a shared ride to two travelers who traveled in the public transit mode only before. Neither of them would provide private vehicles. Another assumption is that all the travelers traveling on the network have full knowledge of the current traffic condition in the network, the queuing status at each bus stop, and number of travelers aboard. Theoretically, such information can be collected by an integrated ITS system, e.g., the current traffic condition can be obtained from the live traffic monitoring system, the queuing status at a bus stop can be estimated from the video captured by the surveillance camera in bus stations, and the number of travelers aboard can be counted by automated passenger counter (APC) devices. But it is unfeasible to require that this information be available throughout the entire city.

Future work in the study of dynamic ridesharing includes applying the proposed agentbased approach to actual networks with different demographics, and measuring the network performance with a spatial and temporal-varied index. In the case study, different types of traffic demands on the same network have been modeled by adjusting the market penetration and group traveler parameters. It is desired to conduct modeling with actual travel demand in a real traffic network. The market penetration and group traveler parameters can be obtained from survey data. The modeling results based on real data can help government and dynamic ridesharing service providers make proper decisions in terms of institutional design and marketing strategy.

Since the distribution of traffic congestion is very uneven spatially and temporally, the performance of the multimodal network should be measured with spatial and temporal-varied index.

5.2 Nonrecurrent Congestion

Nonrecurrent congestion, especially incident-induced delay, is a pronounced contributor for traffic unreliability. There was very limited work that has been accomplished on estimation the IID on the individual incident level. This study of quantifying nonrecurrent congestion has developed a methodology to quantify the delay induced by individual incidents, which is the major contribution of the study. The methodology combines spatiotemporal analysis and pattern recognition to carry out an information construction process, which dynamically uncovers a variety of features associated with any specific incident. The spatiotemporal analysis offers great insights on the convoluted causes of delay and unveils the true impacts of nonrecurrent congestion. The pattern recognition process identified the recurrent congestion in a hypothetical scenario where the incident never happened.

Another contribution of the study of nonrecurrent congestion is secondary incident identification. The methodology proposed in this study integrates multiple criteria to identify secondary incidents in a spatiotemporal extent, including instantaneous delay, contiguity, and shockwave front location. The frequency of secondary incidents reflects the vulnerability of incidents happening under severe congestion, which can serve as a performance measurement of freeway corridors. Hot spot analysis was conducted based on the output of methodology, the frequency of incidents/secondary incidents/primary incidents, and the IID statistics at each freeway spot. The results of hot spot analysis are used for incident mitigation and investment decision-making. For example, when an incident occurs at a bottleneck, the traffic operator can apply speed harmonization to accelerate bottleneck clearance and create a uniform speed.

The major limitation of the study of nonrecurrent congestion quantification is the limited number of historical records under incident-free scenario used in pattern recognition. In the proposed methodology, the recurrent congestion during the incident is estimated KNN method with the congestion under incident-free scenarios with the most similar previous traffic conditions. Due to the uniqueness of traffic conditions, the incident-free scenarios to be searched must be at the same location, TOD, and DOW with the incident. Considering that the traffic conditions may change significantly over the years, the methodology was implemented by data from 2013. The number of available incident-free scenarios for each incident is between 30 and 40. The incident with the least incident-free scenarios had only 25 incident-free scenarios to select the most similar ones. The limited number of records may lead to low accuracy in estimating the recurrent congestion.

Another limitation that impairs the reliability of the methodology is the lack of validation. This is also a major limitation for all the previous studies of nonrecurrent congestion quantification. The most challenging problem in the quantification of IID is to estimate the recurrent congestion under the hypothetical scenario if the incident never happened. Since the recurrent congestion is fluctuating over time, the exact value of recurrent congestion is unknown. Therefore, it is impossible to validate the output of the methodology implementation with ground truth.

Future work of nonrecurrent congestion quantification includes two topics: identifying the features associated with IID, and separating the effects of secondary incidents from the effects of primary incidents.

5.3 Maintenance Infrastructure Sampling

The objective in the study of maintenance infrastructure sampling is to develop asampling method to choose proper segments where the conditions of sampled infrastructures can represent the LOMs of the full inventory within the network. To accomplish this objective, a highdimensional clustering-based stratified sampling method is proposed. The HDCSS method is based on a stratified method, but utilizes high-dimensional cluster analysis to define the similarity between segments. It integrates infrastructure deterioration prediction, localitysensitive hashing, and spectral clustering. The information required by highdimensional clustering is constructed with infrastructure deterioration prediction, which assumes the current condition of a segment is predictable with the historical conditions of the same segment. LSH is used to quantify the similarity between segments with multiple features (types of infrastructures). After the similarities between segments are defined, the segments in the network are classified into several strata with spectral clustering.

The HDCSS method has been tested with infrastructure inspection records collected from the freeway network throughout the State of Utah. Generally HDCSS outperforms SRS by yielding lower errors, especially under the circumstance where LOM varies greatly within infrastructures. Another advantage of applying HDCSS is that it effectively reduces the sample size without compromise in accuracy compared with SRS, leading to significant saving in inspection costs for large scale network inspection. For example, when the sample size with SRS is less than 15%, using HDCSS can reduce the sample size by 1%. When the sample size with SRS is more than 15%, it reduces the sample size by 2%.

One limitation of the HDCSS method lies in the process of infrastructure deterioration prediction. To simplify the sampling process, it is assumed that the deterioration of the

infrastructures on a segment is a linear process, which obviously underestimated the variation of the deterioration. The deterioration prediction process provides essential information for the high-dimensional clustering. An accurate estimation of the infrastructures' current conditions is a prerequisite for accurate sampling results. The infrastructure deterioration prediction may be a major source of error in the sampling method.

Another limitation of the study lies in clustering analysis. In HDCSS, the similarity between segments is quantified by the LSH method, which is quite different from the Euclidean distance. It is difficult to define the physical meaning of the similarity between the segments with features of infrastructures. So the proper number of clusters is unknown. In the case study, the number of clusters was determined empirically when clusters have similar numbers of items. Sensitivity analysis can provide some insights into the relation between sampling accuracy and the number of clusters, but it is questionable if it is applicable to other datasets.

Future work of maintenance infrastructure sampling includes improving the deterioration prediction and applying other high-dimensional cluster analysis. Deterioration matrix construction can significantly influence the accuracy of the sampling method. It is necessary to apply a more rigorous data analysis tool to enhance the estimation of deterioration process. Testing other high-dimensional clustering methods can potentially improve the accuracy of the sampling results.

5.4 Summary

The major objective of this research was to explore the application of data-driven analytics in solving transportation problems which were not solvable with traditional methods. This dissertation developed new methodologies revolutionizing the solutions to problems in transportation planning, traffic operation, and infrastructure maintenance. In the study of simulating dynamic ridesharing competing with public transit in a multimodal network, market penetration parameters are incorporated into the model. By adjusting those parameters in the customized model, transportation agencies can make decisions regarding adopting the dynamic ridesharing system or revising policies to accommodate the service. Meanwhile, dynamic ridesharing service providers can adjust their marketing strategies to increase the exposure of the service and attract potential users. In the study of nonrecurrent congestion quantification, data-driven analytics is applied to estimate the delay induced by an individual incident. Quantifying IID at the individual level provides alternative methods for hot spot identification on a freeway corridor. Different from incident frequency, which is another commonly used measurement for hot spot identification, IID at individual level provides traffic operators a new perception to identify the ill-designed freeway segments or locations based on congestion. By applying IID for hot spot identification, vulnerability to congestion caused by nonrecurrent reasons has been considered as an index for measuring the reliability of freeway performance. The sampling method proposed for infrastructure inspection enables maintenance personnel to incorporate historical IM&R

records into the selection of inspection samples, which can better represent the LOMs of full infrastructure inventory. Compared with the current inspection sampling method, the proposed data-driven method requires small sample size and reduces inspection costs.

From the three examples in this dissertation, it is concluded that data-driven analytics has great potential in revolutionizing transportation problem solving. With data of good quality collected from heterogeneous sources and when novel data mining techniques becoming available, more existing transportation problems will be solved by integrating such data-driven analytics into practice. However, so far there are still several issues in applying datadriven analytics in transportation engineering. The first issue is associated with the validation of analysis results. For example, in the study of quantifying IID, the only data that might be available to validate the IID result is video recording, yet it is almost unrealistic to capture the IID for each and every incident from video recording, let alone for an entire freeway corridor. Another issue is the data quality. All applications of data-driven analytics proposed in this dissertation are established based on data of good quality. Without accurate data, the results can be significantly compromised. Therefore, future efforts on the application of data-driven analytics can be focused on result validation and data quality control.

APPENDIX

SPECTRAL CLUSTERING ALGORITHM

Spectral clustering algorithm *(93)*:

Given a set of points $V = \{v_1, v_2, \ldots, v_n)$, the similarity matrix $S = \{s_{ij}\}$, where s_{ij} refers to the similarity between v_i and v_j

1. Define D to be the diagonal matrix $D_{ii} = \Sigma_{\mathrm{j}} A_{ij}$, and construct the matrix $L = D^{-1/2} A D^{-1/2}$
2. Find $x_1, x_2, \ldots, x_k$, the k largest eigenvectors of L, and form the matrix $X = [x_1, x_2, \ldots x_k]$ by stacking the eigenvectors in columns.
3. Form the matrix Y from X by renormalizing each of X's rows to have unit length (i.e. $Y_{ij} = X_{ij} / \left(\Sigma_j X_{ij}^2\right)^{1/2}$).
4. Treating each row of Y as a point in $\Re^k$, cluster them into k clusters via Kmeans.
5. Finally, assign the original point v_i to cluster j if and only if row i of the matrix Y was assigned to cluster j.

REFERENCES

1. Eisele, W. L., T. Lomax, D. L. Schrank, and S. M. Turner. Lessons Learned for Transportation Agencies Preparing for MAP-21 Performance Management Requirements Related to Mobility and Reliability. Presented at 93rd Annual Meeting of Transportation Research Board, Washington, D.C., 2014.

2. Provost, F., and T. Fawcett. Data Science and Its Relationship to Big Data and Data-Driven Decision Making. *Data Science and Big Data*, Vol. 1, No. 1, 2013, pp. 51–59.

3. Tufte, K. A., B. Elazzabi, N. Hall, M. Harvey, K. Knobe, D. Maier, M. Veronika, and V. Megler. Guiding Data-Driven Transportation Decisions. *Computer Science Faculty Publications and Presentations*, Jan. 2014, p. 128.

4. Donhost, M., and V. Anfara. Data-Driven Decision Making. *Middle School Journal*, Vol. 42, No. 2, 2010, pp. 56–63.

5. Chan, N. D., and S. A. Shaheen. Ridesharing in North America: Past, Present, and Future. *Transport Reviews*, Vol. 32, No. 1, Jan. 2012, pp. 93–112.

6. Schrank, D., B. Eisele, and T. Lomax. *TTI 's 2012 Urban Mobility Report. INRIX Traffic Data*. 2012.

7. Bertini, R., and G. E. McGill. Getting Traffic Moving Again. *Public Roads*, Vol. 67, No. 2, 2003, pp. 14–17.

8. Cambridge Systematics. *Incorporating Reliability Performance Measures into the Transportation Planning and Programming Processes*. 2013.

9. Karlaftis, M. G., S. P. Latoski, N. J. Richards, and K. C. Sinha. ITS Impacts on Safety and Traffic Management: An Investigation of Secondary Crash Causes. *ITS Journal - Intelligent Transportation Systems Journal*, Vol. 5, No. 1, 1999, pp. 39–52.

10. Allen, D. K., S. Karanasios, and A. Norman. Information Sharing and Interoperability: The Case of Major Incident Management. *European Journal of Information Systems*, Vol. 23, No. 4, Jun. 2013, pp. 418–432.

11. Khattak, A., X. Wang, and H. Zhang. Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. *Transportation Research Record: Journal of the Transportation Research Board*,

Vol. 2178, No. 1, Dec. 2010, pp. 128–137.

12. Lou, Y., Y. Yin, and S. Lawphongpanich. Freeway Service Patrol Deployment Planning for Incident Management and Congestion Mitigation. *Transportation Research Part C*, Vol. 19, No. 2, 2011, pp. 283–295.

13. Hegyi, A., B. DeSchutter, J. Hellendoorn, B. De Schutter, and J. Hellendoorn. Optimal Coordination of Variable Speed Limits to Suppress Shock Waves. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6, No. 1, Mar. 2005, pp. 102–112.

14. Hellinga, B., and M. Mandelzys. Impact of Driver Compliance on the Safety and Operational Impacts of Freeway Variable Speed Limit Systems. *Journal of Transportation Engineering*, Vol. 137, No. 4, 2011, pp. 260–268.

15. Durango-Cohen, P. L. A Time Series Analysis Framework for Transportation Infrastructure Management. *Transportation Research Part B: Methodological*, Vol. 41, No. 5, Jun. 2007, pp. 493–505.

16. Mishalani, R. G., and L. Gong. Optimal Sampling of Infrastructure Condition: Motivation, Formulation, and Evaluation. *Journal of Infrastructure Systems*, Vol. 15, 2009, pp. 313–320.

17. Furuhata, M., M. Dessouky, F. Ordóñez, M. E. Brunet, X. Wang, and S. Koenig. Ridesharing: The State-of-the-Art and Future Directions. *Transportation Research Part B: Methodological*, Vol. 57, 2013, pp. 28–46.

18. Di Febbraro, A., E. Gattorna, and N. Sacco. Optimization of Dynamic Ridesharing Systems. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2359, No. 2359, 2013, pp. 44–50.

19. Agatz, N. A. H., A. L. Erera, M. W. P. Savelsbergh, and X. Wang. Dynamic Ride-Sharing: A Simulation Study in Metro Atlanta. *Transportation Research Part B: Methodological*, Vol. 45 `, No. 9, 2011, pp. 1450–1464.

20. Huang, S. C., M. K. Jiau, and C. H. Lin. A Genetic-Algorithm-Based Approach to Solve Carpool Service Problems in Cloud Computing. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 1, 2015, pp. 352–364.

21. Agatz, N., A. Erera, M. Savelsbergh, and X. Wang. Optimization for Dynamic Ride-Sharing: A Review. *European Journal of Operational Research*, Vol. 223, No. 2, 2012, pp. 295–303.

22. Aissat, K., and A. Oulamara. A Priori Approach of Real-Time Ridesharing Problem with Intermediate Meeting Locations. *JAISCR*, Vol. 4, No. 4, 2014, pp. 287–299.

23. Nourinejad, M., and M. J. Roorda. Agent Based Model for Dynamic Ridesharing.

Transportation Research Part C: Emerging Technologies, Vol. 64, 2016, pp. 117–132.

24. Hosni, H., J. Naoum-Sawaya, and H. Artail. The Shared-Taxi Problem: Formulation and Solution Methods. *Transportation Research Part B: Methodological*, Vol. 70, 2014, pp. 303–318.

25. Santos, D. O., and E. C. Xavier. Dynamic Taxi and Ridesharing: A Framework and Heuristics for the Optimization Problem. *IJCAI International Joint Conference on Artificial Intelligence*, 2013, pp. 2885–2891.

26. Santos, D. O., and E. C. Xavier. Taxi and Ride Sharing: A Dynamic Dial-a-ride Problem with Money as an Incentive. *Expert Systems with Applications*, Vol. 42, No. 19, 2015, pp. 6728–6737.

27. Fagnant, D. J., and K. M. Kockelman. Dynamic Ride-Sharing and Fleet Sizing for a System of Shared Autonomous Vehicles in Austin, Texas. *Transportation*, 2016, pp. 1–16.

28. Deakin, E., K. T. Frick, and K. M. Shively. Markets for Dynamic Ridesharing? *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2187, No. 1, 2011, pp. 131–137.

29. Galland, S., L. Knapen, A. U. H. Yasar, N. Gaud, D. Janssens, O. Lamotte, A. Koukam, and G. Wets. Multi-Agent Simulation of Individual Mobility Behavior in Carpooling. *Transportation Research Part C: Emerging Technologies*, Vol. 45, 2014, pp. 83–98.

30. Stiglic, M., N. Agatz, M. Savelsbergh, and M. Gradisar. Making Dynamic Ride-Sharing Work: The Impact of Driver and Rider Flexibility. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 91, 2016, pp. 190–207.

31. Shaheen, S. A., N. D. Chan, and T. Gaynor. Casual Carpooling in the San Francisco Bay Area: Understanding User Characteristics, Behaviors, and Motivations. *Transport Policy*, Vol. 51, 2016, pp. 1–9.

32. Mote, J. E., and Y. Whitestone. The Social Context of Informal Commuting: Slugs, Strangers and Structuration. *Transportation Research Part A: Policy and Practice*, Vol. 45, No. 4, 2011, pp. 258–268.

33. Liu, Y., and Y. Li. Pricing Scheme Design of Ridesharing Program in Morning Commute Problem. *Transportation Research Part C: Emerging Technologies*, Vol. 79, 2017, pp. 156–177.

34. Kramers, A. Designing Next Generation Multimodal Traveler Information Systems to Support Sustainability-Oriented Decisions. *Environmental Modelling and Software*, Vol. 56, 2014, pp. 83–93.

35. Chavis, C., and V. V Gayah. Development of a Mode Choice Model for General Purpose Flexible Route Transit Systems. Presented at 96th Annual Meeting of Transportation Research Board, Washington, D.C., 2016

36. Hussain, I., L. Knapen, A. Yasar, T. Bellemans, D. Janssens, and G. Wets. Negotiation and Coordination in Carpooling : An Agent-Based Simulation Model. Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.

37. Cho, S., A. U. H. Yasar, L. Knapen, T. Bellemans, D. Janssens, and G. Wets. A Conceptual Design of an Agent-based Interaction Model for the Carpooling Application. *Procedia Computer Science*, Vol. 10, 2012, pp. 801–807.

38. Knapen, L., D. Keren, A. U. H. Yasar, S. Cho, T. Bellemans, D. Janssens, and G. Wets. Analysis of the Co-Routing Problem in Agent-Based Carpooling Simulation. *Procedia Computer Science*, Vol. 10, No. 270833, 2012, pp. 821–826.

39. Sanchez, D., S. Martinez, and J. Domingo-Ferrer. Co-utile P2P Ridesharing via Decentralization and Reputation Management. *Transportation Research Part C: Emerging Technologies*, Vol. 73, 2016, pp. 147–166.

40. Bellemans, T., S. Bothe, S. Cho, F. Giannotti, D. Janssens, L. Knapen, C. Korner, M. May, M. Nanni, D. Pedreschi, H. Stange, R. Trasarti, A. U. H. Yasar, and G. Wets. An Agent-Based Model to Evaluate Carpooling at Large Manufacturing Plants. *Procedia Computer Science*, Vol. 10, No. 270833, 2012, pp. 1221–1227.

41. LeBlanc, L. J. Transit System Network Design. *Transportation Research Part B: Methodological*, Vol. 22, No. 5, Oct. 1988, pp. 383–390.

42. Liu, X., G. Zhang, Y. Lao, and Y. Wang. Quantifying the Attractiveness of High-Occupancy Toll Lanes with Traffic Sensor Data Under Various Traffic Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2229, Dec. 2011, pp. 102–109.

43. Gentile, G., S. Nguyen, and S. Pallottino. Route Choice on Transit Networks with Online Information at Stops. *Transportation Science*, Vol. 39, No. 3, 2005, pp. 289–297.

44. Yu, R., Y. Lao, X. Ma, and Y. Wang. Short-Term Traffic Flow Forecasting for Freeway Incident-Induced Delay Estimation. *Journal of Intelligent Transportation Systems*, Vol. 18, No. 3, May 2014, pp. 254–263.

45. Mongeot, H., and J. Lesort. Analytical Expressions of Incident-Induced Flow Dynamics Perturbations Using Macroscopic Theory and Extension of Lighthill-Whitham Theory. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1710, No. 0, Jan. 2000, pp. 58–68.

46. Wirasinghe, S. Determination of Traffic Delays from Shock-Wave Analysis.

Transportation Research, Vol. 12, No. 5, Oct. 1978, pp. 343–348.

47. Li, J., C. Lan, and X. Gu. Estimation of Incident Delay and Its Uncertainty on Freeway Networks. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1959, Jan. 2006, pp. 37–45.

48. Bailie, C., J. a. McFarland, J. a. Greenough, and D. Ranjan. Effect of Incident Shock Wave Strength on the Decay of Richtmyer–Meshkov Instability-Introduced Perturbations in the Refracted Shock Wave. *Shock Waves*, Vol. 22, No. 6, 2012, pp. 511–519.

49. Rakha, H., and W. Zhang. Consistency of Shock-Wave and Queuing Theory Procedures for Analysis of Roadway Bottlenecks. Presented at 84th Annual Meeting of Transportation Research Board, Washington, D.C., 2005.

50. Skabardonis, A., P. Varaiya, and K. F. Petty. Measuring Recurrent and Nonrecurrent Traffic Congestion. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1856, No. 1, 2003, pp. 118–124.

51. Moore, J. E., G. Giuliano, and S. Cho. Secondary Accident Rates on Los Angeles Freeways. *Journal of Transportation Engineering*, Vol. 130, No. 3, 2004, pp. 280–285.

52. Anbaroglu, B., B. Heydecker, and T. Cheng. Spatio-Temporal Clustering for Non-Recurrent Traffic Congestion Detection on Urban Road Networks. *Transportation Research Part C: Emerging Technologies*, Vol. 48, 2014, pp. 47–65.

53. Chung, Y. Quantification of Nonrecurrent Congestion Delay Caused by Freeway Accidents and Analysis of Causal Factors. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2229, No. 1, Dec. 2011, pp. 8–18.

54. Chung, Y., and W. Recker. A Methodological Approach for Estimating Temporal and Spatial Extent of Delays Caused by Freeway Accidents. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 3, 2012, pp. 1454–1461.

55. Snelder, M., T. Bakri, and B. van Arem. Delays Caused by Incidents. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2333, 2013, pp. 1–8.

56. Chung, Y. Identifying Primary and Secondary Crashes from Spatiotemporal Crash Impact Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2386, No. 1, Dec. 2013, pp. 62–71.

57. Yang, H., B. Bartin, and K. Ozbay. Use of Sensor Data to Identify Secondary Crashes on Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2396, No. 1, Dec. 2013, pp. 82–92.

58. Yang, H., K. Ozbay, E. F. Morgul, B. Bartin, and K. Xie. Development of an Online Scalable Approach for Identifying Secondary Crashes. *Transportation Research Record Journal of the Transportation Research Board*, No. 26, 2014, pp. 15–33.

59. Chung, Y. Assessment of Non-Recurrent Congestion Caused by Precipitation Using Archived Weather and Traffic Flow Data. *Transport Policy*, Vol. 19, No. 1, 2012, pp. 167–173.

60. Chung, Y., and W. W. Recker. Spatiotemporal Aanalysis of Traffic Congestion Caused by Rubbernecking at Freeway Accidents. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 3, 2013, pp. 1416–1422.

61. Habtemichael, F. G., and M. Cetin. Methodology for Quantifying Incident-Induced Delays on Freeways by Grouping Similar Traffic Patterns. Presented at 94th Annual Meeting of Transportation Research Board, Washington, D.C., 2015.

62. Park, H., and A. Haghani. Real-Time Prediction of Secondary Incident Occurrences Using Vehicle Probe Data. *Transportation Research Part C: Emerging Technologies*, 2015.

63. Yang, X., and W. Recker. Simulation Studies of Information Propagation in a Self-Organizing Distributed Traffic Information System. *Transportation Research Part C: Emerging Technologies*, Vol. 13, No. 5–6, Oct. 2005, pp. 370–390.

64. Keller, J. M., M. R. Gray, and J. A. Givens. A Fuzzy K-Nearest Neighbor Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-15, No. 4, Jul. 1985, pp. 580–585.

65. Khattak, A., X. Wang, and H. Zhang. Are Incident Durations and Secondary Incidents Interdependent? *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2099, No. 1, Jan. 2009, pp. 39–49.

66. Imprialou, M.-I. M., F. P. Orfanou, E. I. Vlahogianni, and M. G. Karlaftis. Methods for Defining Spatiotemporal Influence Areas and Secondary Incident Detection in Freeways. *Journal of Transportation Engineering*, Vol. 140, No. 1, Jan. 2014, pp. 70–80.

67. Chung, Y. Assessment of Non-Recurrent Traffic Congestion Caused by Freeway Work Zones and Its Statistical Analysis with Unobserved Heterogeneity. *Transport Policy*, Vol. 18, No. 4, Aug. 2011, pp. 587–594.

68. Ord, J. K., and A. Getis. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, Vol. 27, No. 4, 1995, pp. 286–306.

69. Zhang, H., and A. Khattak. What Is the Role of Multiple Secondary Incidents in Traffic Operations? *Journal of Transportation Engineering*, Vol. 136, No.

November, Nov. 2010, pp. 986–997.

70. Smilowitz, K., and S. Madanat. Optimal Inspection and Maintenance Policies for Infrastructure Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 1, Jan. 2000, pp. 5–13.

71. Mishalani, R. G., and L. Gong. Optimal Infrastructure Condition Sampling over Space and Time for Maintenance Decision-Making under Uncertainty. *Transportation Research Part B: Methodological*, Vol. 43, No. 3, 2009, pp. 311–324.

72. Mishalani, R. G., and L. Gong. Evaluating Impact of Pavement Condition Sampling Advances on Life-Cycle Management. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2068, 2009, pp. 3–9.

73. Guillaumot, V. M., P. L. Durango-Cohen, and S. M. Madanat. Adaptive Optimization of Infrastructure Maintenance and Inspection Decisions under Performance Model Uncertainty. *Journal of Infrastructure Systems*, Vol. 9, No. 4, 2002, pp. 133–139.

74. Ben-Akiva, M., F. Humplick, S. Madanat, and R. Ramaswamy. Infrastructure Management under Uncertainty: Latent Performance Approach. *Journal of Transportation Engineering*, Vol. 119, No. 1, 1993, pp. 43–58.

75. Durango-Cohen, P. L., and S. M. Madanat. Optimization of Inspection and Maintenance Decisions for Infrastructure Facilities under Performance Model Uncertainty: A Quasi-Bayes Approach. *Transportation Research Part A: Policy and Practice*, Vol. 42, No. 8, 2008, pp. 1074–1085.

76. Adams, T. M., and B. Winkelman. Statistical Analysis for Assessing Highway Maintenance Level of Service. *Transportation Research Record: Journal of Transportation Research Board.* No. 2551, 2016, pp. 73–81.

77. Medina, R. a., A. Haghani, and N. Harris. Sampling Protocol for Condition Assessment of Selected Assets. *Journal of Transportation Engineering*, Vol. 135, No. 4, 2009, pp. 183–196.

78. De la Garza, J. M., J. C. Piñero, and M. E. Ozbek. Sampling Procedure for Performance-Based Road Maintenance Evaluations. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2044, No. 2044, 2008, pp. 11–18.

79. Bellman, R. E. *Adaptive Control Processes: A Guided Tour*. 1961.

80. Steinbach, M., L. Ertöz, and V. Kumar. The Challenges of Clustering High Dimensional Data. In *New Directions in Statistical Physics*, Springer Berlin Heidelberg, pp. 273–309.

81. Hinneburg, A., and D. A. Keim. Optimal Grid-Clustering: Towards Breaking the Curse of Dimensionality in High-Dimensional Clustering. 1999, pp. 506–517.

82. Aggarwal, C. Re-Designing Distance Functions and Distance-Based Applications for High Dimensional Data. *ACM SIGMOD Record*, Vol. 30, No. 1, 2001, pp. 256–266.

83. Li, C., E. Chang, H. Garcia-Molina, and G. Wiederhold. Clustering for Approximate Similarity Search High-Dimensional Spaces. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 14, No. 4, 2002, pp. 792–808.

84. Datar, M., N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-Sensitive Hashing Scheme Based on P-Stable Distributions. *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, 2004, pp. 253–262.

85. Charikar, M. S. Similarity Estimation Techniques from Rounding Algorithms. *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing - STOC '02*, 2002, pp. 380–388.

86. Jain, P., B. Kulis, and K. Grauman. Fast Image Search for Learned Metrics. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, No. June, 2008, pp. 1–8.

87. Kulis, B., and K. Grauman. Kernelized Locality-Sensitive Hashing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 6, 2012, pp. 1092–1104.

88. Brint, A. T. T. Sampling on Successive Occasions to Re-Estimate Future Asset Management Expenditure. *European Journal of Operational Research*, Vol. 175, No. 2, 2006, pp. 1210–1223.

89. Prozzi, J. A., and F. Hong. Transportation Infrastructure Performance Modeling through Seemingly Unrelated Regression Systems. *Journal of Infrastructure Systems*, Vol. 14, No. June, 2008, pp. 129–137.

90. Kulis, B., and K. Grauman. Kernelized Locality-Sensitive Hashing for Scalable Image Search. *Proceedings of the IEEE International Conference on Computer Vision*, No. Iccv, 2009, pp. 2130–2137.

91. Von Luxburg, U. A Tutorial on Spectral Clustering. *Statistics and Computing*, Vol. 17, No. 4, 2007, pp. 395–416.

92. Schmitt, R. L. R., S. Owusu-ababio, R. M. R. Weed, E. E. V Nordheim, and R. L. R. Schmitt. *Development of a Guide to Statistics for Maintenance Quality Assurance Programs in Transportation*. 2006.

93. Ng, A. Y., M. I. Jordan, and Y. Weiss. On Spectral Clustering: Analysis and an

Algorithm. Advances in Neural Information Processing Systems 14, Vol. 14, 2001, pp. 849–856.