NORTHWESTERN UNIVERSITY

Travel Time Reliability in Stochastic Dynamic Transportation Networks: Modeling, Path Finding and Routing

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Civil and Environmental Engineering

By

Monika Filipovska

EVANSTON, ILLINOIS

December 2021

ABSTRACT

Travel Time Reliability in Stochastic Dynamic Transportation Networks:
Modeling, Path Finding and Routing

Travel time is a key aspect of capturing and evaluating the operational performance and service quality of transportation systems, and travel time improvement is a common objective for travelers, service providers, transportation practitioners and agencies. However, the reliability of travel times, including the probability of unexpected delays, is an important factor for travelers' satisfaction with their transportation network experience, shown to affect their travel choices. Travel time reliability can also play a significant role in the decision making of goods shippers and logistic firms, where variability can have a direct economic impact. Modeling travel time variability is key to quantifying the inherent uncertainty in the knowledge of future events in a transportation network. It provides a more comprehensive representation of the state of the network and allows for making decisions that account for uncertainty and are robust to potential travel time variability.

This dissertation is concerned with modeling, optimization, and analysis problems in stochastic dynamic transportation networks, where link travel times are modeled as random variables with time-varying distributions. Motivated by the need for data-driven and application-oriented modeling and optimization approaches for transportation network analysis that consider travel time reliability, the overarching goal of this thesis is to define, model and present solution approaches for key problems in stochastic dynamic transportation networks. The key objectives of this thesis include (1) modeling the temporal and spatial dependencies in stochastic transportation

networks revealed in observed travel time data, (2) devising solution approaches for the estimation of path travel time distributions considering those spatio-temporal dependencies, (3) defining and solving path finding problems for reliable least-time routing, (4) incorporating en-route information in routing problems and analyzing its impact on travelers' decision making.

To meet the first objective, this thesis presents a comprehensive methodology for modeling the temporal and spatial aspects of stochastic transportation networks, as well as their intersection: the temporal variation of spatial characteristics and vice versa. To address the second objective, this thesis presents, tests, and evaluates a number of distribution estimation approaches that consider the network's spatio-temporal characteristics. The third objective is at the center of three problem classes of concern in this dissertation: (1) a priori reliable least-time paths, (2) trajectory-adaptive reliable least-time strategies, and (3) information-adaptive reliable least-time routing. The fourth and final objective is concerned with a key characteristic of stochastic network models, namely that knowledge of future network states can be adjusted based on information of past and current states. This objective is met via the latter two of these problem classes that consider traveler decision making based on in-vehicle trajectory data availability and connected vehicle information access in a connected environment.

In addressing these core objectives, this dissertation achieves the larger goal of presenting a comprehensive conceptual and methodological framework for modeling, estimation, and optimization in stochastic dynamic networks. The modeling and estimation methods from the first two objectives are key tools in addressing the problems in the following two objectives. Furthermore, the solution approaches for the three routing problems contain a shared component that allows for their solution to be initiated in a single shared procedure.

The problem classes and solution approaches that this thesis is concerned with have several important application areas. The stochastic transportation networks characterization and estimation of path travel time distributions can be applied for performance measurement and monitoring of transportation policies, projects and applications concerned with the reliability performance of transportation systems. Reliable path finding problems have a host of relevant applications, such as reliability-based vehicle routing of freight or mobility service providers, or applications for emerging transportation technologies and services such as electric vehicles, autonomous vehicles, ride-sourcing companies, etc. Real time data access and the increased use of navigation services also call for making reliability-based decision-making adaptive to information.

Acknowledgments

I am grateful and indebted to a great number of people who have given me their support throughout my PhD.

I would like to start by thanking my advisor and mentor, Dr. Hani Mahmassani, for giving me this opportunity, and for his invaluable support, instruction, supervision, and encouragement. I feel extremely honored and grateful to have had the opportunity to listen and learn from his deep insights, holistic vision, and ideas. I am even more privileged to have worked with him on making some of those ideas come to life. I will always be grateful to have been the beneficiary of his guidance through my PhD, continued enthusiasm about my work, encouragement to pursue challenging problems, and immense trust and confidence he has shown in my research and career pursuits. I am privileged to have him as my role model, and I hope to have gained even an ounce of his enthusiasm, wisdom, and diligence.

I am very grateful to my dissertation committee of distinguished researchers, Dr. David Morton, Dr. Yu (Marco) Nie, Dr. Ali Zockaie, and Dr. Jiwon Kim. Their comments and suggestions have been extremely valuable in forming my dissertation. They each brought expertise that was essential to my dissertation work, I am thankful to have had the opportunity to learn from them.

I would also like to thank the other members of the transportation faculty at Northwestern, Dr. Joseph Schofer, Dr. Pablo Durango-Cohen, Dr. Amanda Stathopoulos, and Dr. Ying Chen. Their teaching and insights have broadened my knowledge in transportation tremendously.

To my friends and colleagues, Dr. Marija Ostojic, Dr. Archak Mittal, Dr. Michael Hyland, Dr. Amr Elfar, Dr. Lama Al Hajj Hassan, Dr. Helen Pinto, Eunhye Kim, Haleh Ale Ahmad, and

Moein Hosseini, at the Northwestern University Transportation Center (NUTC), I am thankful to have learned with and from them and to have had their camaraderie through graduate school. I am also thankful to Hoseb, Divyakant, Yuhan, Christopher, Bingy at NUTC, as well as my cohort at the Civil and Environment Engineering department's transportation program, Elisa, Maher, and Jason. I am very lucky to have shared this journey with them. I am thankful for the support of staff at NUTC and the Department of Civil and Environmental Engineering (CEE), including, Diana, Hillary, Cindy, Joan, Andrea, Goldie, Bret, and Melissa. I am also very grateful for the fellowships and funding support from the department, NUTC and my advisor, that made my doctoral studies possible in the first place.

I owe a debt of gratitude to my undergraduate research advisor, Dr. Saif Jabari, and Dean Samer Madanat at New York University Abu Dhabi (NYUAD). I am lucky to have worked with Prof. Saif Jabari, and to be introduced to the field of transportation and academic research, which enticed me to change my graduate school plans. I am grateful to have had both of their mentorship during my final year at NYUAD and their support and encouragement to pursue graduate studies.

I would like to thank my family for their love and support throughout my studies. I will never be able to fully express my gratitude to my parents for their immense sacrifices, for believing in me and trusting me to pursue education away from home since I was sixteen, and for their sustained interest in and support of my work. I am thankful to my brother for his friendship and support, as well as his ideas and suggestions whenever my work could benefit from his knowledge and experience. I am thankful to my parents-in-law and sister-in-law for their continued prayers and encouragement, for their support despite the distance and the language barriers. I am also very thankful for our dog, Avche, whose company comforted me more than I could have imagined, and who ensured I went for a walk even on the busiest days.

I am most thankful to my wonderful husband, Paul, for his love, friendship, encouragement, and support during all stages of graduate school. I am so fortunate to have had you by my side from start to finish. Your understanding and unwavering support through each obstacle and success have kept me going - thank you for always believing in me and lifting me up. I have benefited greatly from the sacrifices you have made and all of the things you have put on hold to allow me to prioritize my work. I would not have made it here without you. Above all, my deepest gratitude belongs to the One whose grace carried me here.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Motivation

The performance of transportation systems, especially their efficiency and quality of service, is important to every level of society. From the impact on the environment and the safety of the communities these systems serve, to the transportation costs of businesses and industries, to the time, resources, and quality of life of individuals, the impact of transportation problems and solutions is ubiquitous. Travel time is a key aspect of assessing the operational efficiency and service quality of transportation systems and improving travel time is often seen as the common objective of individual travelers, service providers, transportation practitioners and agencies. Yet from the perspective of individual users, there are key characteristics of travel time that are traditionally not accounted for in transportation systems evaluations. Firstly, travelers evaluate the transportation system through their experience on entire paths or trajectories for specific origin-destination pairs, rather than in distinct portions of the network or via overall aggregate quantities. Secondly, travelers can be concerned about travel time variability at least as much as they care about mean travel time and have varying levels of sensitivity to travel time reliability.

Transportation networks and their performance are affected by exogeneous factors such as changing weather conditions, work zones, and traffic control devices, which impact the decisions and actions of travelers (Kim and Mahmassani, 2015; Filipovska et al., 2019; Filipovska and Mahmassani, 2020a). Travelers' decisions in turn become part of the operation of the system and endogenously affect the state and performance of the network. The performance of transportation networks is uncertain due to fluctuations in both the exogeneous and endogenous factors that affected it. Such fluctuations cause uncertainty in the transportation network and as a result the travel time between any pair of points in the network can be viewed as a random variable. Moreover, systematic time-dependence and variation of some of the exogenous factors cause the distributions of travel times in the network to be non-stationary and time-dependent. The presence of travel patterns, the structure of the network and the dependence of traffic conditions across links additionally impose spatial and temporal dependencies between travel times in the network, whether on individual links or more generally between two points in the networks.

Travel time reliability has been recognized as an important factor for users' satisfaction with their experience in a transportation network and has been shown to affect the travel and activity choices of individual travelers. Furthermore, it affects the service levels experienced by goods shippers and logistics firms and has a direct economic impact on their decisions. As such, travel time reliability affects the route choices of individual travelers and fleet operators in the network. However, unlike in deterministic networks where travel times are cardinal numbers, users in stochastic networks may have different responses to travel time variability based on their personal preferences and risk tolerance, which may also be non-stationary and vary depending on factors such as their departure time or trip purpose. The presence of different types of users with varying levels of risk tolerance and different reliability-based objectives in the network may also result in different network dynamics.

Reliability-based path finding and routing in stochastic dynamic networks is a multi-faceted problem. Firstly, it requires the modeling of travel time distributions across various paths on the network so as to capture their time-varying nature and intrinsic spatio-temporal dependencies. Secondly, reliability-based objectives call for path finding solutions that extend beyond stationary and deterministic path search and are complicated by the characteristics of stochastic dynamic networks. Furthermore, since solutions to path finding problems are inherently uncertain, accounting for en-route information about revealed travel times in the network may result in a change to the optimal path selection for a given user, which introduces the need for adaptive routing solutions. With the increased reliance on traveler information services and navigation systems, and the increasing availability of trajectory data via geographical positioning systems (GPS), users expect real-time updated information and increasingly use such information for their decision-making.

1.2 Research Overview and Objectives

This dissertation is concerned with problems in stochastic dynamic networks with spatio-temporal dependencies. The overarching objective of this dissertation is to define, model and present solution approaches for key problems in stochastic dynamic transportation networks through a comprehensive conceptual and modeling framework.

Modeling approaches for the estimation of path travel time distributions are developed to capture the progression of traffic over space and time and the non-stationary correlation between travel times on different links. Considering different ways to model stochastic dynamic networks and the spatio-temporal dependencies of travel times in such networks, this study will review existing methods and introduce new approaches for modeling travel time variability in the network, specifically at the path level. The objective for this portion of the dissertation is to provide conclusions and guidelines for unifying the travel time variability modeling approaches with those for stochastic dynamic network modeling. Applications of such a modeling framework may include, but are not limited to, performance measurement, performance monitoring and simulation modeling in the context of transportation policies, projects and applications concerned with the reliability performance of transportation systems.

This dissertation further aims to extend the modeling framework to include path finding approaches in stochastic dynamic networks, by applying the travel time variability and estimation methods, based on different ways of modeling stochastic dynamic networks, and with different en-route information access scenarios. Firstly, assuming no en-route information access, the problem of a priori path finding is addressed with exact and approximate approaches for path generation and under different reliability-based least-time objectives. Secondly, in the case of adaptive routing, problem definitions and solution approaches differ based on different assumptions regarding levels of spatio-temporal correlations in the network, en-route information availability and traveler responses to information. This study proposes exact and heuristic approaches for adaptive routing problems to improve the efficiency of adaptive routing solutions and make them suitable for real-time application. Problem definitions with different information types and different responses to information are considered under a unified methodological framework. Furthermore, since objectives for path finding in stochastic dynamic networks can vary between different users or applications, the routing approaches are suitable for heterogeneous users with different objective types and risk sensitivity levels.

1.3 Stochastic Dynamic Network Problems

This section introduces and defines the problems addressed in this dissertation, discusses the features of the problems and how they will be addressed. The problem definitions are followed by the associated expected contributions in section 1.4.

Consider a transportation network consisting of a set of nodes and directed arcs. The travel times along the arcs in the network are modeled as positive continuous random variables with distributions that vary over time. Furthermore, the link travel times are assumed to be correlated over space and time. The characterization of such generalized stochastic dynamic networks is the first step in defining and formulating the problems in this dissertation and is an extension to the stochastic, time-varying network definitions introduced by Miller-Hooks (1999) and extended by Gao (2004). Miller-Hooks (1999) defines stochastic time-varying networks via a set of discrete time-bins each of equal duration of the smallest increment of time over which a perceptible change in the travel time distributions will occur. Then, for each such time-bin and each arc in the network, a set of non-negative real valued possible travel times with their associated probabilities are given. Gao (2004) extends this framework and models the dependence between the arc travel times via joint time-dependent discrete distributions i.e. the possible joint realizations of travel times on all of the arcs for each time-bin.

1.3.1 Characterization of Stochastic Dynamic Networks

One of the goals of this dissertation is to introduce methods for data-driven characterization of the network where travel time data can be utilized to model the link travel time distributions. This framework will provide approaches to determine the points in time where the link travel time distributions change, or equivalently, to determine segments of time for which the marginal distributions can be considered constant. Furthermore, spatio-temporal dependencies between the arc travel times are to be modeled via a generalized covariance structure that may also change over time. The strength and presence of correlations may change for different spatial and temporal neighborhoods that are to be identified from the data.

Once the stochastic dynamic network with spatio-temporal dependencies is characterized in this manner, the problems addressed in this study pertain to the estimation of path travel time distributions, a priori and adaptive path finding problems, with different information access and response types, and are outlined in sections 1.3.2 to 1.3.5. These problems are further developed as part of the conceptual framework in Chapter 3 and formally addressed in later chapters of this dissertation.

1.3.2 Estimation of Path Travel Time Distributions

Given a stochastic dynamic network where the network link travel times are random variables with time-varying distributions correlated over space and time, the problem is to establish an approach or set of approaches for the estimation of path travel time distributions as the sum of such time-varying correlated random variables. The distribution function of the sum of random variables is determined by solving a convoluting integral, and in this problem the convoluting integral needs to be reformulated to allow for time-dependence and correlations between the random variables. Convolution integrals that account for the time-dependence aspect have been formulated in previous work, but they can only be solved analytically in restricted cases and with certain types of distribution forms. Partial dependence between the link travel times has been incorporated into the convolution formulation when modeled via different states each link could experience and their likelihood of occurrence.

In modeling the path travel time distributions for the case with generalized correlations between the time-varying link travel time distributions, the formulation of the convolution integral itself is challenging and often intractable. Since the general case is not limited to certain distribution types for the link travel times, the convolution integrals do not have analytic solutions.

Therefore, this portion of the study will aim to identify or develop and test different approaches for the estimation and computation of path travel time distributions. These approaches are intended to be applicable to general cases, while also leveraging the specific characteristics of the stochastic dynamic network.

1.3.3 The A Priori Path Finding Problem

Given a stochastic dynamic network with time-varying and correlated link travel time random variables, the a priori path finding problem focuses on determining the optimal paths, for a given set of departure times, from an origin node to all destination nodes in the network for a given reliability-based optimality criterion. The a priori path finding problem requires the estimation of path travel time distributions, which would be methodologically addressed via the problem in Section 1.3.2 in order to perform comparisons of path distributions as part of the search. Additionally, it can be expected that the modeling and characterization of the stochastic dynamic network and its correlation structure will significantly impact the possible solutions for path finding problems

The problem of a priori path finding is made more difficult in the context of stochastic dynamic networks with spatio-temporal dependencies since the characteristics of the network can easily invalidate some commonly used assumptions that exploit the network structure and allow for efficient solution algorithms, such as the first-in-first-out (FIFO) assumption and Bellman's principle of optimality. Without Bellman's principle of optimality, typically used path finding approaches for time-dependent networks become almost equivalent to the comparison and evaluation of all possible paths in the network. Such a solution approach can be very expensive computationally and inefficient, and therefore not suitable for most potential applications.

Therefore, one of the goals of this dissertation in relation to the path finding problem is to develop efficient heuristic path finding approaches that leverage the structure and the characterization of the stochastic dynamic network.

1.3.4 The Trajectory-Adaptive Routing Strategy Problem

In the trajectory-adaptive routing problem, given the stochastic dynamic network with correlated link travel times, the objective is to find an optimal routing strategy (or routing policy) that is designed to allow for changes in the path (or next node selection) at each intermediate node based on en-route revealed information about the network. This adaptive routing problem is not simply a path finding problem since its solution can be seen as a collection of paths, rather than a single path. The trajectory-adaptive problem assumes the simplest and most common type of information access, where a traveler has access solely to information from their own trajectory.

In describing the problem of adaptive strategy finding, it should be noted that in order for this problem to be solved fully it should be addressed from the perspective of a strategic or proactive traveler. A strategic traveler considers the availability of information in all later decision stages (i.e., the information that could be realized) so as to plan and choose their route based on the different likelihoods of choosing each of the paths comprising the routing policy. This problem carries the challenges of the a priori path finding problem, but the number of possible solutions to be evaluated increases significantly with the consideration of en-route decision points. Since the adaptive routing strategy problem directly relates to the use of en-route real-time information it should be expected to be solved in real time, which further amplifies the need for efficient solution approaches that simplify the problem without a significant loss in accuracy.

1.3.5 The Information-Adaptive Routing Problem

The information-adaptive problem considers a more general definition of information access, where a traveler in a connected environment can also have access to the trajectory information of any connected vehicle in the network. In this setting, a strategic traveler would not be realistic because consideration of the possible information they could receive in the future, in order to devise a routing strategy, would require knowing the destinations of the connected vehicles (CVs) in the environment, their objectives, access and response to information. However, the problem of information-adaptive routing for a traveler reactive to information allows for modeling the typical response to large levels of information and understanding the impact that information can have on a traveler's decisions, and specifically on meeting their objectives.

1.4 Contributions

This dissertation makes several methodological and conceptual contributions to the scientific literature.

First, this thesis considers the problem of stochastic transportation network modeling. It presents a taxonomy of existing stochastic dynamic network models and extends the taxonomy to inform future research in this domain. It develops a data-driven and application-oriented approach for the characterization of stochastic dynamic networks with spatio-temporal dependencies to reveal the network characteristics from data, instead of imposing them as a priori assumptions. The approach includes methods for modeling the spatio-temporal dependencies between the link travel times via a generalized covariance structure, testing for the strength, presence and change in those correlations for different spatial and temporal neighborhoods based on patterns revealed in the data.

Second, this thesis addresses the problem of estimation of path travel time distributions, identifying and testing approaches that can be used for this problem in the context of stochastic dynamic networks with generalized correlations. Solving this problem contributes to the literature on estimating path travel time distributions by comparing and contrasting different modeling approaches and providing conclusions for the methods suitable for a variety of application purposes. This contribution extends the unified framework for modeling stochastic dynamic networks and provides approaches that can be used for data driven solutions in path finding problems, not restricted by distribution types or state-based correlation modeling.

Third, this thesis deals with the problem of a priori path finding and contributes to the existing work in two related ways. To maintain the ability for efficient path finding by exploiting the network structure, modifications of Bellman's optimality principle or stochastic dominance are presented to allow for approximate solution approaches with adjustable risk-tolerance levels. Then, using those modified principles and approximation approaches it presents a heuristic approach with improved efficiency for path finding in stochastic dynamic networks with correlations.

Fourth, this dissertation presents a solution approach for trajectory-adaptive strategic routing that unifies this problem with the problem of priori path finding. Finally, it considers the problem of information-adaptive reactive routing via a general definition of information access and presents a solution approach that can be used for different special case problem under the general definition. Under this problem, it considers the stochastic dynamic network as a connected environment and solves the information-adaptive routing problem for different connectivity levels by varying the penetration of connected vehicles.

1.5 Organization

Chapter 1 introduces the dissertation by providing the motivation for the topic, presenting the research overview and objectives, and describing its contributions. **Chapter 2** reviews the literature on the topics relevant to this dissertation. **Chapter 3** presents the conceptual framework for the thesis and addresses each of its components and how they connect to one another within the overall conceptual framework. **Chapter 4** focuses on the stochastic dynamic network characterization and presents an approach for data-driven modeling of its spatio-temporal dependencies. **Chapter 5** is concerned with the problem of modeling and estimation of path travel time distribution in stochastic dynamic networks with correlations.

Chapter 6, Chapter 7 and Chapter 8 each present a different optimization problem in stochastic dynamic networks using the modeling approaches from Chapter 4 and Chapter 5. **Chapter 6** is focused on the a priori problem of reliable least-time paths (RLTP). **Chapter 7** considers the problem of trajectory-adaptive reliable least-time strategies (TA-RLTS). **Chapter 8** focuses on the information-adaptive reliable least-time routing (IA-RLTR) problem. **Chapter 9** concludes the dissertation with a summary of contributions, applications, and future research areas.

Chapter 2 Literature Review

This chapter presents a review of the literature relevant to this dissertation. As the topics of travel time reliability and the modeling and analysis of stochastic networks are quite broad, the literature review presented in this section is not meant to be exhaustive.

The literature review is separated into four sections that categorize the literature according to the problems introduced in section 1.3. First, relevant literature is presented regarding the approaches for stochastic dynamic network modeling in section 2.1. Then, literature on travel time variability modeling is presented in section 2.2 as it relates to the problem of estimation of path travel time distribution. The literature on path finding is separated into two closely related parts according to the problem definitions for the a priori path finding and adaptive routing. Section 2.3 provides an overview of the a priori path finding problems in the literature and section 2.4 reviews the body of literature on adaptive routing problems. Studies on problems with different types of information access and traveler response to information are all included in section 2.4 as this portion of the body of literature is relatively small.

2.1 Stochastic Dynamic Network Modeling

The modeling or characterization of stochastic transportation networks is considered in the literature primarily as part of problems for optimization or analysis in the contest of the stochastic network, rather than seeking to understand and model the network as a problem of its own.

The relevant studies, most of which will be considered again in other sections of this literature review, are summarized here, and categorized in terms of their approach to modeling the stochastic transportation network. Specifically, the approaches to modeling time variation or time dependence in the network and spatial dependence between link travel time distributions are used as the primary dimensions for comparison. The summary of the categorized literature is presented in Table 2-1.

Table 2-1. Summary of temporal and spatial dependence characterization of stochastic transportation networks

Temporal Dependence Characterization	Spatial Dependence Characterization	References
Time-invariant network assumption	Independent link travel times assumed	Nie and Wu (2009a), Chen et al. (2013), Chen et al. (2016)
	Partial link travel time dependencies with a Markovian link state model	Xing and Zhou (2011), Xing and Zhou (2013)
	Full link travel time dependencies with a stationary joint link travel time distribution	Zockaie et al. (2013), Zockaie et al. (2014)
	Full link travel time dependencies with a stationary general correlation structure	Prakash and Srinivasan (2014), Srinivasan et al. (2014), Prakash and Srinivasan (2015), Zeng et al. (2015), Chen et al. (2018)
	Partial link travel time dependencies for neighboring links only	Fan et al. (2005)
Link travel times vary during a peak period with fixed time intervals of 1 time unit	Independent link travel times assumed	Miller-Hooks and Mahmassani (1998a), Miller-Hooks and Mahmassani (2000a), Miller-Hooks and Mahmassani (2003a)
Link travel times vary during a peak period with fixed time intervals	Independent link travel times assumed	Wu and Nie (2009), Zhang et al. (2010), Nielsen et al. (2014), Pretolani (2000), Miller-Hooks (2001a), Miller-Hooks and Mahmassani (2003a), Opsanon and Miller-Hooks (2006), Prakash and Srinivasan (2017), Prakash et al. (2018)
	Partial link travel time dependencies with a Markovian link state model	Nie and Wu (2009b)

	Partial link travel time dependencies for neighboring links in an 'impact area'	Chen et al. (2012)
	Dependence, but additive mean variance assumed	Ji et al. (2011)
	Link travel time dependencies with a scenario-based representation	Yang and Zhou (2014)
	Full link travel time dependencies with a stationary joint link travel time distribution	Gao and Chabini (2006), Pretolani et al. (2009), Huang and Gao (2012), Zockaie et al. (2016), Huang and Gao (2018)
	Full link travel time dependencies with a stationary joint link travel time distribution: correlated log-normal	Chen et al. (2020)
	Full link travel time dependencies with a stationary general correlation structure	Yang and Zhou (2017), Zhang et al. (2017)

The literature summary in Table 2-1 reveals a few major categories by simply grouping the studies based on their assumptions in modeling the stochastic networks. Temporally, networks are primarily classified as time-invariant (or static) stochastic networks and time-varying stochastic networks with fixed time intervals. Spatially, different dependence assumptions for the link travel time random variables are encountered, primarily models assuming independent link travel time distributions, those assuming partial dependence, and models with full link travel time dependence assumptions. These categories are further expounded and classified in the taxonomy section of the corresponding Chapter 4.

2.2 Travel Time Variability Modeling

The literature addresses the problem of modeling and estimation of travel time distributions at a few different aggregation levels, such as the network, origin-destination (O-D), path and link level. Additionally, travel time variability has been defined from different perspectives, such as day-to-day, within-day (or time-of-day) and vehicle-to-vehicle variability.

At the network level, Mahmassani et al. (2012) demonstrate that the mean and standard deviation of network travel times (per unit distance) are highly positively correlated. Hunter et al. (2013) present a method for estimating travel time distributions in a network based on probe vehicle data, under the assumption that link travel times follow a multivariate Gaussian distribution. This study further presents a path travel time model that learns the network travel time distributions and makes inference on the travel time distributions for arbitrary paths in the network. Westgate and coauthors (2013) proposes statistical methods to estimate travel time distributions on a road network using Markov chain Monte Carlo approaches using GPS-enabled ambulance data. However, being a special type of network travel, ambulance trip data may misrepresent the features of urban traffic networks. Kim and Mahmassani (2014, 2015) propose a compound distribution representation of travel time distributions at the network level to capture both vehicle-to-vehicle and day-to-day variability, which is then integrated with mixture modeling techniques to model unobserved heterogeneity due to daily roadway conditions. At a general level, a few studies examine travel time variability analytically and statistically. Pu (2011) explores the mathematical relationships and interdependencies of a number of reliability measures under the assumption of lognormally distributed travel times. The study finds that the coefficient of variation is a good proxy for several other reliability measures. Fosgerau and Fukuda (2012) apply nonparametric statistical techniques to travel time data and show that though the mean and standard deviation of travel time change with the time of day, standardized travel time is ‘roughly independent' of the time of day. Even further, the authors conclude that, when using standardized travel times, the independence assumption 'could be reasonable'.

The question of estimating path travel time distributions can be answered in several ways, depending on the available information. A very large sample of trajectory data may allow for distributions to be constructed from experienced travel times for all users who have traveled the path of interest. However, to construct travel time distributions along any user-specified path, including those with very few observed vehicle traversals, one needs to identify vehicle traversals of segments or links along the path so as to synthesize the path's travel time distribution. Researchers have focused on the use of different types of data and various methods for estimating travel time distributions. Rahmani et. al (2013) point out some potential pitfalls when estimating path travel time distributions, specifically focusing on trajectories with incomplete traversal of the route, non-uniform coverage of the route in terms of the number of observations and using a non-representative vehicle sample. Ramezani and Geroliminis (2012) use probe vehicles' travel times along all links in an arterial route and estimate the route travel time distributions with a Markov chain approach where spatial correlations between successive links on a path are captured using transition probabilities. Such an approach is limited by the assumption that travel-time dependencies exist only between consecutive links and that such transitions between different links are conditionally independent. Another study synthesizes route travel time distributions based on segment-level temporal and spatial distributions by adding percentile-by-percentile values of the travel times (Isukapati et al., 2013). However, this technique may not be appropriate for generalized correlation structures. In terms of estimating the moments of a travel time distribution, Eisele and coauthors devise a method for estimating the mean and variance of route travel times

(Eisele et al., 2015), while a study by Chen and Osorio (2014) presents an analytic approach for approximating the standard deviation of travel times. Chen et al. (2017) introduced a copula-based model for estimating path travel time distributions on urban arterials which was shown to be superior to convolution and distribution fitting methods. However, the authors point out that the inputs to the proposed model are segment travel times, the marginal distributions for which need to be specified or estimated separately. A few other relevant studies use Markov Chain and Gaussian mixture models and specifically focus on incorporating signalized intersections (Ma et al., 2017; Yildirimoglu and Geroliminis, 2013).

2.3 A Priori Path Finding in Stochastic Networks

Reliability-based stochastic routing has been studied primarily with a focus on a priori path problems, and with early research assuming stationary stochastic networks, later extended to the time-varying case. The pivotal work on this topic (Frank, 1969) presents a closed form solution for the travel time probability distribution on shortest paths in stochastic stationary networks. Subsequently, procedures for discrete solutions and using utility functions to represent decision makers' preferences were presented (Eiger et al., 1985; Mirchandani, 1976; Sigal et al., 1980). In stochastic time-varying (STV) networks, a series of studies by Miller-Hooks and Mahmassani (2000b) focus on finding least expected time (LET) paths, least possible travel time paths (1998b), and propose label correcting algorithms using definitions of optimality based on first-order stochastic dominance (FSD) and definite stochastic dominance (2003b, 1998b). As indicated in the previous section, routing in stochastic networks has important implications for decision making under uncertainty and optimality conditions may be defined based on application-specific objectives or the preferences and risk tolerance of different user groups. Beside the LET criterion, studies have used reliability-based rules such as the shortest path problem with on-time arrival probability (SPOTAR) introduced by Nie and Wu (2009b, 2009a) or minimum travel time budget paths (MTTBP) (Zockaie et al., 2013, 2015, 2016; Fakhrmoosavi et al., 2018, 2019). Besides the fundamental difference in assumptions of static or time-varying travel time distributions, stochastic networks are modeled differently based on assumptions of dependency between link travel times.

The literature on a priori path finding problems in stochastic networks is summarized in Table 2-2, which separates the studies based on whether they consider time-invariant or time-varying stochastic networks. For each reference, the table indicates whether correlations between travel times are considered and if so, indicates how those correlations are modeled. Additionally, the type or approach for the solution algorithm are given, the solution type and the objective function or criterion used for path finding.

Table 2-2. Studies on a priori path finding problems in stochastic network

Reference	Correlations	Solution Algorithm	Solution Type	Objective
Stochastic Time-Invariant Network Assumption				
Nie and Wu (2009a)	No	Label-correcting algorithm	All O to 1 D	Shortest w Max on-time arrival prob.
Xing and Zhou (2011)	Yes, partial – Markovian link state model	Lower-bound approximate method	O-D pair	Most reliable path
Chen et al. (2013)	No	Multi-criteria label-setting algorithm and A* algorithm (heuristic)	All O to 1 D	Reliable shortest path, α-conf.
Xing and Zhou (2013)	Yes, partial – Markovian link state model	Heuristic using Lagrangian relaxation (previous paper)	O-D pair	Absolute and percentile robust shortest path

Zockaie et al. (2013)	Yes, via joint link travel time distribution	2-stage Monte-Carlo Simulation approach	All O to 1 D	Shortest w given on-time arrival prob.
Prakash and Srinivasan (2014)	Yes, general correlation structure	Pruning algorithm based on label-correcting procedures	O-D pair	Minimum Robust Cost Path
Srinivasan et al. (2014)	Yes, general correlation structure	Bounds-based sufficient optimality criterion algorithm Exact	O-D pair	Maximum reliability path
Zockaie et al. (2014)	Yes, via joint link travel time distribution	Outer approximation method	All O to 1 D	Min path travel time budget
Prakash and Srinivasan (2015)	Yes, general correlation structure	Bilevel minimization problem Exact algorithm	O-D pair	Maximum reliability path
Zeng et al. (2015)	Yes, general correlation structure	Lagrangian relaxation-based approach	O-D pair	α -reliable path problem
Chen et al. (2016)	No	Deviation path algorithm and A* approximation	O-D pair	K α-reliable paths
Chen et al. (2018)	Yes, general correlation structure	A moment-matching-based hybrid genetic algorithm (MHGA) and heuristic constraints	O-D pair	α -reliable path problem
Stochastic Time-Varying Network Assumption				
Miller-Hooks and Mahmassani (1998a)	No	Modified label-correcting algorithm	All O to 1 D, set of departure times	Least possible time
Miller-Hooks and Mahmassani (2000a)	No	Label-correcting algorithm, Exact	All O to 1 D, set of departure times	Least-expected time
Miller-Hooks and Mahmassani (2003a)	No	N/A (measures for comparison discussed)	N/A	N/A
Wu and Nie (2009)	No	Label-correcting algorithm	All O to 1 D, single departure time	Shortest w Max on-time arrival prob.
Nie and Wu (2009b)	Yes, partial – Markovian	Label-correcting algorithm and approximation	All O to 1 D, set of departure times	Shortest w given on-time arrival prob.

	link state model			
Zhang et al. (2010)	No	Chaotic Immune Particle Swarm Optimization	O-D pair	Multi-objective (mean and variance)
Ji et al. (2011)	Yes, but additive mean variance assumed	Simulation-based multi-objective genetic algorithm (Approximate)	O-D pair	Min travel time budget with given reliability level
Chen et al. (2012)	Yes, for neighboring links in an 'impact area'	Multi-criteria A* algorithm	All O to 1 D	K shortest α-reliable paths
Huang and Gao (2012)	Yes, joint distribution representation	Exact label-correcting algorithm	All O to 1 D, set of dep. times	Min Expected Disutility
Yang and Zhou (2014)	Yes, via scenario-based representation	Lagrangian relaxation-based lower bound approximation	O-D pair	Least expected travel time
Nielsen et al. (2014)	No	Best-first branch and bound method	O-D pair	Min expected cost
Zockaie et al. (2016)	Yes, joint distribution representation	2-stage Monte-Carlo Simulation approach	All O to 1 D	Min path travel time budget
Yang and Zhou (2017)	Yes, general correlation structure	Mathematical formulation (to be solved with LP algorithms)	O-D pair	On-time arrival probability and percentile travel time
Zhang et al. (2017)	Yes, general correlation structure	Lagrangian relaxation (LR) based algorithm	O-D pair	Reliable shortest path (RSP) problem
Chen et al. (2020)	Yes, as correlated log-normal distributions	Dynamic moment-matching method and approximation	O-D pair	Reliable Shortest Path

2.4 Adaptive Routing in Stochastic Networks

Adaptive routing problem formulations carry the time-dependence and correlation modeling questions, and also differ based on assumptions of information access. At the root of adaptive path finding is the idea that en-route revealed information about travel times in the network changes the knowledge of future travel time distributions in the network and potentially leading to a change in the optimal path for the user. Some of the comprehensive studies on the optimal routing strategy problems in STV networks (Gao and Chabini, 2006; Gao and Huang, 2012) present a taxonomy of the definitions of such problems. The different assumptions on link travel time dependency and information access in the literature are addressed in turn here, primarily with a focus on adaptive routing in STV networks, referencing the a-priori problems or the time-invariant counterparts where appropriate.

Studies of routing in STV networks differ in their assumptions and modeling of link travel time dependencies. Gao and Chabini define complete dependency as the case when travel times on all links and across all time-intervals are correlated, while on the other end of the spectrum is the assumption of no spatial or temporal dependence (Gao and Chabini, 2006). Limited dependency assumptions could also be made in order to capture some dependence and attempt to mitigate some of the potential problems and complications of the general dependency case.

The assumption of no dependency between link travel times is prevalent in the literature both for the case of stochastic static (i.e., time-invariant) and STV networks. A series of studies by Miller-Hooks and Mahmassani are based on the assumption of no dependency for the problem of a priori path finding (1998b, 2003b) and adaptive path finding (2001b; 2003a; 2006). For the a priori path finding problem, studies address the shortest reliable path problem (Chen et al., 2013, 2016; Nie and Wu, 2009b) and minimum travel time budget path finding problem (Zockaie et al., 2014). A few studies in the STV context assume limited dependency, and their representation of such dependencies varies across the literature. For example, in the a priori path finding context Nie and Wu (2009b) extend their work to include limited spatio-temporal dependencies between link travel times. The assumption of complete generalized dependency is the most general case, where the limited dependency and no dependency assumptions can be seen as special cases. Thus, solution methods for path finding or routing problems under the complete dependency assumption are directly applicable for solving the other types of problems, while the reverse does not hold. STV networks with complete link travel time dependencies in the literature are modeled in a few different ways. In the context of a priori path finding Zockaie et al. (2013, 2016) use a complete joint travel time distribution representation for the entire network. The studies addressing adaptive path-finding problems with different assumptions of travel time dependencies are further discussed below along with their assumptions on information access.

In terms of access to information, Gao and Chabini (2006) categorize adaptive routing problems in three groups. Perfect online information assumes the knowledge of link travel times across the entire network and for all past time periods; no online information is the case where travelers have no knowledge of previously realized travel times anywhere in the network, while partial online information assumes the knowledge of travel times on a portion of the network links and past time periods.

The modeling implications of assumptions on information access go hand in hand with the assumptions on travel time dependency. In the context of networks that assume no dependencies between link travel time distributions, adaptive routing is based on time alone (also referred to as time-adaptive routing) and the problem definition and solution method remain unchanged under the various information access assumptions. A number of studies address such problems in the literature, such as (Miller-Hooks, 2001; Miller-Hooks and Mahmassani, 2003a; Prakash and Srinivasan, 2017; Pretolani, 2000) to name a few. Some studies have also addressed the problems with partial online information and partial network dependency. A study by Fan and coauthors (Fan et al., 2005) addresses the adaptive path finding problem in static stochastic networks with limited correlations, where links can experience one of two states and conditional probabilities are introduced to address the correlation between the states on adjacent links.

Complete network dependency is assumed in a few studies with different assumptions of information access. In the context of static stochastic networks, studies by Srinivasan and coauthors (Prakash and Srinivasan, 2018, 2015; Srinivasan et al., 2014) present a few approaches for modeling correlations in link travel time distributions. Additionally, studies by Gao and coauthors (Gao and Chabini, 2006; Gao and Huang, 2012) assume complete temporal and spatial stochastic dependence of link travel times, where a joint distribution of discrete link travel times is applied. Various assumptions on information access are made with information scenarios such as delayed global information, real-time local information, and pre-trip global information.

Trajectory-adaptive routing is a special case of partial information that has been considered in some studies (Huang and Gao, 2018; Opasanon and Miller-Hooks, 2006; Pretolani et al., 2009) based only on the knowledge given user's trajectory. Specifically, trajectory information assumes the knowledge of travel times for the links traversed by the user and for the time periods during which they were traversed on their current route. Such an assumption is appropriate under decentralized routing systems which may have access to historical information, but only utilize the current trip information for en-route decisions. Specifically, a multicriteria adaptive path finding study (Opasanon and Miller-Hooks, 2006) presents a label-correcting algorithm, but under the assumption of no travel time distribution dependencies. Subsequently, a study by Pretolani and coauthors (Pretolani et al., 2009) provides an in-depth comparison of the time-adaptive and trajectory-adaptive problem in multi-criterion optimal routing. More recently, a study by Huang and Gao (Huang and Gao, 2018) presents an approach for trajectory-adaptive routing with complete dependency between link travel times. In this study, the complete stochastic dependency between link travel times is represented via a joint discrete travel time distribution across the links in the network and a peak travel period, and the authors present an exact solution algorithm for such a problem. A limitation of this study is the assumption of discrete travel times and the need to maintain a finite number of support points in the event space (i.e., the number of possible realizations of travel times in the network). Additionally, the authors demonstrate that the running time of the algorithm grows exponentially with the network size.

Table 2-3 summarizes the literature on adaptive path finding problems in stochastic networks. In addition to all of the information contained in the previous table, Table 2-3 also points out the problem type, which indicates the type of information-adaptive problem considered for each reference.

Table 2-3. Studies on adaptive path finding problems in stochastic network

Reference	**Problem Type**	**Correlations**	**Solution Algorithm**	**Solution Type**	**Objective**
Stochastic Time-Invariant Network Assumption					
Fan et al. (2005)	Trajectory-Adaptive	Yes, partial – states for neighboring links	Picard's Method of Successive Approximation	All O to 1 D	Least expected travel time
Stochastic Time-Varying Network Assumption					

Pretolani (2000)	Time-adaptive	No	N/A Formulation only	O-D pair (formulation)	Minimum expected and Min-max cost
Miller-Hooks (2001a)	Time-adaptive	No	Modified label-correcting algorithm	All O to 1 D, set of departure times	Least expected time
Miller-Hooks and Mahmassani (2003a)	Time-adaptive	No	N/A (measures for comparison discussed)	N/A	N/A
Opasanon and Miller-Hooks (2006)	Time-adaptive	No	2 stage approach: generation and selection of hyperpath	All O to 1 D, set of departure times	Multicriteria – Least disutility
Gao and Chabini (2006)	Full information, Full history-adaptive	Yes, via joint distribution representation	Decreasing Order of Time Algorithm, Exact	All O to 1 D, set of departure times	Least expected travel time
Pretolani et al. (2009)	Time-adaptive and history-adaptive	Yes, via joint distribution representation	N/A (Formulation / representation only)	O-D pair	N/A (Formulation/ representation only)
Prakash and Srinivasan (2017)	Time-adaptive	No	Iterative bounds-based algorithm	O-D pair	Minimum robust-cost strategy
Prakash et al. (2018)	Time-adaptive	No	Label-Correcting Network-Pruning Procedure	O-D pair	Minimum Robust-Cost strategy
Huang and Gao (2018)	Trajectory-Adaptive	Yes, via joint distribution representation	Decreasing order of time algorithm	All O to 1 D, set of departure times	Minimum Expected Disutility

2.5 Conclusion

A review of the literature relevant to this dissertation is presented in this chapter. The literature review sections 2.1 to 2.3 cover the literature related to the problems introduced in sections 1.3.1 to 1.3.3, respectively. The review of literature on adaptive routing problems in section 2.4 jointly covers the literature relevant to the problems introduced in sections 1.3.4 and 1.3.5. This chapter identifies gaps in the existing literature that are to be addressed in this dissertation. The specific problems, research questions and gaps that this dissertation is concerned with are first presented via the conceptual framework in Chapter 3 and then addressed individually in Chapter 4 through Chapter 8.

Chapter 3 Conceptual Framework

One goal of this dissertation is to develop a comprehensive framework that incorporates the approaches for modeling and optimization in stochastic dynamic networks. This chapter presents the conceptual framework, describing each of the components in turn, in sections 3.1 to 3.5 as they correspond to the problems identified in sections 1.3.1 to 1.3.5 in the Introduction, and later presented individually in Chapter 4 through Chapter 8. The overall conceptual framework is illustrated in Figure 3-1.

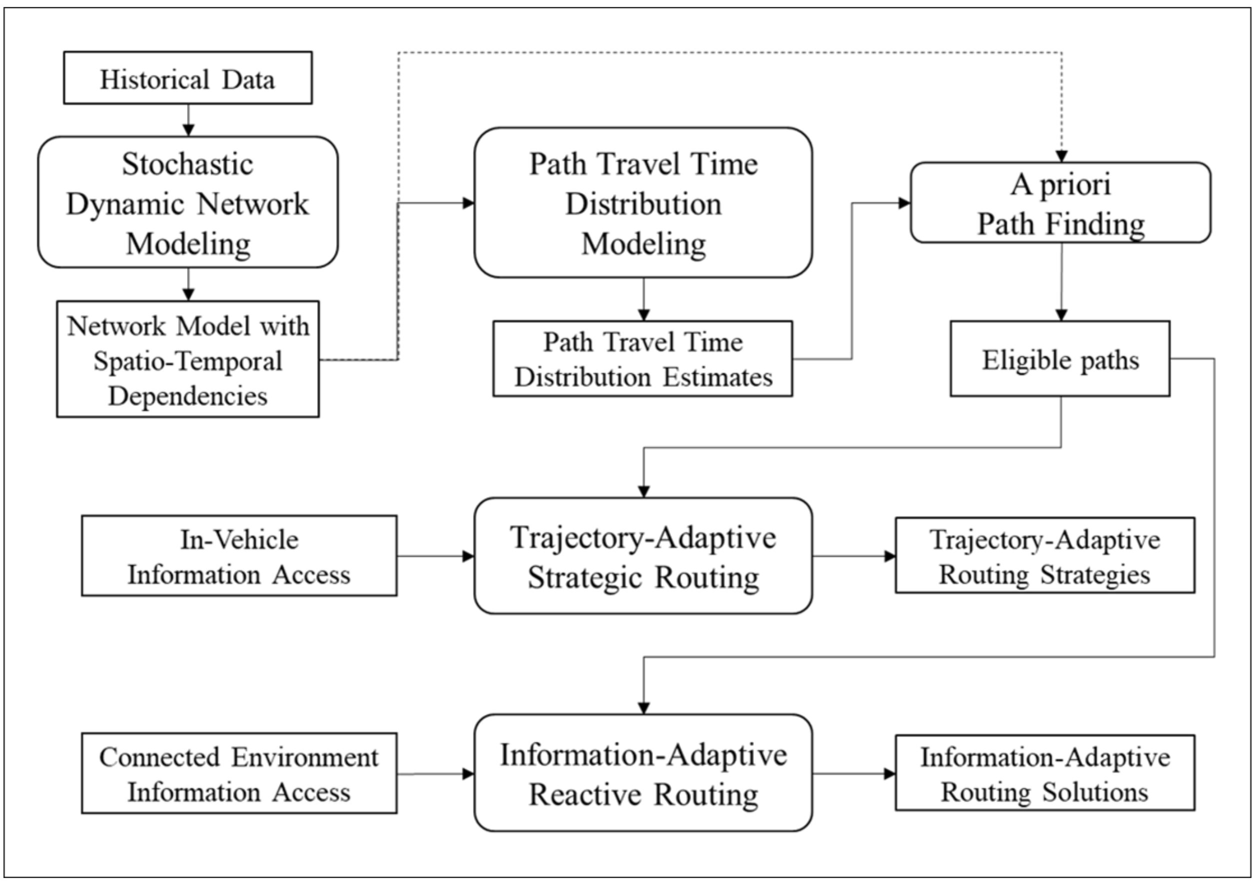

Figure 3-1. Dissertation Framework

The illustrated dissertation framework in Figure 3-1 shows the connections between the various components, either directly from one component to the next, or via an intermediate component.

3.1 Stochastic Dynamic Network Characterization

An important aspect of modeling and optimization in stochastic networks is the ability to identify the characteristics of the network accurately and effectively. The goal of this component of the thesis is to present an approach for the characterization of stochastic dynamic networks that would correctly capture the dynamics and interdependencies in the network and be useful for analysis and optimization.

In characterizing the stochastic dynamic network, a few important questions are asked, that will have a significant impact on the modeling of path travel time distributions and the optimization problems to follow.

1. Is the network time-invariant or time-varying?

   Time-varying network models, as mentioned in Section 1.1 and in the literature review in Chapter 2, are necessary for representing the dynamics of the network and capturing the changes in the network's performance over time. If using a time-varying network model, the next question is to determine the appropriate duration of time-intervals with which travel time distributions vary.

2. Are the travel times on network links correlated?

   The structure of the network itself and the presence of travel patterns lead to dependence of the traffic conditions on different links in the network, and consequently lead to correlations between the travel time distributions. If link travel times can in fact be modeled as correlated random variables, then a range of questions need to be answered on the type of correlations that exist between link travel times and their nature within the context of stochastic dynamic networks, which are further explored in Chapter 4.

3. Are travel times distributions impacted by endogenous and exogeneous factors in the network?

   The performance of transportation networks has been shown to be affected by a range of factors, ranging from different demand patterns due to special events, work zones, incidents or accidents, or weather events that change the driving behavior of travelers at a large scale. This question would aim to evaluate if and how the time-variability and interdependence of network travel times may change due to the impact of such factors.

The characterization of the stochastic dynamic network is the basis for the remaining components of this thesis, from the modeling of path travel time distributions, the assumptions that can be made for a priori path finding or the type of information relevant for adaptive routing.

3.2 Path Travel Time Distribution Modeling and Estimation

Establishing an approach (or set of approaches) for modeling path travel time probability distributions is essential to devising a cohesive framework for optimization in stochastic dynamic networks, and specifically for path finding and adaptive routing problems. As mentioned in Chapter 1, user-centric measures in the transportation network focus on full trips, paths, or trajectories that a given user may experience. In the context of stochastic dynamic networks, that translates into modeling the variability of travel time along entire paths for a given origin-destination pair, which is captured via the path travel time distribution.

Figure 3-1 shows that on the input side of path travel time probability modeling are the modeling of the stochastic dynamic network, and through that the availability of historical data. These two components are closely related to one another and also affect the type of models that would be appropriate or necessary for modeling path travel time distributions, as they become inputs for each following component, as shown in Figure 3-1.

In addition to these questions on the input side concerning the modeling of the stochastic dynamic network and the available data, there is important work that can be done in terms of the modeling of path travel time distributions. As mentioned in the literature review, determining the distribution of path travel times in stochastic dynamic networks is equivalent to the problem of determining the distribution of the sum of random variables. The sum of random variables is typically determined via a convolution integral. However, in this case, those travel time variables can be time-varying and correlated over space and time, which complicates the convolution integral and, in some cases, makes it impossible for the integral to be formulated in the first place. In cases when the integral can be formulated, numerical approximation approaches could be used to estimate the analytically intractable integrals. Monte-Carlo simulation approaches, which are typically used for the estimation of intractable integrals, can be implemented, and modified, for the specific problem at hand. Other sampling approaches suitable for sampling correlated random variables, such as Gibbs or the Metropolis-Hastings sampling, may be more appropriate for this application. Parametric methods can also be considered and compared to sampling-based approaches.

An important question that arises from all of the considerations in this section is – what level of detail is actually necessary in order to model path travel time distributions with sufficient accuracy? Which details can be forgone in order to make the approach simpler, more intuitive, computationally efficient, and applicable? Devising an approach for modeling path travel time distributions that is suitable given the stochastic dynamic network and the available input data will require a complete experimental design that accounts for these factors. The numerical experiments will need to consider the accuracy and computational effort for the possible approaches in order to identify one or more of these methods as suitable for certain definitions of the stochastic dynamic network.

3.3 A Priori Path Finding

With the path travel time modeling approaches devised based on section 3.2, the distribution estimates are to be used for path finding. As introduced in the literature review Section 2.3, the problem of path finding in stochastic dynamic networks is significantly different from the corresponding problems in deterministic networks. Given the stochastic dynamic network, decisions are made under uncertainty and as such they can only be optimal at an aggregate or overall level. The optimality of a specific solution is itself uncertain, for example a least expected travel time optimal path is expected to have the shortest travel time, but there is some likelihood that once travel times are revealed there may have been other better paths.

The literature review in Section 2.2 describes the existing work on a priori and adaptive path finding in stochastic networks and summarizes all of the references in Table 2-2. There is a lot of variation in the literature in terms of the assumptions of time-dependence and the modeling of correlations for such problems. Therefore, the modeling of the stochastic dynamic network, per section 3.1 and specifically the dependence assumptions regarding link travel time distributions will significantly impact the possible solutions for path finding problems. One goal in this thesis is to unify the formulation of the path finding problem with the problem of modeling travel time probabilities and estimation of path travel time distributions via their shared formulation of the stochastic dynamic network.

Assumptions necessary for the accurate modeling of path travel time distributions will need to remain and extend into the path finding problem, and as such will have an impact on the path finding solution approaches. Some important aspects of the assumptions typically made in the path finding problem need to be revised and re-assessed with respect to the context of stochastic dynamic networks.

One complicating factor is the first-in-first-out (FIFO) assumption that in deterministic networks specifies that vehicles entering a given link at a later time must also exit that link later. This assumption is critical for establishing the acyclicity criterion for optimal paths, i.e., that an optimal path will not traverse the same node more than once, thus preventing cycling in the network. The FIFO assumption has been modified for different networks, for example in networks that account for turning movements, vehicles entering a link at a later time may be allowed to exit it earlier if they are making a different turning movement. In the context of stochastic dynamic networks, the FIFO criterion has been adjusted and substituted with a stochastic FIFO criterion that is defined differently in different studies.

Another difficulty is introduced by the non-applicability of Bellman's principle of optimality, typically critical in the design of path finding solution algorithms that make use of the network structure. The principle says that any sub-path of the optimal path for a given origin-destination pair is itself optimal for the intermediate nodes. The optimality principle needs to be adjusted for stochastic dynamic networks in a few different ways. For networks with independent link travel times, Bellman's principle has been successfully used when applied to the dominance, rather than optimality, of paths and sub-paths using the first order-stochastic dominance (FSD) criterion. However, for stochastic dynamic networks with correlated link travel times, Bellman's principle has also been shown to be invalid with the FSD criterion. Thus, modifications of this principle are further investigated for a dominance criterion that can be used with Bellman's principle (i.e., at intermediate nodes) in stochastic dynamic networks with correlations. Even further, probabilistic versions of the criterion may be possible, for example ones where the likelihood of a sub-path's dominance can be quantified. Such modifications form the basis of algorithms for generating eligible paths for optimal a priori path finding in Chapter 6.

3.4 Trajectory-Adaptive Strategic Routing

The trajectory-adaptive routing problem in the context of a strategic traveler is a type of information-adaptive routing problem, previously discussed in the literature review in Section 2.4 and is summarized in Table 2-3. The problem difficulties for stochastic path finding problems, described in the previous section, are also relevant here. However, information-adaptive routing problems are defined by two key characteristics: the type of information access and the type of traveler response to information.

Trajectory-adaptive strategic routing considers access to in-vehicle information i.e., traveler's own trajectory and a strategic (i.e., proactive) response to that information. The problem specific are defined in more detail in Chapter 7. The important aspects of this problem with regard to the cohesive conceptual framework are that it is defined on the basis of the stochastic dynamic network model per section 3.1, uses the path travel time distribution estimation approaches according to section 3.2 and the eligible path generation criteria and algorithms from section 3.3.

3.5 Information-Adaptive Reactive Routing

The information-adaptive routing problem in the contest of a reactive traveler considers the other side of adaptive routing, there the traveler has access to external information via other vehicles' trajectories in a connected environment and the traveler has a reactive response to that information. With rapidly changing types and access to online information in transportation networks, the case of a vehicle operating in a connected environment is now conceivable and may allow for more fluid definitions of partial online information access for adaptive routing problems.

One aim of this component of the thesis is to unify the formulation of the adaptive routing problem with the problems of modeling travel time probabilities and estimating path travel time distributions via their shared formulation of the stochastic dynamic network. Furthermore, the goal is to integrate those aspects of the problem with different types of information access under one broad definition. Previous studies have considered the extreme cases of full information access, from the perspective of a centralized operator, or trajectory access, where an in-vehicle information system may only have access to that vehicle's trajectory only. However, modeling information access in a connected environment can allow for a fluid problem definition that can be used with varying levels of information availability. Under this definition, the full information case, the trajectory-information, or spatial neighborhood information can all be seen as special cases of a general information access definition.

Thus, this component of the dissertation also unifies the a priori and adaptive path finding problems via a solution approach that can be used to solve both problems. Additionally, this component aims to provide efficient solution approaches, as previous studies in the literature have recurrently recognized the need for approximate solution approaches, heuristics or hybrid approaches for adaptive path finding.

3.6 Conclusion

This chapter provides the conceptual framework for the proposed dissertation, centered around the goal of developing a comprehensive framework to incorporate the approaches for modeling, path finding and routing in stochastic dynamic networks. The input, methodology and output side for each of the components of the conceptual framework are addressed in sections 3.1 through 3.5. The next five chapters of the thesis, Chapter 4 through Chapter 8, present the problems, methods, and results on each of the conceptual framework components presented here.

Chapter 4 Characterization of Stochastic Dynamic Networks

4.1 Overview

The modeling of path travel time reliability, path finding and routing in stochastic dynamic transportation networks are rooted in the idea that a transportation network can be represented as a stochastic dynamic system. Consequently, it is important to characterize and specify the transportation network as a stochastic dynamic system accurately and effectively. An accurate specification of the network that correctly captures the network characteristics, including its dynamics and interdependencies, is crucial to having a useful model of the real network. Furthermore, an effective characterization that can be used for analysis and optimization is important to having a functional model of the network. There are often trade-offs between these objectives: accurate modeling of the network may involve modeling the complexity of its dynamics and interdependencies to a level of specificity that could be at odds with the usefulness of the resulting model, where simplicity is important for its applicability.

This chapter presents an approach for characterizing the spatial and temporal features of stochastic dynamic transportation networks. Starting with a problem definition in section 4.2, this chapter then presents a systematic taxonomy for stochastic dynamic network characterization in section 4.3 along with a review of literature from several relevant domains. The methodology in section 4.4 presents the approaches selected for the spatial and temporal characterization of stochastic dynamic networks. Section 4.5 presents a sensitivity analysis performed to assess the network characterization results and their robustness across several main dimensions related to parameters in the characterization methodology and changes in the network data due to varying demand and weather conditions.

4.2 Problem Statement

This section presents the problem of characterization of stochastic dynamic networks addressed in this chapter. Section 4.2.1 provides an outline of the research questions answered through this chapter, while section 4.2.2 provides a more formal problem definition for those questions.

4.2.1 Research Questions

In defining the stochastic dynamic network, several questions on how the network is to be modeled will have a significant impact on the modeling of path travel time distributions and the optimization problems to be solved in this system. These research question expand on the general questions introduced in the conceptual framework in section 3.1.

1. Is the network considered to be time-invariant or time-varying?

As mentioned in Section 1.1 and in the literature review Section 2, time-varying networks are necessary for representing the dynamics of the network and capturing the changes in the network's performance through time. Stochastic dynamic networks are typically modeled with time intervals, where travel time distributions can be considered static within each time interval but vary from one time interval to the next. Thus, an important aspect of modeling a stochastic time-varying network is the question of determining the appropriate time interval duration.

2. Are there dependencies between travel times on different links?

The structure of the network itself and the presence of travel patterns lead to dependence of the traffic conditions on different links in the network, and consequently lead to correlations between the travel time distributions. If link travel times can in fact be modeled as correlated random variables, then a range of questions need to be answered on the type of correlations that exist between link travel times and their nature within the context of stochastic dynamic networks.

- Are there dependencies between all or some of the links' travel times in the networks?
- Are there groups of links within which dependencies exist and outside which dependencies can be considered negligible?
- Are link travel times correlated over time? Given that traffic takes time to progress over space, there may be some time lag in the dependencies leading to link travel times being correlated across time.
- Are link travel time correlations stationary or do they vary over time? If temporal variation is observed, the duration of the corresponding time intervals also needs to be estimated.

3. Are the temporal and spatial dependencies stationary over varying network conditions?

The performance of transportation networks has been shown to be affected by a range of exogeneous factors, ranging from new demand patterns due to special events, work zones impacting the flow of traffic, incidents or accidents, weather events that change the driving behavior at a large scale, etc. The question is whether the temporal and spatial dependencies of travel time distributions in the network vary with any of these exogeneous factors. Specifically, the following questions can be asked:

- Are time-intervals for time-variable networks stationary over space and time? Over space, one needs to examine if link travel time distributions vary at the same rate across all links (or groups of links) in the network. Across time, the question is whether time time-interval duration for temporal dependence changes based on the time of day.
- Are link travel time dependencies stationary over time? That is, one needs to examine whether link travel time correlations change over time and to determine the corresponding time-interval.
- More broadly, which of the aspects of the stochastic dynamic network, modeled as part of answering questions 1 and 2, can be considered robust with respect to the impact of some or all exogeneous factors considered? This question aims to understand if the modeled characteristics can be considered robust attributes of the network or if they are dependent on external factors.

Answering these specific questions, as well as devising approaches for how to find those answers for different networks is important in deciding how to model the travel time distributions across the network and may be essential in devising methods for the estimation of path travel time distributions. As mentioned at the opening of this section, however, each of these questions should be supplemented with the question of whether each added level of complexity can be modeled and whether it needs to be modeled. That is, which added details can be eliminated to make the network model more intuitive and applicable while retaining a sufficient level of accuracy?

4.2.2 Problem Definition

To summarize the problem statement introduced through the research questions in the previous section, a more formal definition of the problem is presented here. Considering the links in the network and their random variable travel times, represented as a time series through a given day or time period, the problems for spatial and temporal characterization can be defined as follows.

The spatial characterization aims to find groups of links on the network, spatially adjacent to one another, within which link travel times are highly dependent but can be considered independent from those outside their group. An important step in finding such groups is to define an appropriate goodness of fit measure, accounting for the similarity of link travel times (as they vary with time) and then find groups, clusters, or communities of links so as to maximize the goodness of fit measure.

The temporal characterization considers the time series of the travel times on any given network link and aims to find the points in time where a change in the underlying link travel time distribution is likely to have occurred. Namely, the problem is to find (continuous) periods of time where the link travel times can be considered to be observations from a single stationary underlying distribution, but a change in the distribution is detected from one period to the next. If link travel time dependencies exist, based on the spatial characterization, the aim is to find the points where link travel time distributions change jointly for groups of dependent (i.e., jointly distributed) link travel times, so as to capture the changes in the mean and covariance of the joint distribution.

4.3 Taxonomy for Stochastic Dynamic Network Modeling

As introduced in the literature review, the modeling and optimization problems in stochastic dynamic transportation networks are not entirely new. A stochastic dynamic transportation network is a model of the transportation network in which the travel times across the network links are random variables with time-varying distributions that may also exhibit spatio-temporal dependencies.

The existing literature on modeling and optimization problems in stochastic dynamic networks is extensive and diverse. However, most of the studies in this domain, as presented in the literature review in Chapter 2, approach stochastic dynamic networks from the perspective of different optimization problems within the network rather than with the goal of modeling the network itself. Models of the network underlying the problem definitions are typically treated as assumptions imposed to the network, rather than being uncovered or exposed from data or network characteristics.

Table 4-1. Taxonomy of stochastic dynamic network models

	Temporal Dimension	Spatial Dimension
General Classifications & Existing Taxonomic Categories	Do travel time distributions in the network vary over time? • Time-invariant networks, • Time-varying networks with a fixed time interval.	Are link travel times dependent random variables? • Independent link travel times, • Partial Markovian dependence between neighboring links, • Partial dependence for links in impact neighborhoods with fixed radius, • Full dependence via known joint link travel time distribution or correlation structure.

<table>
<tr><td rowspan="5">Additional Dimensions for Further Classification of Stochastic Dynamic Network Models</td><td>In time-varying networks:</td><td>In networks with spatial dependencies:</td></tr>
<tr><td colspan="2">Temporal Variation</td></tr>
<tr><td>Do link travel time distributions' time-intervals change over time?
• Time-intervals have fixed durations over time,
• Time-interval durations change over time.</td><td>Does the presence or strength of dependencies between link travel times change over time?
• Link travel time correlations are static,
• Link travel time correlations vary over time with the link travel time distributions,
• Link travel time correlations vary over time with intervals different from the link travel time distributions.</td></tr>
<tr><td colspan="2">Spatial Variation</td></tr>
<tr><td>Do link travel time distributions' time-intervals change over space?
• Time-intervals have fixed durations for the entire network,
• Time-interval durations vary over space (i.e., across the network links).</td><td>Does the presence of dependencies between link travel times change over space?
• Link travel time correlations exist for equal-size neighborhood for each link, fixed over time,
• Link travel time correlations exist for equal-size neighborhood for each link, that vary over time or changing conditions,
• Link travel time correlations exist within spatial neighborhoods of varying size and distance, fixed over time,
• Link travel time correlations exist within spatial neighborhoods that vary over time or changing conditions.</td></tr>
</table>

This taxonomy aims to present a unified overview of the modeling dimensions that can be considered in characterizing a transportation network as a stochastic dynamic system. Some of these dimensions have been extensively explored in existing literature, several have been primarily treated as assumptions, while others are new dimensions of modeling that are introduced here to give a more comprehensive model of the stochastic dynamic transportation network. Table 4-1 displays the general classification of stochastic network models and introduces the additional dimensions for further classification based on the questions introduced in the problem definition.

4.3.1 Existing Taxonomic Categories

Stochastic transportation networks are typically classified based on the temporal and spatial dependencies, as seen in the literature in section 2.1. Some additional observations to identify the existing taxonomic categories based on Table 2-1 are presented here.

In the temporal dimension, networks are classified based on whether the travel time distributions vary over time and two major categories are encountered: time-invariant or static stochastic networks and time-varying stochastic networks with fixed time intervals. The time intervals for the temporal variation of travel time distributions are fixed across time, (i.e., they have the same duration for all times of day), across space (i.e., they are equal for all network links) and are considered stationary under changing network conditions.

In the spatial dimension, the literature classifies different models of stochastic transportation networks based on dependence assumptions for the link travel time random variables. Three primary categories emerge: (1) models assuming independent link travel time distributions, (2) models assuming partial dependence, and (3) models where full dependence between link travel times is assumed. The second category can be further separated into models assuming Markovian dependence and those assuming partial dependence for neighboring links in an impact area of pre-specified radius. In the third category, a few subcategories exist: cases where dependence is modeled via a given correlation structure or via a given joint link travel time distribution, either of pre-specified distribution functional form or as discrete distributions with finite support.

The existing taxonomic categories cover the main questions in modeling stochastic transportation networks and the major assumptions that can be made in answering them. Three important limitations should be pointed out:

(1) The existing categories do not consider cross-sectional categories that may emerge from considering the temporal and spatial dimension simultaneously.

(2) The existing categories are imposed onto a given network or data set rather than uncovered from the network structure and its data.

(3) The existing categories have emerged as rather rigid classes where assumptions hold equally over the entire spatial and temporal domain of the network.

4.3.2 Extended Taxonomic Categories

An important premise of this chapter is to consider stochastic networks in the general sense and allow for underlying assumptions, such as those regarding the time-variability or spatial dependence, to vary across the temporal and spatial domain considered. The existing taxonomic categories can be seen as special cases of this unifying generalized modeling framework for stochastic dynamic networks. The taxonomic categories define here are presented in a flexible manner to allow for a data-driven modeling framework, as presented in section 4.4 below.

The bottom portion of Table 4-1 outlines the additional dimensions for further classification of stochastic dynamic networks. These categories are concerned with allowing for cross-sectional classification considering both the temporal and spatial dimension, and thus allow for fluidity of the assumptions across the network and the temporal domain under consideration. Some of the categories have been considered in previous work (Filipovska et al., 2021; Filipovska and Mahmassani, 2020b), although not explicitly presented as distinct from the existing taxonomic categories or classified according to the taxonomy presented here.

The additional dimensions for further classification of stochastic dynamic networks are addressed in the temporal and spatial dimensions with sub-categories for cross-sectional classification.

Firstly, for time-varying networks temporal and spatial variation of the time intervals can be considered. In terms of temporal variation, in asking the question of whether link travel time distributions' time-intervals change over time, two categories can be added:

- Time-varying networks with time-interval durations fixed and equal over time.
- Time-varying networks with time-interval durations that vary over time.

In terms of spatial variation, in asking the question of whether link travel time distributions' time-intervals change over space, two categories can be added:

- Time-varying networks with time-interval durations fixed and equal over the entire network.
- Time-varying networks with time-interval durations that vary across the network (i.e., may be specific to each link or neighborhood of links).

Secondly, for networks with spatial dependencies, temporal and spatial variation of the spatial dependencies (both in their strength and existence) can be considered. In terms of temporal variation, in asking the question of whether the presence or strength of correlations between link travel times change over time, three categories can be added:

- Networks with static (over time) link travel time dependencies.

- Time-varying networks with link travel time dependencies (or correlations) that vary with the link travel time distributions.
- Time-varying networks with link travel time dependencies (or correlations) that vary intervals different from those of the link travel time distributions.

Thirdly, in terms of spatial variation, in asking the question of whether the presence of dependencies between link travel times changes over space, four categories can be added:

- Networks where link travel time correlations exist for equal-size neighborhood for each link, fixed over time, where the size can be determined based on physical or network-based distances.
- Networks where link travel time correlations exist for equal-size neighborhoods for each link, that vary over time or with changing conditions.
- Networks where link travel time correlations exist within spatial neighborhoods of varying size and distance, fixed over time.
- Networks where link travel time correlations exist within spatial neighborhoods of varying size and distance, that vary over time or with changing conditions.

The following section presents a review of methods relevant for spatio-temporal network characterization to provide the basis for a methodology to address each of the limitations of existing taxonomic categories.

4.3.3 Review of Methods Relevant for Spatio-Temporal Network Characterization

This section presents a review of methods relevant for spatio-temporal network characterization that can be used for obtaining data-driven models of stochastic dynamic networks according to the extended taxonomic categories presented above.

4.3.3.1 Review of Community Finding Approaches from Network Science

To give a brief overview of the notion of community structures and community finding in networks, this section reviews some introductory and classical works in this domain. Community structure detection is a data analysis technique used for understanding large-scale network data where the goal is to find groups that a network can best be divided into, assuming that it naturally divides into subgroups based on some characteristics (Newman, 2010, 2006).

An important concept in finding community structures in networks is the notion of modularity, defined by Newman and coauthors (2004) as a property of the network and a specific proposed community structure that measures the goodness of fit for the proposed structure to the underlying network. Modularity has been the basis for the development of a whole class of community finding clustering algorithms, specifically those targeted towards large networks (Clauset et al., 2004).

The Louvain Modularity-Maximization Approach for Community Finding

A key modularity maximization approach, developed for highly practical applications and suitable for large networks, is the Louvain method introduced by Blondel and coauthors (2008). The Louvain is a heuristic introduced to address some of the challenges of the fastest modularity maximization approximation algorithm proposed by Clauset and coauthors (2004). The Louvain method recognizes that there may be several natural divisions and organization levels in large networks (where communities can be divided into sub-communities) and is thus intended to reveal this hierarchical structure. The details on the Louvain method and the corresponding algorithm can be found in the original work (Blondel et al., 2008), but some key aspects, such as the general structure of the algorithm and the modularity computation, are presented in the methodology in section 4.4.1 for completeness.

As can be expected, the Louvain approach is not readily applicable to the problem of community detection in transportation networks. There are key differences in how transportation networks are modeled and how the notions of adjacency and proximity translate for the application to transportation networks. The methodology section presents these key differences and the approach taken in this study to adequately adapt the Louvain algorithm for modularity optimization to the context of transportation networks.

Advancements in Network Community Finding

Some more recent advancements in the domain of network community detection would allow for more sophisticated approaches for community finding in transportation networks. A study by Mucha and coauthors (Mucha et al., 2010) focuses on finding community structures in time-dependent, multiscale and multiplex networks, which can be useful when considering applying community finding approaches to dynamic transportation networks. However, these methods require further adaptation to be suitable for application to transportation networks and are not suited for large-scale network implementation.

A recent review paper on the challenges and opportunities in community structure in complex networks emphasizes the relevance of time evolving networks and the extraction of communities in complex networks where nodes and links can disappear or be introduced, or attributes of the nodes and links may be variable (Cherifi et al., 2019). The authors consider snapshot-based approaches, evolutionary algorithms, online community finding, and predicting community evolution. Some extensions of the generative models of communities in complex networks may be useful for the further study and analysis of transportation network characterization. The present study aims to introduce the notion of community finding for the characterization of stochastic dynamic transportation networks, and as such the focus is on implementing and testing the most prevalent and applicable approaches and to adequately adapt them for application to transportation networks. The idea of having evolving network communities that may change over time or with the change of some exogenous or endogenous factors are considered via numerical experiments for sensitivity analysis, but it may be worthwhile to investigate more sophisticated methods in future work.

Measuring Robustness of Community Structure in Networks

To conclude this brief review, it is important to consider the notion of evaluating and comparing network community structures. As indicated at the start of this section, modularity is typically used to evaluate the goodness of fit of a given community structure on the underlying network of interest. By using modularity optimization, the approach implemented in this study aims to maximize modularity in order to obtain the best fit community structure for a given network and data. However, adjusting the parameters of the community finding algorithm or changing the conditions of the network itself may lead to different community structures of the transportation network that all maximize modularity for the problem at hand.

For an effective sensitivity analysis of the applied methodology with respect to varying parameters of the algorithm or varying network conditions, it is important to compare the resulting community structures and evaluate whether community structure tends to be robust or highly varying with respect to those factors. The notion of robustness of community structures in networks can be used for this purpose, along with approaches to quantify network perturbation, differences in community structure, and use existing information-theoretic measures for community structure comparison (Karrer et al., 2008).

The measures of similarity or difference between community assignments are typically categorized as (1) pair counting measures – evaluating the number of pairs of nodes in the same or different communities in two assignments; (2) cluster matching measures comparing the best match for each cluster in two different community assignments; and (3) information-theoretic measures for clustering. This study will use a few different measures to perform the sensitivity analysis for network community assignments in the network characterization. Specifically, pair counting measures used in this study are the adjusted Rand score and Fowlkes-Mallows score, normalized mutual information score - which a cluster matching measure, while homogeneity and completeness are information-theoretic measures. Further detail on each of the measures and their interpretation are given in the methodology section.

4.3.3.2 Review of Time Series Similarity Measures

Time series analysis is the study of time series data defined as sequences of measurements over time describing the behavior or state of systems. This section provides a brief review of methods for comparison of time series and retrieval of similar time series. An extended introductory survey on this topic is presented by Last at al. in a book on data mining in time series data bases (2004), and it should be noted that this review is oriented towards approaches suitable for this study.

Most robust distance measures for time series comparison utilize signature-based similarity search. An initial approach introduced by Agrawal et al. (1993) is the F-index in which a 'signature' of the time series is extracted from its frequency domain which preserves Euclidean distance. Piecewise constant approximation was introduced based on the idea of dividing each time series into k equal-length segments and comparing the average values of each segment as a coordinate for the k-dimensional signature vector (Keogh et al., 2001a; Yi, B and Faloutsos, 2000).

Landmark models introduced significant progress in this domain, where landmarks are identified in the time series as points of great importance. The approach introduced by Perng et al. (2000) has been shown to be invariant with respect to a number of transformations, including shifting, uniform amplitude scaling, uniform time scaling, non-uniform time-scaling (time warping) and non-uniform amplitude scaling.

For the application to transportation networks, specifically to comparing link travel times as time-series, an important challenge in comparing time-varying travel times on a pair of links comes from the difficulty in defining the time-scaling or shifting necessary to match the time series in a way that would be experienced by a vehicle on the network. If one were to consider a vehicle traversing a pair of links on the network (consecutive or otherwise) that vehicle will experience travel times on the two links with a certain lag in between. However, considering a string of vehicles, one might observe that the time-lag changes over time due to shifting traffic patterns. This complexity becomes difficult to model from a mathematical or physical perspective, and data mining methods may be useful to find and extract existing patterns.

Dynamic Time Warping (DTW) has been suggested as a technique for more robust distance calculations for time series comparison. The DTW technique was introduced for the purpose of finding patterns in time series data by Berndt and Clifford (1994). The authors demonstrated the usefulness of the approach and pointed out its limitations to being implemented in large data bases. DTW for time series data analysis has been advanced significantly via techniques to improve on its computational limitations, specifically by modifying DTW to operate on higher-level abstractions of the data via piecewise aggregate approximation (Keogh and Pazzani, 2000) and exact indexing of DTW (Keogh and Ratanamahatana, 2005). Details on the DTW technique, the

Piecewise DTW (PDTW) algorithm, and exact indexing approximation can be found in the original papers (Berndt and Clifford, 1994; Keogh and Ratanamahatana, 2005; Keogh and Pazzani, 2000). In summary, DTW aims to find an alignment of two time series that may have an overall similar shape but are misaligned in a non-linear manner in the time axis. DTW finds the most appropriate shifting and scaling (even if non-uniform) so that the patterns in the time-series are aligned. Then, distance can be calculated for two aligned time series observations, often referred to as the minimum DTW distance. More rigorous treatment of DTW and how it used in this study is provided in the methodology section below.

4.3.3.3 Review of Time Series Change Point Detection Approaches

In dynamic and complex systems, and change point detection (CPD) is the problem of finding the points where changes in time series data occur, often viewed as changes in the underlying probability distributions due to changes in the state or the system behavior (Aminikhanghahi and Cook, 2017; Last et al., 2004). CPD is closely related to the idea of time series segmentation but has been defined more broadly and studied in the fields of data mining, statistics, and computer science.

Classification of Change Point Detection Approaches

Change point detection algorithms are typically classified as batch (i.e., offline) or online (Last et al., 2004). In addition to having different computational requirements, the key difference is that offline algorithms consider the entire time series all at once, while online algorithms observe the process as it occurs and aim to detect a change point (often associated with an anomaly or attack on the system) as soon as possible after its occurrence.

Another important distinction is between supervised and unsupervised methods for change point detection. Like with other machine learning or statistical approaches, supervised learning requires labeled data, and a model is trained to accurately replicate the known labels and accurately assign labels to new observations. Unsupervised methods, on the other hand, discover patterns in unlabeled data, which in the context of change point detection would mean finding change points based on statistical features of the data itself without knowing the 'true' change points beforehand (Aminikhanghahi and Cook, 2017).

Most change point detection methods can be classified into three main categories: sliding windows, top-down, and bottom-up approaches (Last et al., 2004). Sliding window approaches start with a small segment that is grown until some error bound is exceeded and then starts a new window with the next data point. Top-down approaches recursively partition the time series until stopping criteria are met, while bottom-up methods start with the finest possible segmentation and merge segments until stopping criteria are met (Keogh et al., 2001b). Each of these algorithms have their advantages and shortcomings and the details can be found in some classical texts and review studies on the topic (Aminikhanghahi and Cook, 2017; Keogh et al., 2001b; Last et al., 2004).

Change Point Detection Models

In addition to considering is the application and available data are suitable for supervised or unsupervised modeling, online or offline applications, and choosing the specific algorithm to use, one also needs to choose the model. In CPD methods, the model is typically implemented as a cost function that then defines the stopping criteria for the chosen algorithm. Further, there are a variety of changes that may be considered in multi-dimensional time-series, typically depending on the application at hand. For example, one may wish to find a change in the probability distribution of any one of the variables and thus consider them jointly, or simply to find change in the distribution of the target variable and then see the corresponding changes in other variables.

A comprehensive review of CPD models and cost functions in the context of offline CPD methods is presented by Truong et al. (2020). The authors classify cost functions as parametric and non-parametric and provide an overview of the most commonly used models in each category. Parametric models are based on assumptions about the data at hand, while non-parametric models can be more robust when features of the data are not known or the research question is posed as a data mining problem.

To maintain the generality in characterizing stochastic dynamic networks, this study will focus primarily on non-parametric methods. Early non-parametric models have focused on least absolute deviation detecting changes in the median of a signal without assuming an underlying distribution (Bai, 1995) or similarly least squared deviation detecting mean-shifts in the signal. An interesting group of models for multivariate data CPD are those based on rank statistics. Using rank statistics is a popular strategy to deriving distribution-free statistics in statistical inference. This notion has been applied to finding multiple change points to multivariate time-series so as to measure the joint behavior of the marginal rank statistics of each coordinate (Lung-Yut-Fong et al., 2015; Truong et al., 2020). Another common set of approaches is kernel-based detection, which require a user-defined kernel function (Arlot et al., 2019; Truong et al., 2020). The advantage of the kernel transformation is that the change point is then applied to detect mean-shifts in the appropriately transformed time-series. This study compares these four main models for CPD methods and ultimately selects the rank statistics model as the most general approach that allows for changes in the full rank of multivariate time series to be considered and detected. A question that may be interesting for future work is to consider if some of the simpler and more restricted models may be sufficient from an accuracy perspective.

Measuring Robustness of Change Point Detection Results

To conclude this review on CPD approaches, this section considers the evaluation and comparison of change point detection results. As previously seen in the case of community structures, in order to perform the sensitivity analysis and evaluate if and how the change points detected may vary under certain conditions measures of similarity or robustness for change-point models on a time series are needed. For this purpose, the comparison of change point models is seen as a comparison of clustering configurations, similarly to the case of community characterization. Therefore, the main ideas for measuring robustness remain the same and the same measures are applicable. In this study, the Rand index and Fowlkes-Mallows Index were used as the CPD robustness measures.

4.4 Methodology

This section presents the methodology for the problem of characterizing a stochastic dynamic transportation network considering the spatial and temporal dimensions, along with cross-sectional spatio-temporal aspects of the network, based on network characteristics revealed from travel time data in the network and according to the extended taxonomic categories presented in section 4.3.2 and Table 4-1. The methodological framework is complete with methods from network theory and time series analysis that can be used for data-driven stochastic dynamic network characterization.

This section is broken down into three subsections, each focusing on an aspect of network characterization and the corresponding methodology. Firstly, section 4.4.1 considers spatial characterization of the network, where the goal is to detect the existence of a community structure of the transportation network. In this study, the application of existing network community detection methods to transportation networks requires a few additional steps due to some differences in how a network is viewed and modeled.

Secondly, section 4.4.2 considers temporal characterization of the network, where the goal is to detect time-intervals within a given temporal domain so that link travel time distributions can be considered constant within each time-interval and changing from one interval to the next. Equivalently, the aim is to identify points in time when the link travel time distribution changes, given that travel times on a given link over a specified temporal domain form a time series. This question will be approached from the standpoint of time-series analysis, where the study of change point detection focuses on this problem precisely: to detect if and when changes may have occurred in the probability distribution of a stochastic process or time series.

Finally, section 4.4.3 considers the intersection of the spatial and temporal domain, bringing together the components from the earlier two to provide a complete methodology for spatio-temporal stochastic dynamic network characterization.

4.4.1 Spatial Characterization via Network Community Detection

This section considers the spatial characterization for stochastic dynamic networks, aiming to identify neighborhoods of network links such that link travel times within a given neighborhood are dependent, while dependencies can be considered negligible outside a links' neighborhood. For this purpose, community detection approaches from network science are considered as the starting point. Most community detection approaches are in fact clustering methods that account for the underlying structure of the network so that clusters are formed by considering adjacent nodes. As mentioned above, the application of network community detection methods to stochastic dynamic transportation networks requires some adaptations of existing approaches due to some differences in how a network is viewed and modeled.

4.4.1.1 Transportation Network Structure Adaptation

Network community structure is typically built considering the network vertices as objects to be clustered based on their connectivity and closeness (or similarity) to one another considering from the edges (i.e., links) in the network. The most immediate difference with transportation networks comes from the fact that in transportation networks travel is primarily experienced on links (i.e., edges) and nodes (or vertices) are seen the connectors between those links. Additionally, travel time is associated with links and in stochastic dynamic networks link travel times are the random variables to be modeled. Edge community finding approaches exist in network science but are not as common, varied, or scalable as those for node community detection.

Therefore, unlike typical network community finding approaches based on node clustering, in this study the goal is to identify groups or communities of edges rather than vertices. In order to allow for the implementation of common network clustering approaches, the transportation network is to be transformed, so that the network links can be modeled as graph vertices that hold certain properties and are to be placed in communities. Thus, the transportation network nodes are the objects that hold link adjacency information, and the adjacency of links is determined by converting the transportation network nodes to graph edges so that neighboring links can remain adjacent.

A note on terminology is added here to clarify these distinctions in upcoming sections. For the sake of clarity, the transportation network will be referred to as network and its elements nodes and links, while the network science notion of network will be termed graph with its elements vertices and edges. Thus, the transformation of the transportation network into a graph is performed by viewing the network links as vertices of the graph, and the network nodes as edges of the graph.

4.4.1.2 Link Travel Time Similarity for Community Detection

The second important difference when considering community finding in transportation networks relates to the idea of proximity or similarity of the network nodes. Network science applications typically model the distance between vertices in the graph so that community finding is based on the density of connections between vertices and the proximity between a pair of vertices contributes to how densely connected a community would be. Clustering approaches typically use an adjacency matrix that contains information on the adjacency of graph vertices or a weighted adjacency matrix where vertex connections also have a weight that is interpreted as a stronger connection.

In this application, however, the transportation network links are to be grouped based on the dependence or similarity of link travel times, rather than any measure of physical proximity. While it is important that the links' adjacency remains under consideration, the implementation of community detection approaches would also require a measure of similarity between link travel times in lieu of the network science notion of proximity or adjacency weights.

It should be noted that some community discovery algorithms have been developed to specifically focus on clustering graph vertices with attributes so as to account for both the network structure and the vertex attributes. Modeling the link travel times as attributes of the network links and implementing network attribute community clustering approaches could be one option to addressing the challenge at hand. However, this idea would be insufficient for dynamic transportation networks where the travel times are time-varying attributes of the network links.

Furthermore, considering vectorized link travel times (i.e., time-series) as node attributes in these approaches would be equivalent considering snapshots of travel times across network links for a given time, which also has significant limitations. Transportation network dynamics are much more complex, specifically when considering time-varying link travel times, and the similarity of travel times along a pair of links should be considered over time and ideally account for the time-lag between a given vehicle's traversal of the two links.

This study considers more suitable measures of link travel time similarity that are to be derived based on the time-series of link travel times, i.e., travel times experienced on each link across time. As reviewed previously in section 4.3.3.2, a number of time series similarity measures exist in time series analysis, each for different applications and capturing different aspects of time series comparison. To maintain the generality of the approaches and avoid imposing characteristics onto the time-series comparisons, this study considers dynamic time warping (DTW) as a method to calculate the optimal match between two time series and then determine the distance between the two aligned time series.

4.4.1.3 Louvain Community Finding Approach with Dynamic Time Warping

<u>Overview of Modularity Optimization</u>

As introduced in the literature review on network community structures in section 4.3.3.1, a class of community finding approaches have been developed around the notion of modularity optimization (or maximization). Modularity measures the goodness of fit of a specified community division or assignment for a given network (Clauset et al., 2004).

Originally, Newman and coauthors defined modularity (2004) with the notion of adjacency of network nodes: if A is an adjacency matrix for a given network, then the adjacency element A_{vw} is 1 if vertices v and w are connected and 0 otherwise. If the vertices are divided into communities such that the vertex v belongs to community c_v then the fraction of edges that lie in the same community can be defined as

$$\frac{\sum_{vw} A_{vw}\delta(c_v, c_w)}{\sum_{vw} A_{vw}} = \frac{1}{2m}\sum_{vw} A_{vw}\delta(c_v, c_w)$$

where $\delta(c_v, c_w)$ is 1 if $c_v = c_w$ and 0 otherwise, and m is the number of edges in the graph $m = \frac{1}{2}\sum_{vw} A_{vw}$.

It can be observed that this fraction will be large if there are many within-community edges, thus large for good divisions of the network. A useful measure is obtained if from this fraction is subtracted its expected value for a randomized network, since on its own this fraction will be maximized by assigning all vertices into a single community. By letting k_v be defined as the degree of the vertex v i.e., the number of edges incident upon it $k_v = \sum_w A_{vw}$ the modularity Q is defined as follows:

$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\delta(c_v, c_w),$$

where the probability of an edge existing between vertices v and w in a randomized network with the same vertex degrees is $k_v k_w/2m$.

Therefore, in this definition of modularity, nonzero values all represent deviation from randomness and the authors point out that a value above 0.3 is a good indicator for significant community structure in a network.

The Louvain Algorithm for Modularity Optimization in Large Networks

The community finding approach used in this study is the Louvain algorithm (Blondel et al., 2008), initially introduced in the literature review section, which expands on and develops a fast heuristic suitable for large networks based on the initial work of Newman and coauthors

(Clauset et al., 2004; Newman and Girvan, 2004). Blondel and coauthors use the same definition of modularity Q as defined above, often referred to as the Newman measure of modularity.

The Louvain method uses a simple algorithm of two phases repeated iteratively. This section will provide a brief summary of the method, but the details can be found in the original work (Blondel et al., 2008). An important advantage of the Louvain algorithm is that it is extremely fast with linear complexity on linear and sparse data. An additional benefit to the application of transportation networks that is used here is the opportunity to use weighted network edges – and the Louvain algorithm is defined assuming a weighted network of N vertices, where the simple 1-0 non-weighted network can be seen as a special case.

Phase 1:

Assign each vertex to its own community, i.e., there are as many communities as there are vertices. For each node v:

Consider the neighboring vertices $w \in N$, $w \neq v$ and $A_{vw} \neq 0$.
Evaluate the gain in modularity if v is removed from its community and assigned to the community of w.
Place v in the community for which the gain is maximum and positive.

Stop when no individual move will improve modularity.

Phase 2:

Build a new graph whose vertices are the communities found in Phase 1. The weights of the edges between the new vertices are the sum of weights of edges between the nodes in the two corresponding communities.
Go to Phase 1 for the next iteration.

According to the authors, part of the algorithm efficiency is due to the ease of computing the gain in modularity for moving a single node v into a community C, as follows:

$$\Delta Q = \left[\frac{\sum_{in}\cdot + k_{v,in}}{2m} - \left(\frac{\sum_{tot}\cdot + k_v}{2m}\right)^2\right] - \left[\frac{\sum_{in}\cdot}{2m} - \left(\frac{\sum_{tot}\cdot}{2m}\right)^2 - \left(\frac{k_v}{2m}\right)^2\right]$$

Where $\sum_{in}\cdot$ is the sum of the weights of edges inside C, $\sum_{tot}\cdot$ is the sum of edges incident to nodes in C, k_v is the sum of weights on the edges incident to v and $k_{v,in}$ is the sum of weights on the edges from v to nodes in C and m is the sum of all weights of all edges in the graph.

The implementation of a weighted adjacency matrix in the community finding algorithm is an important aspect of implementing these approaches in the context of transportation networks. The computation of these weights can be specific to the application at hand, but it is important to note that when using the adjacency matrix for modularity optimization higher edge weights indicate a stronger relationship between the pair of vertices relative to lower edge weights. The remainder of this section focuses on applying the Louvain method to transportation networks adapted as described in the previous sections.

Dynamic Time Warping for Link Travel Time Similarity

In order to appropriately define the weight of edges in the adapted network, where transportation network links are seen as graph vertices and the nodes between them serve as connecting edges this study uses dynamic time warping (DTW) due to its generality and robustness, to avoid imposing characteristics or limitations on the time series comparisons and instead extract the information from the data itself. The review of the DTW technique is extended here so as to explain its application to the present study and is based on the paper introducing the exact indexing of DTW approximation used in this study (Keogh and Ratanamahatana, 2005).

Let Q and C be two time series, potentially of different lengths n and m respectively. The two sequences are aligned by constructing an n-by-m matrix where the (i, j) element contains the distance between the two points and corresponds to an alignment between the points q_i and c_j. Then a warping path W is a set of matrix elements that defines a mapping between the time series where the k_{th} element of W is defined as $w_k = (i, j)_k$ and the total number of elements in W is at least equal to $\max(m, n)$ and at most equal to $m + n - 1$. In addition, the warping path is subject to boundary conditions, continuity, and monotonicity constraints. As there may be several eligible warping paths, DTW alignment is based on the path that minimizes warping cost:

$$DTW(Q, C) = \min \sqrt{\sum_{k=1}^{K} w_k} \,.$$

This definition allows for Euclidean distance to be seen as a special case of DTW where $w_k = (i, j)_k, i = j = k$ and as can be expected the two sequences and the warping path must all have the same length T.

To the DTW distance as a similarity metric in this study, the minimum DTW distance is converted to a similarity measure by using the inverse. In this study DTW distance never equals 0 except for the trivial case when finding the distance between two of the same time series, thus the inverse – potentially scaled by a constant depending on the order of magnitude, were used as weights in the adjacency matrix.

4.4.2 Temporal Characterization via Time-Series Change Point Detection

This section considers the temporal characterization for stochastic dynamic networks, aiming to identify time-intervals of the temporal domain where link travel time distributions are constant. Change point detection approaches from time series analysis focus on this problem and are designed to identify points in time when the probability distribution of a time series changes. For a time period between two change points, the observations can be considered to be drawn from the same distribution, different from the distribution that the observations beyond the change points.

CPD approaches would allow for the time varying aspect of travel times in stochastic networks to be modeled mathematically and in a data-driven manner. In CPD, the resulting time intervals are not restricted to uniform duration across time or across the different links. Namely, link travel time distributions can vary with time intervals that change over time (e.g., with the time of day) and across space (e.g., from link to link, or from neighborhood to neighborhood).

A brief review on the classes of change point detection (CPD) algorithms and models were given in the literature review. This section provides the problem formulation and expands on that brief review in order to introduce and explain the specific CPD methodology used in this study.

4.4.2.1 Change Point Detection Algorithms for Link Travel Times

CPD approaches can be applied to for multi-variable time series change point detection when the time series can be seen as observations from a joint multivariate distribution. In the case of this study, since link travel times are considered to be correlated random variables, multivariate time series analysis is needed. An important consideration when using multivariate CPD approaches is that the random variables are seen as jointly distributed, so a change in the underlying distribution for one of the variables translates into a change of the joint distribution and thus the change points are the same across all jointly distributed variables. This approach has obvious shortcomings in that it may create an unnecessarily high number of change points due to changes in any one of the time series. On the other hand, the joint distribution cannot be ignored as it is requisite for detecting changes in the covariance between variables that would be missed in the case of single variable time series CPD. Therefore, an important question in implementing multivariate CPD is to decide when to model variables as jointly distributed and when any potential dependencies can be ignored in order to avoid an excessively large number of change points.

This study models link travel times as jointly distributed for links in the same community based on the community detection results from the previous methodology section. With this approach a single set of change points (or time series segmentations) are obtained for each community of links that are spatially near one another and exhibit significant similarity as measured by the inverse DTW distance.

Consider a multivariate non-stationary random process $y = \{y_1, \dots, y_T\}$ with values in $\mathbb{R}^d (d \geq 1)$ and T samples. In CPD, the signal y is assumed to be piecewise stationary, meaning that it can be separated into time-periods so that within each time period the underlying random process is stationary and comes from a single distribution, and then the characteristics of the process change abruptly at some unknown time instants $t_1^* < t_2^* < \cdots < t_K^*$. CPD consists of estimating the indexes t_k^* and in some cases the total number of changes K, and is a model selection problem of choosing the best possible segmentation $\mathcal{T} = \{t_1, \dots, t_K\}$ so as to minimize a quantitative criterion $V(\mathcal{T}, y)$.

A number of CPD algorithms were considered for this study, based on those included in the literature review section, including kernel-based approaches, sliding-window, bottom-up and binary segmentation approaches. To select the preferred algorithm this study relied on insight from the literature and a small sensitivity analysis that showed that all algorithms yielded the same change points for fixed parameters that define the stopping criteria, but there were some differences in computational run time. The bottom-up algorithm was selected due to its superior performance in terms of computational run times. Some benefits of bottom-up segmentation are the low complexity, its ability to perform single or multiple change point detection, and that it can work with a known or unknown number of change points. Algorithmic analysis of the bottom-up approach is presented by Keogh et al. (2001b) and (Fryzlewicz, 2007).

The algorithms used and tested here were implemented in Python according to the implementations presented by Truong et al. (2020) and the Python package ‘ruptures’ presented in this survey paper. In these algorithms, the assumption is that the criterion function $V(\mathcal{T}, y)$ for a particular segmentation of a signal y is the sum of costs of all segments that define the segmentation:

$$V(\mathcal{T}, y) \coloneqq \sum_{k=0}^{K} c(y_{t_k \dots t_{k+1}})$$

where $c(\cdot)$ is a cost function measuring goodness-of-fit for the sub-signal to a specific model.

4.4.2.2 Implementation of Change Point Detection Algorithms

One of the benefits of the bottom-up algorithm for CPD, as indicated above, is that it can work whether the number of change points is known beforehand or not. The vast majority of CPD methods are better or only suited for problems where the number of different regimes, and thus the number of change points, in the time-series is known and the task of the algorithm is to find the best locations for those known number of change points. However, for most data mining applications the number of change points is part of the unknowns in the problem and thus the task of the algorithm is a more complex one – to determine both how many change points there are and when they occur.

Implementing algorithms to search for the number and (temporal) location of the time series change points requires different types of stopping criteria and specifications. Without proper specifications, the tendency of the algorithm is to detect a change point where there is even the smallest variation in the time series values. Three key parameters can be modified to obtain the desired solution quality for most CPD algorithms, including the bottom-up approach. In the simplest case the algorithm takes an input of the number of break points in the time series. Alternatively, one can specify a complexity penalty parameter p or a threshold on the residual norm ϵ. According to Truong et al. (2020) the complexity penalty parameter works as a constraint that balances the goodness-of-fit term and the actual choice of the complexity penalty relates to the amplitude of the changes to detect. A very low penalty can lead to the detection of too many change points, even ones that are actually the result of noise, while too much penalization may detect only the most important changes or none at all. For problems with unknown number of changes, the problem presented above is modified so as to have the objective function $\min_{\mathcal{T}} V(\mathcal{T}) + p(\mathcal{T})$. The most common and popular choice is the linear penalty (also known as the l_0 penalty) where a smoothing parameter β needs to be calibrated and $p(\mathcal{T}) = \beta|\mathcal{T}|$. The penalty used in this study is based on the Bayesian Information Criterion and Akaike Information Criterion, defined as:

$$p(\mathcal{T}) = |\mathcal{T}| \log(T)\, \sigma^2$$

as presented and explained by Truong et al. (2020). In this notation σ is the standard deviation of the signal noise, where the signal would be the observed link travel times. Including this value in the penalty term ensures that changes in link travel time are ignored if they are likely be due to the standard deviation of the underlying noise in observation.

4.4.3 Spatio-Temporal Stochastic Dynamic Network Characterization

Having presented the methodology for spatial characterization of the transportation network via community structures and for temporal characterization via time series change point detection, this section briefly presents the overall methodology used for spatio-temporal stochastic dynamic network characterization that brings these two components together. Opportunities exist for further exploration and implementation of more complex or extended versions of the presently used methods to obtain potentially more realistic models of the transportation network. This methodology aims to provide the starting point for implementing these methods in the domain of transportation networks. The methodology is separated into two stages, as presented below.

The first stage consists of community finding in the transportation network, given full 24-hours of observations. The underlying assumption is that the community structure or community assignment of network links will be fixed over time and developed on the basis of the inverse of the minimum dynamic time warping (DTW) distance for any pair of full-day link travel time observations in the form of time series. The resulting community structure and corresponding community assignments are then interpreted as spatial neighborhoods (continuous on the network itself) that are composed of links with high link travel time similarity according to the inverse minimum DTW distance. Therefore, the assumption is that link travel times are considered dependent for any pair of links within the same community, but dependencies can be disregarded for links placed in different communities.

The second stage builds on the community structure and consists of change point detection (CPD) for all sets of jointly distributed multivariate link travel time distributions as time series observed over a full 24 hours. The CPD is applied to each community in the network assuming the joint distribution of link travel times in the community. CPD for multiple variables implies that the change is detected for the joint distributions and the change points are thus the same across all variables considered simultaneously. For this application, that would translate to having change points and corresponding time intervals that are the same across all links within a given community, but different communities may have different change points.

4.5 Sensitivity Analysis for Stochastic Dynamic Network Characterization

The methodology for characterizing stochastic dynamic transportation networks was tested and evaluated via a sensitivity analysis set up to answer two key research questions. Firstly, by implementing the methodology on a given network and data set with varying modeling parameters, the sensitivity analysis evaluates how the network characterization varies with the main modeling parameters and if it can be considered robust with respect to those parameters. This aspect of the sensitivity analysis also aimed to assess for values of ranges for the key parameters that are especially suitable for the network at hand. Secondly, by implementing the methodology with fixed modeling parameters over different operational conditions in the network, the sensitivity analysis was set up to evaluate how the network characterization varies with changing network conditions. Specifically, with varying demand patterns and weather conditions were used to evaluate their impact on the network characterization and test for its robustness.

4.5.1 Study Sites and Data

The sensitivity analysis for this study was performed on the large-scale Chicago network consisting of 4805 links and 1578 nodes. This study used simulated data that would allow for a sensitivity analysis to test for the robustness of the network characterization over varying exogenous and endogenous changes. Simulations were performed using a state of the art simulation-based dynamic traffic assignment tool DYNASMART-P (Mahmassani et al., 2004), previously calibrated to perform simulations for a varying operational conditions based on varying weather and demand conditions from real historical data from the network of Chicago (Yelchuru et al., 2017). Defining and calibrating these operational conditions is outside the scope of this study. For this study, simulated data were obtained for a total of 9 cases: a base case of 'regular' weekday demand pattern and clear weather conditions, four cases of varied demand patterns with clear weather conditions, and four weather cases with the regular weekday demand pattern to be compared to the base case.

Link travel times were extracted for the full day for each case with a frequency of one observation per minute, resulting in 1440 travel time observations for each link for each of the nine cases. The simulation output allows for travel time information to be extracted in two ways: (1) using vehicle trajectory data and considering the start and end times on the given link for any vehicle that traverses that link at the given time, or (2) using detector data and converting the aggregated speed observation into travel time for a given link and at the given time. For this study, travel time data were extracted primarily from vehicle trajectory data and supplemented with 1-minute detector data for times where no vehicle was detected on a given link.

4.5.2 Sensitivity Analysis of Modeling Parameters

The first part of the sensitivity analysis was designed to understand and evaluate the robustness of the network characterization while varying the modeling parameters for the spatial and temporal characterization methods introduced in the methodology section.

4.5.2.1 Parameter Selection and Values

The Louvain algorithm for community detection has two primary parameters that could be varied to potentially result in different network community assignments. The first parameter, optimization tolerance ϵ_{opt}, is the minimum increase in the objective function to enter a new optimization pass (i.e., first stage) in the Louvain algorithm. The second parameter, aggregation tolerance ϵ_{agr}, is the minimum increase in the objective function to enter a new aggregation pass (i.e., second stage). If the objective function cannot be increased by at least ϵ_{opt} with another optimization pass or ϵ_{agr} with another aggregation pass the algorithm terminates. Based on preliminary tests on the network data, the following values for the two parameters were selected for the sensitivity analysis.

$$\epsilon_{opt} \in \{0.00001, 0.0001, 0.001, 0.01, 0.1\};\ \epsilon_{agr} \in \{0.00001, 0.0001, 0.001, 0.01, 0.1\}$$

In the bottom-up approach for change point detection two parameters can be varied for different change point results. The first is considered a parameter in the implementation according to Truong at al. (2020) but is in fact the choice of model, as described in the literature review section. Different models measure different types of changes in the underlying travel time distribution and were included in this sensitivity analysis as a way of seeing if the use of a more complex model would have better sensitivity and result in more change points. Four different models were tested in this study: l_1 i.e., least absolute deviation, l_2 i.e., least squared deviation, K_{rbf} i.e., kernelized mean change with the radial basis function (rbf) kernel, and c_{rank} i.e., a rank-based model with the empirical mean and covariance matrix for the complete rank signal. The second parameter is the penalty term, previously presented in the methodology in section 4.4.2. The penalty term was implemented as a constraint specified by the penalty parameter p, a function of known variables which may vary between spatial and the assumed standard deviation of noise σ. The value of σ specified the amplitude of small changes in travel time observations that can be ignored as being due to noise of observations rather than the underlying distributions. Therefore, in this analysis $p(\mathcal{T}) = |\mathcal{T}|\log(T)\,\sigma^2$ was varied based on values of σ to be tested, $\sigma^2 \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

Based on these parameters and their specified values, this portion of the case study considered 5 values for each of the parameters ϵ_{opt} and ϵ_{agr} resulting in 25 different community specifications. For each of those, CPD was tested with a total of 40 parameter combinations from the 4 models and 10 values for the penalty parameter, resulting in a total of 1000 cases for the complete spatio-temporal specification of the network. For the sensitivity analysis with exogenous and endogenous changes in section 4.5.3, the 1000 cases were tested using the data for each of the 9 days, for a total of 9000 experiments.

4.5.2.2 Sensitivity Analysis Results for Network Community Structure

In this section, the sensitivity analysis results for the network community structure are considered while varying the Louvain algorithm parameters ϵ_{opt} and ϵ_{agr}. Three sets of results are presented: (1) characteristics of the network community structures for each combination of parameters, (2) cluster comparison using the robustness scores, and (3) visual representations of the community structures for a few of the cases.

Network Community Structure Characteristics

The network community characteristics included in this portion of the analysis are the numbers and sizes of communities in each community assignment resulting from the varied parameter values. Table 4-2 shows the number of communities obtained from each of the parameter combinations, a total of 25 network community structures obtained, and Table 4-3 shows information on the size of the communities, including mean, maximum, and minimum community size for each of the 25 community structures.

Table 4-2. Number of communities for 25 network community structures varying tolerance parameters for aggregation and optimization.

		ϵ_{agr} values				
		0.00001	**0.0001**	**0.001**	**0.01**	**0.1**
ϵ_{opt} values	**0.00001**	60	60	68	68	194
	0.0001	60	60	69	69	194
	0.001	60	60	69	69	194
	0.01	60	60	70	70	200
	0.1	58	59	64	96	240

Table 4-3. Mean, maximum, and minimum community size for 25 network community structures varying tolerance parameters for aggregation and optimization.

Mean community size						
		ϵ_{agr} values				
		0.00001	**0.0001**	**0.001**	**0.01**	**0.1**
ϵ_{opt} values	**0.00001**	80.08	80.08	70.66	70.66	24.77
	0.0001	80.08	80.08	69.64	69.64	24.77
	0.001	80.08	80.08	69.64	69.64	24.77
	0.01	80.08	80.08	68.64	68.64	24.03
	0.1	82.84	81.44	75.08	50.05	20.02

Maximum and minimum community size											
		ϵ_{agr} values									
		0.00001		**0.0001**		**0.001**		**0.01**		**0.1**	
ϵ_{opt} values	**0.00001**	217	2	217	2	217	2	217	2	154	2
	0.0001	217	2	217	2	217	2	217	2	154	2
	0.001	217	2	217	2	217	2	217	2	154	2
	0.01	219	2	219	2	219	2	219	2	148	2
	0.1	237	4	237	2	237	2	214	1	137	1

Comparing the characteristics of the network community structures obtained from the sensitivity analysis in Table 4-2, the number of communities for most cases is close or equal to 60. For the cases as ϵ_{opt} varies from 0.01 to 0.00001 and ϵ_{agr} goes from 0.0001 to 0.00001, the number of communities is 60 and the maximum, mean and minimum size of resulting communities are also equal. This shows that there may possibly be a convergence in the community structures, since reducing the tolerance parameters did not yield further aggregation of the communities or improvement in the optimization of modularity.

Increasing the value of ϵ_{opt} to 0.1 results in a slight decrease in the number of communities for $\epsilon_{agr} \in \{0.00001, 0.0001\}$, showing possible robustness of the community characteristics with respect to the optimization phase. However, as ϵ_{agr} increases, the number of communities can increase to close to 70 and then close to 200 for $\epsilon_{agr} = 0.1$, signifying that 0.1 may be too large of a value for this parameter as it does not have the ability to further aggregate the community structure.

From the values in Table 4-2 and Table 4-3, the 6 cases with the same community characteristics in fact had the same modularity as modularity optimization had reached a limit at approximately 0.94781 (and modularity is bounded by 1). Decreasing the tolerance for aggregation to 0.01 from 0.1 resulted in an increase in modularity close to 0.1 but a significant decrease in the number of communities through aggregation, a relatively high value for modularity can be achieved with a large number of small communities but the final marginal increase of modularity by 11% had a significant impact on the community structure and yielded a potentially robust structure. Further aggregations did not have significant improvements in modularity i.e., less than $1E^{-5}$ equivalent to 0.001% increase.

Community Robustness Measures

The same set of community structure results were further compared via clustering similarity and robustness measures that primarily pertain to quantifying the overlap of different cluster community structures. Three different measures were used for community robustness evaluation, as introduced in the literature review and methodology sections:

- Adjusted Rand Score (ARS) is the ratio of the number of pairs clustered in the same way (i.e., either in the same community or in different communities) in both assignments to the total number of pairs, adjusted with the expectation of the null (random) model.
- Fowlkes-Mallows Similarity (FMS) coefficient is the geometric mean between the precision and recall and is the number of pairs of points in the same clusters is accounted for while not accounting for the value of points being placed in different clusters, which would have resulted in a skew towards 1 in the ARS.
- Normalized Mutual Information score (NMI) is a normalization of the mutual information score where the similarity between two labels of the same data is compared independently of the values of cluster labels.

Since similarity measures were computed for all pairs of the 25 community assignments, there are $(25^2 - 25)/2$ values calculated for each measure, and the results were separated across the tables below. For easier reading of the results, the identifiers assigned to the 25 cases are shown in Table 4-4 to be used to refer to these 25 cases in the corresponding results.

Table 4-4. Reference identification numbers for the 25 cases of community structures

		ϵ_{agr} values				
		$\mathbf{0.00001}$	$\mathbf{0.0001}$	$\mathbf{0.001}$	$\mathbf{0.01}$	$\mathbf{0.1}$
ϵ_{opt} values	$\mathbf{0.00001}$	0	1	2	3	4
	$\mathbf{0.0001}$	5	6	7	8	9
	$\mathbf{0.001}$	10	11	12	13	14
	$\mathbf{0.01}$	15	16	17	18	19
	$\mathbf{0.1}$	20	21	22	23	24

Table 4-5 shows the cases in which significant robustness was observed (highlighted in yellow in Table 4-4) with their values for ARS, FMS and NMI. Values along the diagonal are not relevant, and the tables are symmetric.

Table 4-5. ARS, FMS and NMI values for 8 cases of community assignments

ARS									
		ID 2							
ID 1		0	1	5	6	10	11	15	16
	0	–	1.00	0.99	0.99	0.98	0.98	0.97	0.97
	1		–	0.99	0.99	0.98	0.98	0.97	0.97
	5			–	1.00	0.99	0.99	0.96	0.96
	6				–	0.99	0.99	0.96	0.96
	10					–	1.00	0.97	0.97
	11						–	0.97	0.97
	15							–	1.00
	16								–
FMS									
		ID 2							
ID 1		0	1	5	6	10	11	15	16
	0	–	1.00	0.99	0.99	0.98	0.98	0.97	0.97
	1		–	0.99	0.99	0.98	0.98	0.97	0.97
	5			–	1.00	0.99	0.99	0.96	0.96
	6				–	0.99	0.99	0.96	0.96
	10					–	1.00	0.97	0.97
	11						–	0.97	0.97
	15							–	1.00
	16								–
NMI									
		ID 2							
ID 1		0	1	5	6	10	11	15	16
	0	–	1.00	1.00	1.00	0.99	0.99	0.99	0.99
	1		–	1.00	1.00	0.99	0.99	0.99	0.99
	5			–	1.00	0.99	0.99	0.99	0.99
	6				–	0.99	0.99	0.99	0.99
	10					–	1.00	0.99	0.99
	11						–	0.99	0.99
	15							–	1.00
	16								–

From Table 4-5, NMI values for most pairs were equal to 1 or 0.99 and similarly ARS and FMS values were very close to 1, confirming that the community structure for the network is robust for these 8 cases with the threshold parameters $\epsilon_{opt} \in \{0.01, 0.001, 0.0001, 0.00001\}$ and $\epsilon_{tol} \in \{0.0001, 0.00001\}$.

To compare these fairly robust clustering assignments with those with different characteristics, the next set of results compare case 6 with the cases for which $\epsilon_{agr} = 0.1$, in the far-right column highlighted in green in Table 4-4. The FMS values are shown in Table 4-6, since the ARS and NMI values showed similar results.

The FMS values show some robustness of the less aggregated cases, i.e., all except case 6. Comparing the case 6 with the less aggregated clustering results show relatively high values for FMS, above 0.84 and for three cases equal to 0.88, thus showing that even though the characteristics of these assignments were very different in terms of the number and size of resulting communities, the actual community assignments remained similar. Since the robustness scores measure similarity of the clustering assignments based on overlaps and the Louvain approach performs hierarchical clustering, it is likely that the more aggregated structures were obtained by the merging of clusters from the less aggregated cases.

Table 4-6. FMS values for 6 cases of network community structures

		ID 2					
		6	**4**	**9**	**14**	**19**	**24**
ID 1	**6**	–	0.88	0.88	0.88	0.87	0.84
	4		–	1.00	1.00	0.99	0.96
	9			–	1.00	0.99	0.96
	14				–	0.99	0.96
	19					–	0.96
	24						–

Visual Inspection of Network Communities

Some visual representations illustrating the different community assignments are included in this section in Figure 4-1 through Figure 4-4. Since the community assignments for the 8 cases shown in Table 4-5 have a large number of communities but do not exhibit a lot of variation from one another, they are difficult to compare visually. Thus, some cases with stronger differences are selected to be shown here and will be referred to using the reference identification numbers from Table 4-4. Specifically, cases 6, 12, 18 and 24, where $\epsilon_{opt} = \epsilon_{agr} = 0.0001, 0.001, 0.01,$ and 0.1 respectively, are shown in Figure 4-1 through Figure 4-4.

Each of the four figures shows the network communities from the resulting network structure of the Louvain algorithm with the corresponding threshold parameters. In the network represented on the map, each connected set of links represented in the same color form a community. In reading these figures it should be noted that it is likely that the same color has been reused to color different communities within one figure since all of the cases in the figures have at least 60 communities, and links belong in the same community only if they have the same color and are connected. In comparing Figure 4-1 through Figure 4-4, the matching of colors between separate figures should not be considered as the colors are simply associated with cluster numberings.

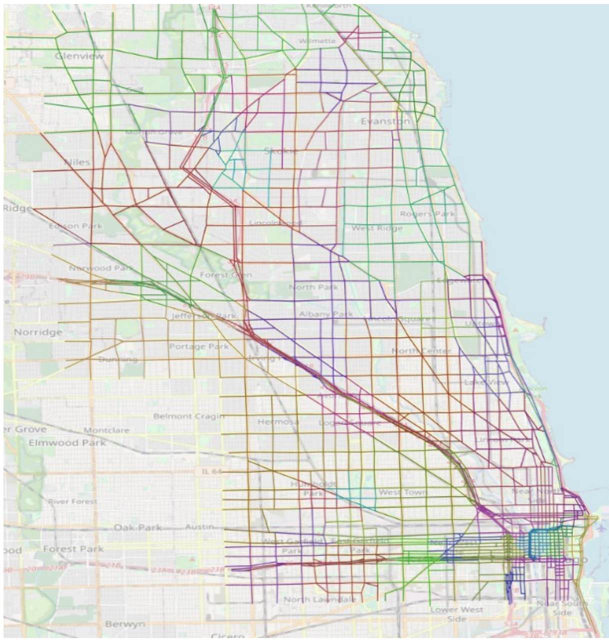

Figure 4-1. Mapped network communities for case ID = 6, $\epsilon_{opt} = \epsilon_{agr} = 0.0001$

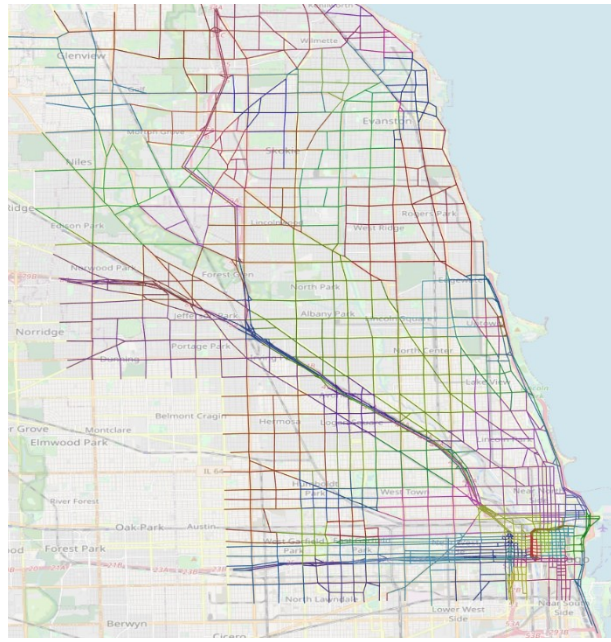

Figure 4-2. Mapped network communities for case ID = 6, $\epsilon_{opt} = \epsilon_{agr} = 0.001$

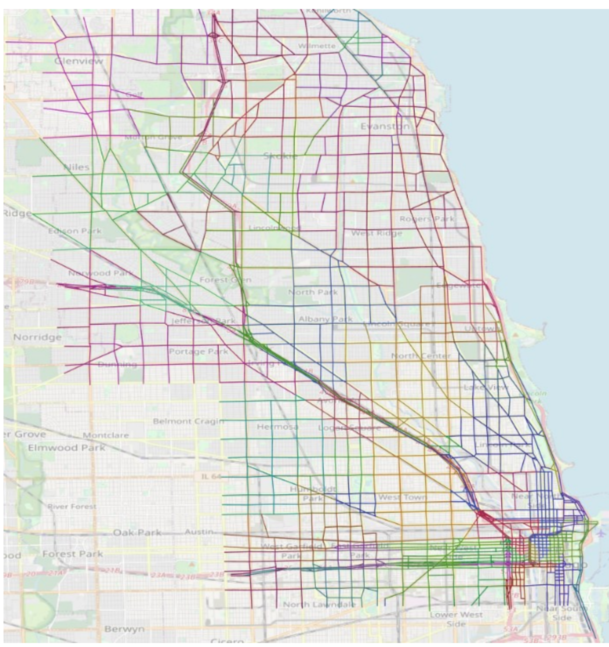

Figure 4-3. Mapped network communities for case ID = 6, $\epsilon_{opt} = \epsilon_{agr} = 0.01$

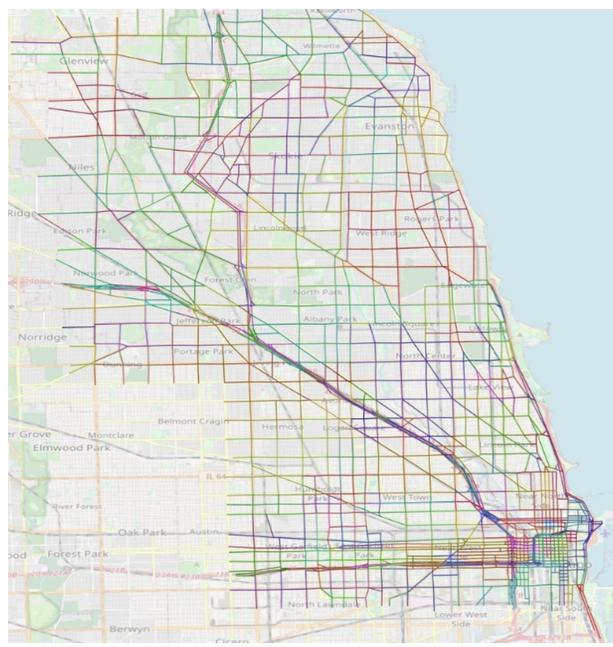

Figure 4-4. Mapped network communities for case ID = 6, $\epsilon_{opt} = \epsilon_{agr} = 0.1$

A few observations can be made from the presented figures. Comparing Figure 4-1, Figure 4-2 and Figure 4-3, it can be seen that many of the same groups of links remain through all three cases, but noting that Figure 4-2 and Figure 4-3 show more segmentation than can be seen in Figure 4-1. However, Figure 4-4 has a significantly larger number of separate communities. Even so, many of the community structures are similar to those observed in the other cases and careful inspection reveals that these are similar to the communities in the other cases, simply broken down further. Quantitative analysis of these results was provided in the previous section and should be used as the primary way for evaluating the sensitivity analysis results.

4.5.2.3 Sensitivity Analysis Results for Temporal Change Point Detection

The results for the sensitivity analysis on the CPD results consists of 40 cases for each of the 25 community assignments, resulting in a total of 1000 cases. This section focuses on comparing the change point results for a single fixed community assignment case where $\epsilon_{opt} = \epsilon_{agr} = 0.0001$, to focus the analysis on the impact of the change-point results. The 40 cases for CPD robustness tests were created by combining 4 different CPD models and 10 values for the penalty parameter by specifying the noise standard deviation. This section presents two sets of results: section 0 presents the characteristics of the change point intervals for the 40 cases, and section 0 presents the change point comparison using the robustness tests.

Change Point Results Characteristics

The characteristics of the change point detection results included in this analysis are the numbers of change points and sizes of time intervals for the travel time data on each of the communities, using the different models and parameter values. Change point detection was performed for each of the communities separately, therefore fixing the parameters for community detection to $\epsilon_{opt} = \epsilon_{agr} = 0.0001$ resulted in 60 communities, meaning that CPD is to be evaluated across 60 multivariate time series for each of the 40 cases. Table 4-7 shows the maximum, mean and minimum number of time bins detected for the 60 multivariate time series for the 40 cases of varying the CPD models and the penalty parameter p are.

Important observations about the models can be made from the results in each of the rows in Table 4-7, showing how the number of change points differ from one model to another.

Table 4-7. Maximum, mean, and minimum number of change-points or time-intervals for the 40 cases varying the model and parameter values.

		CPD Models											
		l_1			l_2			c_{rank}			K_{rbf}		
		max	mean	min	max	mean	min	max	mean	min	max	mean	min
σ values specifying p	0.1	288	132.20	3	288	127.13	2	288	281.28	224	144	16.23	3
	0.2	288	115.22	1	288	120.77	1	288	275.10	195	93	7.65	1
	0.3	288	106.57	1	288	117.67	1	288	266.57	170	74	5.23	1
	0.4	288	101.40	1	288	115.22	1	288	256.12	157	55	4.17	1
	0.5	288	97.22	1	288	113.67	1	288	243.52	146	54	3.85	1
	0.6	288	93.62	1	288	111.97	1	288	228.97	134	45	3.40	1
	0.7	288	90.08	1	288	110.88	1	288	214.52	121	40	3.13	1
	0.8	288	87.43	1	288	109.65	1	288	199.48	112	33	2.95	1
	0.9	288	85.10	1	288	108.48	1	285	185.77	110	31	2.77	1
	1	288	82.98	1	288	107.38	1	281	173.32	106	25	2.42	1

Seeing that the rank-based model, accounting for changes in both the mean and the covariance, has detected a significantly larger number of change points compared to those detected by the other models confirms the hypothesis that there are in fact changes in the covariance of the link travel time random variables and they may not necessarily coincide with the occurrence of changes in the mean. Using the lower order models would detect change points in the mean of multivariate time series, but not the dependence.

Considering the columns in Table 4-7, variations in the number of change points when varying the penalty parameter p for each of the models can be observed. As the value of the penalty increases, linear in σ^2, the mean number of change points detected decreases. This is consistent with the interpretation of the penalty parameter based on σ^2: if a larger value for the underlying noise of the observations is assumed, a larger portion of the potential change points are attributed to variations due to noise and a smaller number are detected as true change points. On the other hand, the minimum and maximum number of change points remain robust for all except the rank model and for all except the largest value of p.

Additional data on the duration of the resulting time intervals or regimes from change-point detection are shown in Table 4-8, presenting the maximum, mean and minimum duration of detected time intervals in minutes. This data shows information corresponding to that in Table 4-7 where the durations of those intervals instead of their number are shown. All cases where 288 change points were detected over the 24-hour period had resulting time intervals of 5 minutes.

Table 4-8. Maximum, mean, and minimum duration [minutes] of time interval or regime for the 40 cases varying the model and parameter values.

		CPD Models											
		l_1			l_2			c_{rank}			K_{rbf}		
		max	mean	min	max	mean	min	max	mean	min	max	mean	min
σ values specifying p	0.1	1035	75.65	5	1040	148.66	5	50	5.14	5	1095	218.48	5
	0.2	1440	138.76	5	1440	297.16	5	70	5.28	5	1440	353.94	5
	0.3	1440	182.58	5	1440	348.68	5	110	5.50	5	1440	492.35	5
	0.4	1440	211.59	5	1440	390.47	5	230	5.80	5	1440	611.48	5
	0.5	1440	219.69	5	1440	415.59	5	270	6.15	5	1440	632.39	5
	0.6	1440	227.78	5	1440	421.03	5	415	6.59	5	1440	693.98	5
	0.7	1440	254.35	5	1440	437.65	5	415	7.08	5	1440	743.40	5
	0.8	1440	278.11	5	1440	470.86	5	420	7.65	5	1440	765.53	10
	0.9	1440	286.43	5	1440	485.22	5	420	8.25	5	1440	823.17	10
	1	1440	331.59	5	1440	489.52	5	490	8.86	5	1440	941.36	10

Focusing on the rank model, which both according to the theory and the data is most sensitive to detecting change points, the time interval durations ranged from 5 to 50 minutes with the mean close to 5 minutes, for the case most sensitive to noise (i.e., lowest σ and p values). For the cases robust to noise, i.e., larger values of p, the durations of time interval increased up to 490 minutes (i.e., 8 hours and 10 minutes) with the mean at 8.86 minutes.

Robustness Measures for Change Point Detection Results

To continue the comparison of the CPD results, this section considers robustness measures making use of clustering similarity measures not much different than those considered in the previous section. The two similarity measures considered here were the Rand Index and the Fowlkes-Mallows Index defined similarly to the Rand score and Fowlkes-Mallows score robustness measures for community clusters. The similarity scores were computed pairwise for each pair of the 40 cases considered for CPD. Since each of the 40 cases included performing CPD for the 60 communities for the sample case of community structure results, the scores were computed as mean values over all links in the network corresponding to the cluster size weighted average of the scores for each community. Since these results contain a total of $(40^2 - 40)/2$ values for each of the similarity measures, only some of the more interesting cases were considered here.

Having seen that the rank-based model had the highest sensitivity and exhibited a lot of variation in terms of the number and duration of time intervals detected, Table 4-9 focuses on the similarity measure values for this model across the 10 values for the penalty parameter. As the values for p are farther apart the similarity between the change point results also decreases. For close and low values of p where $\sigma \in \{0.1, 0.2, 0.3\}$ the RI values are equal to 1 and remain high as the values increase. Only after increasing σ to 0.8 or 0.9 RI values lower than 0.995 and 0.992 are encountered, demonstrating very strong robustness for the CPD results for large ranges of the penalty parameter p, perhaps indicating that assuming large values for σ decreases robustness due to the removal of some change points that are significant, and not necessarily due to noise.

Table 4-9 Average RI and FMI values for pairwise comparison of 10 cases using the rank based CPD model.

Average Rand Index values											
		σ values specifying p									
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
σ values specifying p	**0.1**	–	1.000	1.000	0.999	0.998	0.997	0.995	0.992	0.988	0.984
	0.2		–	1.000	0.999	0.998	0.997	0.995	0.992	0.989	0.984
	0.3			–	0.999	0.999	0.997	0.995	0.992	0.989	0.984
	0.4				–	0.999	0.998	0.996	0.993	0.989	0.985
	0.5					–	0.999	0.997	0.994	0.990	0.986
	0.6						–	0.998	0.995	0.992	0.987
	0.7							–	0.997	0.994	0.989
	0.8								–	0.997	0.992
	0.9									–	0.995
	1.0										–
Average Fowlkes-Mallows Index values											
		σ values specifying p									
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
σ values specifying p	**0.1**	–	0.990	0.984	0.975	0.958	0.935	0.908	0.877	0.850	0.822
	0.2		–	0.986	0.977	0.961	0.938	0.911	0.880	0.852	0.825
	0.3			–	0.979	0.965	0.942	0.916	0.885	0.857	0.829
	0.4				–	0.969	0.948	0.923	0.892	0.865	0.837
	0.5					–	0.953	0.931	0.903	0.877	0.849
	0.6						–	0.940	0.917	0.892	0.867
	0.7							–	0.928	0.909	0.887
	0.8								–	0.924	0.907
	0.9									–	0.922
	1.0										–

Considering the FMI values from Table 4-9, a similar pattern can be observed, where the FMI values are very high for the low values of σ and decrease even more significantly as larger values are considered. In addition, it should be observed that a pairwise comparison of cases with high σ values, such as comparing the cases where $\sigma \in \{0.8, 0.9, 1.0\}$, robustness is also low indicating that these solutions were not robust even compared to relatively similar cases. The results presented here are a small portion of the results obtained from the total cases considered, however the main observed patterns remain the same.

Some conclusions from the observations in these results are that the rank based CPD model may be needed as it does capture significantly more change points by accounting for changes in the covariance matrix, relative to the models that are based on mean or median values only. The CPD results are robust for lower values of σ, the assumed standard deviation of noise in the observations, thus inferring that large σ values can lead to highly different CPD results by omitting potentially significant change points.

4.5.3 Sensitivity Analysis to Exogenous and Endogenous Changes

The second part of the sensitivity analysis was designed to understand and evaluate the robustness of the network characterization with varying network conditions. Specifically, this study considers exogenous changes in the network via varying weather conditions and endogenous changes via varying demand patterns. The network characterization sensitivity analysis was performed on a base case for a 'regular' weekday demand pattern and clear weather and 8 additional cases: 4 cases with different weather and 4 cases with different demand patterns. Thus, the 1000 cases from the sensitivity analysis of modeling parameters were now tested over the 9 days, resulting in a total of 9000 cases. Nevertheless, the goal of this portion of the sensitivity analysis is to understand how a single characterization of the network, with fixed, may change when applied to data from the different weather or demand cases.

The sensitivity analysis results are separated across two parts of this section, considering the effect of changes in weather and demand patterns, respectively. Those sections are each organized like the previous one, considering the sensitivity analysis in terms of summary information and robustness measures for the community structures and change points or intervals. This results section focuses only on the robust cases considering a smaller range for the parameters where $\epsilon_{opt} \in \{0.00001, 0.0001, 0.001\}$ and $\epsilon_{agr} \in \{0.00001, 0.0001\}$ for community detection, and the rank-based model and $\sigma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ for CPD.

4.5.3.1 Sensitivity Analysis Results for Varying Weather Conditions

Sensitivity Analysis of Community Structures

For the five different weather conditions, this analysis considers the changes in some of the characteristics of resulting communities and the robustness measures of community structures.

Table 4-10 shows a summary of the characteristics of the community structures for 6 combinations of values for the tolerance parameters $\epsilon_{opt} \in \{0.00001, 0.0001, 0.001\}$ and $\epsilon_{agr} \in \{0.00001, 0.0001\}$ over the 5 days with varying weather conditions (including the base clear weather case). The number and mean size of the resulting communities are stable across the different days, but the consistency among clustering assignment will be further evaluated through the robustness measures.

The cluster robustness measure considered here is the normalized mutual information (NMI) score and Table 4-11 shows the NMI values, pairwise, for the 5 weather cases, with the threshold parameters fixed to $\epsilon_{opt} = \epsilon_{agr} = 0.0001$.

Table 4-10. Number of communities and mean community size for 6 combinations of ϵ_{opt} and ϵ_{agr} values, across 5 days with varying weather conditions

$\epsilon_{agr} =$		$\mathbf{0.00001}$		$\mathbf{0.0001}$	
		Number	**Mean size**	**Number**	**Mean size**
Weather Case IDs		$\epsilon_{opt} = \mathbf{0.00001}$			
	1	58	82.84	58	82.84
	2	56	85.80	56	85.80
	3	56	85.80	56	85.80
	4	55	87.36	55	87.36
	5	55	87.36	55	87.36
		$\epsilon_{opt} = \mathbf{0.0001}$			
	1	58	82.84	58	82.84
	2	56	85.80	56	85.80
	3	56	85.80	56	85.80
	4	55	87.36	55	87.36
	5	55	87.36	55	87.36
		$\epsilon_{opt} = \mathbf{0.001}$			
	1	58	82.84	58	82.84
	2	56	85.80	56	85.80
	3	56	85.80	56	85.80
	4	55	87.36	55	87.36
	5	55	87.36	55	87.36

Table 4-11. NMI for the 5 different weather cases for a fixed penalty parameter value

		Weather Case IDs				
		1	**2**	**3**	**4**	**5**
Weather Case IDs	**1**	–	0.877	0.874	0.877	0.852
	2		–	0.889	0.883	0.896
	3			–	0.884	0.894
	4				-	0.874
	5					–

The NMI values show statistical robustness as all values are well above 0.5 on the 0 to 1 scale. However, relative to the NMI values obtained for the sensitivity analysis across parameters in section 4.5.2.2, where the NMI scores were all greater than or equal to 0.99, robustness is notably lower. This indicates that, while there is significant consistency in the community structures, weather conditions do have a notable effect on the network community structures.

Sensitivity Analysis of Change Point Detection Results

For the five different weather conditions, this analysis considers the changes in some of the characteristics of resulting change points and the robustness measures of CPD results. Again, the values are considered for a single case with $\epsilon_{opt} = \epsilon_{agr} = 0.0001$ and considering the robust penalty parameter values where $\sigma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

First, the number of change points and the duration of resulting time intervals, averaged across all network communities were considered for the 6 parameter values and over the 5 different weather cases. The results shown in Table 4-12 show that as the value for p changes there is some change in the characteristics of the change point results, however the results are more stable when considered for any given value of p, the characteristics remain relatively stable across the different days.

Table 4-12. Mean number of change points (n) and interval durations (d) in minutes for 6 values of the penalty parameter p across 5 weather cases.

Case ID		1		2		3		4		5	
		n	d	n	d	n	d	n	d	n	d
σ values specifying p	0.1	281.12	5.14	282.34	5.12	277.11	5.21	281.15	5.13	280.00	5.17
	0.2	274.72	5.29	276.00	5.26	269.98	5.36	275.91	5.24	271.40	5.37
	0.3	266.97	5.50	268.48	5.43	263.91	5.49	270.15	5.37	260.35	5.65
	0.4	257.02	5.77	259.09	5.66	255.68	5.69	262.78	5.54	247.24	6.01
	0.5	244.10	6.15	247.07	5.99	246.55	5.93	254.40	5.75	232.85	6.44
	0.6	230.33	6.55	234.34	6.35	237.25	6.19	243.87	6.04	218.85	6.90

To assess the robustness of the change point results, the similarity measures used in the previous sections were considered: Rand Index and Fowlkes-Mallows index. To perform the comparison, the values for the penalty parameter was fixed by setting $\sigma = 0.1$ and the results are shown in Table 4-13. These results show very high values for both the RI and FMI scores, even compared to the cases where simply the parameters were varied, showing that in terms of the temporal change point results, the characterization was very robust with respect to varying weather conditions encouraging the possibility that in fact network characterizations may not need to be re-defined and fitted separately across different weather operational conditions.

Table 4-13. RI and FMI values for a single community and CPD case, across 5 weather cases

RI						
		Weather Case IDs				
		1	2	3	4	5
Weather Case IDs	1	–	0.9999	0.9995	0.9998	0.9999
	2		–	0.9995	0.9998	0.9999
	3			–	0.9995	0.9996
	4				–	0.9998
	5					–
FMI						
		Weather Case IDs				
		1	2	3	4	5
Weather Case IDs	1	–	0.9918	0.9871	0.9912	0.9917
	2		–	0.9868	0.9910	0.9916
	3			–	0.9846	0.9861
	4				–	0.9898
	5					–

4.5.3.2 Sensitivity Analysis Results for Varying Demand Patterns

Analogous to the previous section, here the five different demand patterns are considered to evaluate the changes in the spatial and temporal characterization of the network.

<u>*Sensitivity Analysis of Community Structures*</u>

The analysis in this considers the changes in some of the characteristics of resulting communities and the robustness measures of community structures with the changing demand patterns.

Table 4-14. Number of communities and mean community size for 6 combinations of ϵ_{opt} and ϵ_{agr} values, across 5 days with varying demand patterns

$\epsilon_{agr} =$		0.00001		0.0001	
		Number	Mean size	Number	Mean size
Demand Case IDs		$\epsilon_{opt} = 0.00001$			
	1	58	82.84	58	82.84
	2	63	76.27	63	76.27
	3	66	72.80	66	72.80
	4	74	64.93	74	64.93
	5	67	71.72	67	71.72
		$\epsilon_{opt} = 0.0001$			
	1	58	82.84	58	82.84
	2	63	76.27	63	76.27
	3	66	72.80	66	72.80
	4	74	64.93	74	64.93
	5	67	71.72	67	71.72
		$\epsilon_{opt} = 0.001$			
	1	58	82.84	58	82.84
	2	63	76.27	63	76.27
	3	66	72.80	67	71.72
	4	75	64.07	75	64.07
	5	69	69.64	69	69.64

Table 4-14 shows a summary of the characteristics of the community structures for 6 combinations of values for the tolerance parameters $\epsilon_{opt} \in \{0.00001, 0.0001, 0.001\}$ and $\epsilon_{agr} \in \{0.00001, 0.0001\}$ over the 5 days with varying demand patterns. It can be observed that the number and mean size of the resulting can somewhat vary with the changing demand conditions. The consistency among clustering assignment can be evaluated through the robustness measures.

The cluster robustness measure considered here is again the normalized mutual information (NMI) score and Table 4-15 shows the NMI values, pairwise, for the 5 demand cases considered, where the threshold parameters were fixed to $\epsilon_{opt} = \epsilon_{agr} = 0.0001$.

Table 4-15. NMI for the 5 different demand cases for a fixed penalty parameter value

		Demand Case IDs				
		1	2	3	4	5
Demand Case IDs	1	–	0.875	0.872	0.856	0.852
	2		–	0.905	0.882	0.889
	3			–	0.892	0.892
	4				–	0.891
	5					–

The values for NMI comparing the clustering assignments across different demand cases show some level of robustness statistically as all values are well above 0.5 on the 0 to 1 scale. Again, relative to the NMI values obtained for the sensitivity analysis across parameters, where the NMI scores were all greater than or equal to 0.99, robustness has been reduced. This indicates that while there is significant consistency in the community structures, demand conditions can have some effect on the network community structures. In comparison to the results observed when considering different weather cases, however, the robustness is slightly stronger in this case.

Sensitivity Analysis of Change Point Detection Results

Similar to the corresponding section for the weather cases, for the five different demand patterns, the changes in some of the characteristics of resulting change points and the robustness measures of change point detection results are considered. Here, the values are considered for a single case again, setting $\epsilon_{opt} = \epsilon_{agr} = 0.0001$ and considering the robust penalty parameter values where $\sigma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

First, the number of change points detected and the duration of resulting time intervals, averaged across all network communities were considered for the 6 parameter values and over the 5 different demand cases. The results shown in Table 4-16 show that as the value for p changes there is some change in the characteristics of the change point results, however when considered for any given value of p as the characteristics remain relatively stable across the different demand cases, similarly to what was observed for the weather case analysis.

Table 4-16. Mean number of change points (n) and interval durations in minutes (d) for 6 values of the penalty parameter p across 5 demand cases

Case ID		1		2		3		4		5	
		n	d	n	d	n	d	n	d	n	d
σ values specifying p	**0.1**	281.12	5.14	278.49	5.22	275.30	5.35	277.31	5.30	276.48	5.28
	0.2	274.72	5.29	272.21	5.36	269.71	5.49	270.85	5.46	270.30	5.42
	0.3	266.97	5.50	264.22	5.57	262.56	5.69	263.84	5.64	261.85	5.64
	0.4	257.02	5.77	253.00	5.88	253.09	5.96	255.00	5.89	252.07	5.90
	0.5	244.10	6.15	239.22	6.28	241.58	6.32	243.43	6.22	240.31	6.23
	0.6	230.33	6.55	226.22	6.68	227.64	6.77	228.55	6.70	225.87	6.68

To assess the robustness of the change point results, the Rand Index and Fowlkes-Mallows index are considered again, and also the value for the penalty parameter was fixed by setting $\sigma =$

0.1 and the results are shown in Table 4-17. Again, these results show very high values for both the RI and FMI scores, similarly to the results seen when comparing the weather cases.

Therefore, in terms of the temporal change point results, the characterization was very robust with respect to varying demand conditions again. Through these results and those seen for the sensitivity analysis on the weather cases it can be concluded that the temporal characterizations of the network are very robust, while the spatial characterizations may show some non-negligible variations but can be considered statistically robust.

Table 4-17. RI and FMI values for a single community and CPD case, across 5 weather cases

RI						
		Demand Case IDs				
		1	2	3	4	5
Demand Case IDs	1	–	0.9999	0.9997	0.9998	0.9998
	2		–	0.9997	0.9998	0.9998
	3			–	0.9997	0.9997
	4				–	0.9998
	5					–
FMI						
		Demand Case IDs				
		1	2	3	4	5
Demand Case IDs	1	–	0.9925	0.9923	0.9919	0.9915
	2		–	0.9922	0.9917	0.9917
	3			–	0.9915	0.9917
	4				–	0.9913
	5					–

4.6 Conclusion

This chapter is focused on the problem of characterizing stochastic dynamic networks and their underlying spatial and temporal dependencies. A systematic taxonomy for stochastic dynamic network characterization is presented, featuring the main dimensions in defining a stochastic dynamic network: temporal variation of network characteristics, temporal and spatial dependencies and their existence, strength, range, and variation, and stationary vs. non-stationary characteristics of the network.

The approaches considered for stochastic dynamic network characterization are drawn from network science, time series analysis and data mining aimed at developing data-driven models of stochastic transportation networks. The presented methodology consists of two stages: (1) community detection using network clustering approaches with similarity measures derived from dynamic time warping, and (2) temporal change point detection across the joint travel time distributions for each of the resulting communities. The network communities are interpreted as neighborhoods where link travel time dependencies are considered high within each community, but negligible between them. The change point detection results consist of time intervals where joint link travel time distributions are stationary and change points where the underlying distribution changes, for each community of links.

A sensitivity analysis was performed to test for the robustness of the network characterization across two main dimensions. Firstly, the modeling parameters were varied to evaluate ranges or values for each of the parameters where the network characterization was robust or better specified. Secondly, for fixed parameter values a sensitivity analysis was performed using different network data resulting from varying weather and demand conditions in the network to understand and test the robustness of the network characterization with respect to exogenous and endogenous factors. Overall, the network characterization was shown to be robust, especially in terms of the temporal characterization. The spatial characterization shows some practical variation with changes in the weather and demand cases but was also demonstrated to be statistically robust.

This study opens a number of additional questions and considerations that have the potential to serve as basis for future work. Methodologically, sophisticated algorithms for community structures and change point detection can be applied. Different community finding approaches can be considered, such as algorithms for finding overlapping communities. Such an extension would also require a change in the temporal characterization approaches as well. Furthermore, change point detection can be performed on the community structure itself, determining the points in time when the community structure changes significantly and thus capturing temporal variation in the existence of link travel time dependencies.

Chapter 5 Estimation of Path Travel Time Distributions in Stochastic Dynamic Networks with Correlations[1]

5.1 Overview

This chapter focuses on the problem of estimating path travel time distributions in stochastic dynamic networks with capture generalized dependencies between link travel times in the network. It presents a systematic taxonomy of problem situations related to the main dimensions in modeling stochastic travel times in the network and estimating path travel time distributions. A comprehensive methodology of sampling-based convolution estimation algorithms is presented, paired with data analysis approaches, considering the factors that define the taxonomy. The chapter also includes extensive numerical experiments with a complete experimental design considering the computational effort and accuracy of the various methods.

5.2 Taxonomy for Path Travel Time Distribution Estimation

The estimation of path travel time distributions encapsulates a range of problems that can be very diverse from a number of perspectives. The types of available data, the assumptions made about the network characteristics, the type of results the estimation aims to achieve, and the characteristics of relevant applications, are all important considerations that will shape the characteristics of the problem and solution approaches. This taxonomy aims to organize relevant information to help define the types of problem that may arise along with the relevant solution approaches.

---

[1] Portions of this chapter parallel sections of (Filipovska and Mahmassani, 2020b, 2020c, 2021)

The literature on travel time reliability modeling and the estimation of travel time distributions in transportation networks considers several different aggregation levels and definitions of variability. Travel time distributions have been estimated at the network, origin-destination (O-D), path and link level. This chapter is specifically concerned with the estimation of travel time distributions at the path level.

Table 5-1. Taxonomy of path travel time distribution estimation problems

Assumptions	Input Factors	Output
Underlying Network • Graph / Virtual network • Test road network • Real road network Network Congestion and Travel Times Temporal Aspect • Static • Time-dependent Spatial Dependence • Independent • Partial stationary dependence • Full stationary dependence • Generalized dependence Sources of Variability • Time • Driver behavior • External conditions • Mixed	Data Source / Type • Vehicle trajectory data • Detector-based data • Both data types Data Quality • Full observations • Missing data points Data segmentation • Link level • Fixed sub-path level • General sub-path level Supplemental Information • Time information • Vehicle information • Date information • Exogenous factors	Estimated Distribution Form • Parametric • Non-parametric Path-Finding Application • None • A priori • Adaptive Distribution Timeframe • General • Single departure time • Departure period Distribution Application • General • Case-specific • User-specific

Travel time variability in the literature has been defined from different perspectives, including day-to-day, within-day (or time-of-day) and vehicle-to-vehicle variability. This study further classifies the categorizations for formulating problems and solutions for the estimation of path travel time distributions. The taxonomy considers three types of factors: underlying assumptions, input side factors, and output type.

The three taxonomic categories are interrelated. For example, a problem definition based on certain underlying assumptions may be incompatible with some input categories, or a problem definition based on output requirements may also be restricted by some of the underlying assumptions.

5.2.1 Taxonomic Factors Related to Underlying Assumptions

This section presents factors of the taxonomy of travel time distribution estination problems that are related to the underlying assumptions for the problem and the environment.

5.2.1.1 Underlying Network

Three levels are considered for this factor. A *graph or virtual network* is an abstraction of a real network, where the links and nodes represent the roads and their intersections, respectively. *Test networks*, while not as abstract as graphs, are simplistic networks designed for the purpose of analysis to have some specified characteristics. *Real road networks* model the characteristics and structure of physical real-world road networks.

5.2.1.2 Network Congestion and Travel Times

In modeling transportation networks, the network links may include congestion which can be modeled differently based on temporal and spatial dependence assumptions. Networks are generally categorized as deterministic vs. stochastic; this study only considers stochastic networks.

From the temporal aspect, static networks are ones where congestion remains constant through the analysis period and thus travel time distributions are also constant. Time-dependent networks are ones where congestion varies based on the time of day and thus the travel time distributions vary with time. In stochastic networks, this category is designated as stochastic time-varying (STV) networks or stochastic time-dependent (STD) networks. Mixed network models also exist, where typically a peak period is modeled with time dependence, and the network is considered static outside the peak period. This thesis, in accordance with Chapter 4 views a mixed network as a special case of a time-varying network where the time interval changes based on the time of day.

In terms of spatial dependence, travel times across the transportation network can be modeled as independent or dependent. This taxonomy further categorizes the network based on the type of dependence. Stationary dependence assumes fixed dependence, through time or other varying conditions. Partial stationary dependence assumes that dependence exists but is restricted spatially, often to adjacent or neighboring links. Full dependence assumes that dependence exists across all link travel times in the entire network. Again, partial dependence can be seen as a special case of full dependence with certain quantities set to zero.

Generalized dependence is introduced as a new category (factor level) that aims to bring together extensions to current dependence models. Generalized dependence is a non-stationary full dependence (where partial dependence neighborhoods can be modeled as a special case) that varies over time, possibly with the same time variation structure as the link travel time variables.

5.2.1.3 Sources of Variability

In modeling travel time variability, assumptions about sources variability can significantly impact the type of relevant models. Time as the source of variability can capture day-to-day or within-day variation. Driver behavior as the source of variability considers vehicle-to-vehicle variability (thus requiring vehicle trajectory data). External conditions as the source of variability can be considered explicitly via mixture models using the likelihood of occurrence for a certain scenario, or implicitly by extracting the variation from the data without having information of the external factors may have caused it.

5.2.2 Taxonomic Factors Related to Input Characteristics

This section presents factors of the taxonomy that are related to the the input side of the problem, primarily relating to the available data.

5.2.2.1 Data Source or Type

The data sources or types can include Lagrangian (mobile) or Laplacian (stationary) measurements, typically as vehicle trajectory data or detector-based data, respectively. Vehicle trajectory data are a type of mobile data, which can be obtained from a range of different sources, where observations on actual vehicles' traversed trajectories are available typically as time and location observations with a certain frequency. Travel time information can be extracted or estimated by mapping vehicle trajectory data onto a network. Detector-based data are a type of stationary data typically associated with a fixed location that provide observations on the occupancy and speed of vehicles traversing the detector location, aggregated over short periods of time. The type of available data is important in defining and approaching the problem of path travel time estimation. If both data types are available, the analyst may choose to define the problem so as to utilize both data types.

5.2.2.2 Data Quality

The quality of the available input data, whether full observations are available or there are missing data points, plays an important role in defining the goal of estimation. When full data is available, the path travel time estimation is primarily formulated as a mathematical problem of probability modeling. Missing data points would require the formulation of a data mining component for the imputation of missing information.

5.2.2.3 Data Segmentation

The segmentation of the data is classified as: link-level segmentation, fixed (sub-)path segmentation, and general (sub-)path segmentation. Data segmented at the link level allow for travel time information to be obtained for individual network links only. Typically, detector-based data is segmented at the link level, but vehicle trajectory data, if fully anonymized and disaggregated can also be segmented at the link level. Fixed (sub-)path segmentation allows for travel time data to be obtained aggregated at the level of specific network paths or sub-paths. Aggregated vehicle trip data can come at this level of segmentation if it contains information on the traversed path and travel time for each trip, but not the travel times (or time of arrival) for any intermediate point en-route. General (sub-)path segmentation is data that can be segmented at the link level but also provide travel times for any covered (sub-)path and can be extracted from disaggregated vehicle trajectory data.

5.2.2.4 Supplemental Information

Different types of supplemental information may be available in the input data. The categories included here are those most relevant to the problem at hand and are not exhaustive or mutually exclusive. Date and time information provide the exact date and time of day for each observation, respectively. These supplemental information types are necessary for modeling day-to-day and within day variation, respectively. Vehicle information, relevant only to vehicle trajectory data types, includes identification of the vehicle producing each observation, necessary for modeling vehicle-to-vehicle variation. Information on exogeneous factors is a more general category that can include information on weather, special events in the network or any other exogenous factors relevant to the model and would be necessary if the analyst or researcher intends to model event or scenario-based variation.

5.2.3 Taxonomic Factors Related to the Output Requirements

This section presents taxonomic factors related to the output requirements for the problem at hand.

5.2.3.1 Estimated Distribution Form

The estimated distribution can have parametric or non-parametric form. Parametric distribution output is obtained when an assumption is made on the functional form of the distribution of the underlying data and thus the estimation is reduced to estimating the specific parameters of that function to obtain a good fit to the data. Non-parametric distribution output makes no assumptions about the functional form of the distribution, they typically take a collection of samples and estimate an empirical mass function that summarizes the data for which key characteristics of the distribution can then be estimated.

5.2.3.2 Path-Finding Application

The path travel time distribution estimation can be performed for the purpose of being utilized for path finding applications or not. A priori path finding applications determine the path for a given origin-destination pair and departure time(s). Adaptive path finding applications determine a collection of paths for a given origin-destination pair and departure time(s), so that the selected path is based on revealed information.

5.2.3.3 Distribution Timeframe

A path travel time distribution can be estimated without a specified departure time, for a single departure time, or for a departure period. Each of the different timeframe categories would require different ways of accounting for within-day or day-to-day variation. A distribution for a departure period will require that the possible departure times within a given period be modeled and the resulting distribution to account for any within-day variation.

5.2.3.4 Distribution Application

The application of the path travel time distribution can be general, when simply a good distribution of the travel times along the network is needed, or case- or user-specific. Case-specific distributions can be based on exogenous or endogenous factors, such as, for example, the estimation of travel time distribution under specified weather conditions. User-specific distributions may need to account for the behavior of the specific driver or vehicle and potentially omit vehicle-to-vehicle variation.

5.3 Problem Definition

The primary focus of this chapter is the problem of efficient and accurate estimation of travel time distributions on user-specified paths across stochastic dynamic networks with generalized spatio-temporal travel time dependencies. Stochasticity of travel times in the transportation network can be due to a number of factors and capturing generalized spatio-temporal dependencies between link travel times relies on the availability of trajectory data. Therefore, assuming the availability of trajectory data, the methodological challenge of estimating path travel time distributions on any given route in the network is two-fold.

The first methodological challenge is one of probability modeling, where given the link travel time distributions the goal is to estimate the path travel time distribution, i.e., given the distributions of individual random variables, the goal is to estimate the distribution of their sum. The distribution of a sum of random variables is represented with a convolution integral, thus the fundamental problem is to formulate and solve a convoluting integral where random variables can be correlated and have varying distribution forms.

Suppose a path P is composed of K consecutive segments $\{l_1, l_2, \ldots, l_K\}$ where travel time on segment $l_i \; \forall \; i \in \{1, 2, \ldots, K\}$ is a random variable θ_i whose distribution is denoted by π_i. In this notation the segments comprising path P can be sub-paths of varying lengths, where a special case would be to consider the individual links comprising the path. The distribution of travel times Θ_P on the path P is obtained by solving the appropriate convolution integral. In the simplest case, assuming independent random variables, the convolution integral for two random variables (that is, $K = 2$) can be written as:

$$\left(\pi_i * \pi_j\right)(T) = \int_{-\infty}^{\infty} \pi_i(\tau)\pi_j(T - \tau)d\tau$$

If the random variables have time-varying distributions, so that π_i is different at each time t, the convolution integral at time t becomes:

$$\left(\pi_i * \pi_j\right)^t(T) = \int_{-\infty}^{\infty} \pi_i^t(\tau)\pi_j^{t+\tau}(T - \tau)d\tau$$

For the case of travel time random variables, π_i are supported only on $[0, \infty)$ in general, or some specific range $[t_{min}, t_{max}]$, and the integral can be appropriately truncated.

Dependencies between segment travel times can be specified depending on the application. Distributions may vary based on exogeneous conditions, such as weather or seasonal attributes, or different operational states that may be defined based on day of the week or demand level. Without loss of specificity, let ζ be an operational condition indicator, and suppose that operational conditions may be different, yet dependent, on different segments, then the integral is further specified by:

$$\left(\pi_i * \pi_j\right)^t(T) = \int_{-\infty}^{\infty} \pi_i^{\zeta_i,t}(\tau)\pi_j^{\zeta_j|\zeta_i,t+\tau}(T-\tau)d\tau$$

To generalize even further, suppose $\theta_i^{t_i}, \theta_j^{t_j}$ are jointly distributed in a time-varying manner so that that the joint distribution will vary over the combination of times t_i, t_j and a nonstationary correlation can be estimated for θ_i, θ_j over all combination of times t_i, t_j.

The second methodological challenge in this problem is concerned with identifying appropriate segments that may be used to best synthesize the travel time distribution of a path, and as such is a challenge of data mining. Identifying the appropriate segments depends on available trajectory data and the travel patterns captured in it, as well as the spatio-temporal dependencies in the network. Using link-level travel time distributions is one special-case solution to this methodological challenge. However, if the data allows, in some cases it may be useful to consider using sub-paths of the path in question, each comprised of one or more links, that are stitched together to estimate the path travel time distribution. The benefit of this alternative may be that some of the correlations between link travel times can be implicitly captured within known sub-path travel time distributions and thus improve both the computational requirement and accuracy of the resulting estimation.

5.4 Methodology

This section presents a few groups of approaches for the estimation of path travel time distributions, that can be classified according to the output requirement categories previously introduced in the taxonomy summarized in Table 5-1. To provide a comprehensive set of methods, separate approaches or modifications are presented for each method type for their application to a different category of output requirements.

An overview of the path travel time estimation methods is presented in Table 5-2, grouping the approaches into four types and categorizing them according to each of the five classifications based on the output requirements. Four types of estimation approaches are presented: Monte-Carlo Simulation (MCS) based approaches, Metropolis-Hastings Simulation (MHS) based methods, Normal to anything (NORTA) distribution and Lognormal distribution approximation methods. This methodology expands on the initial work and approaches by Filipovska et al. (2021).

Table 5-2. Estimation approaches, types, and categories according to the taxonomy

<table>
<tr><th></th><th></th><th>Distribution Form</th><th>Path-Finding Application</th><th>Distribution Timeframe</th><th>Distribution Application</th></tr>
<tr><td rowspan="4">MCS based</td><td>MCS-I</td><td rowspan="4">Non-parametric</td><td>A Priori</td><td rowspan="4">Adjustable for each type</td><td rowspan="4">Adjustable for each type</td></tr>
<tr><td>MCS-TD-I</td><td>A Priori or Time-Adaptive</td></tr>
<tr><td>MCS-TD-S</td><td>A Priori or Adaptive</td></tr>
<tr><td>MCS-TD-ST</td><td>A Priori or Adaptive</td></tr>
<tr><td rowspan="2">MHS based</td><td>MHS-S</td><td rowspan="2">Non-parametric</td><td rowspan="2">A Priori or Adaptive</td><td rowspan="2">Adjustable for each type</td><td rowspan="2">Adjustable for each type</td></tr>
<tr><td>MHS-TD</td></tr>
<tr><td rowspan="3">NORTA</td><td>NORTA-S</td><td rowspan="3">Parametric</td><td rowspan="3">A Priori or Adaptive</td><td rowspan="3">Adjustable for each type</td><td rowspan="3">Adjustable for each type</td></tr>
<tr><td>NORTA-TD</td></tr>
<tr><td>NORTA-TD-ST</td></tr>
<tr><td rowspan="2">Lognormal approximation</td><td>Lognormal</td><td rowspan="2">Parametric</td><td rowspan="2">Adjustable for each type</td><td rowspan="2">Adjustable for each type</td><td rowspan="2">Adjustable for each type</td></tr>
<tr><td>Lognormal TD</td></tr>
</table>

In the following subsections, suppose a path P is composed of K consecutive segments $\{l_1, l_2, \ldots, l_K\}$ where travel time on segment l_K $\forall\, i \in \{1, 2, \ldots, K\}$ is a random variable θ_i whose distribution is denoted by π_i. In reading this section, one can assume the simplest case where $\{l_1, l_2, \ldots, l_K\}$ is the set of K links comprising the path P with their respective link travel time random variables θ_i and distributions $\pi_i \; \forall \; i \in \{1, 2, \ldots, K\}$. However, the methods are developed so that, either based on the input data structure or the analyst's desires, they can be applied to a different type of segmentation of the path. Let Θ_P be the travel time random variable for path P, with a corresponding cumulative distribution function (CDF) U_P such that for a given travel time T, $U_P(T) = P(\Theta_P \leq T)$.

5.4.1 Monte-Carlo Simulation (MCS) Based Estimation Approaches

Four variations of the Monte-Carlos Simulation (MCS) based approach are presented in this section, starting with an approach assuming independent random variables and then relaxing the assumptions in each sub-sequent modification.

5.4.1.1 MCS-based Approach Assuming Static and Independent Random Variables (MCS-I)

Under the assumption that travel time distributions are time-invariant, then U_P can be written with the following convolution integral:

$$U_P(T) = \int_0^\infty \cdots \left(\int_0^\infty \left(\int_0^\infty \pi_1(\tau_1) \cdot \pi_2(\tau_2 - \tau_1) \, d\tau_1 \right) \cdot \pi_3(\tau_3 - \tau_1 - \tau_2) d\tau_2 \right) \cdots \pi_N \left(T - \sum_{j=1}^{N-1} \tau_j \right) d\tau_{N-1}$$

The MCS approach for static and independent segment travel times, MCS-SI, is as follows.

- Sample link travel time observations $\tilde{\theta}_i^j$ according to $\pi_i \; \forall \; i \in \{1, 2, \ldots, K\}$ for iterations $j \in \{1, \ldots, M\}$
- Compute the path travel time samples $\widetilde{\Theta}_P^j = \sum_{i=1}^N \tilde{\theta}_i^j$ for iterations $j \in \{1, \ldots, M\}$

Then the samples $\widetilde{\Theta}_P^j \; \forall \; j \in \{1, \ldots, M\}$ form a simulated estimate the distribution of Θ_P.

5.4.1.2 MCS-based Approach Assuming Dynamic (Time-Varying) and Independent Random Variables (MCS-TD-I)

In this modification, link travel time distributions are still independent random variables, but unlike the original case above, they are now dynamic or time varying. To apply the MCS approach, a planning horizon of length T_P is considered for the departure times and the link travel time distributions are considered across a set of time intervals $\mathcal{T}$. Segment travel times then have a stationary distribution within each time-interval, given as $\pi_i^t \; \forall \; i \in \{1, \dots, K\}, t \in \mathcal{T}$. A few important notes should be added when considering a distribution across a given horizon.

- A travel time distribution across the planning horizon T_P would require data for a longer time duration (captured by $\mathcal{T}$) so as to cover the departure times across downstream sections of the path.
- To determine the path travel time distribution across the planning horizon T_P, sampling across the possible departure times is needed.
- If T_P is the departure time period, the discretization into time-intervals equal to those in $\mathcal{T}$ might not be appropriate. For generality, suppose that departure times are a random variable ϑ distributed according to ρ with a continuous support over the planning horizon of duration T_P.
- The choice of the distribution of departure times across the planning horizon T_P can vary, especially based on the distribution application. A user-specific distribution would account for the likelihood of a users' departure across the given horizon, which may be a uniform distribution. A general or network-based distribution would be based on the distribution of departure times seen in the data, which would be an empirical distribution.

- The durations of time intervals in $\mathcal{T}$ can have varying durations, depending on how the changes in travel time distributions are modeled. For generality, let $\phi(t) = \mathcal{t} \in \mathcal{T}$ be a function that converts a time $t \in T_P$ into the corresponding time interval in $\mathcal{t} \in \mathcal{T}$.

The MCS approach is extended to its time-dependent modification, MCS-TD-I, as follows:

For iterations $j \in \{1, \dots, M\}$:

- Sample a starting time $\tilde{\vartheta}_j$ is from the distribution ρ and find the corresponding time interval $\mathcal{t}_j = \phi(\tilde{\vartheta}_j)$.
- Sample a travel time $\tilde{\theta}_1$ from $\pi_1^{t_1}$ where $t_1 = \mathcal{t}_j$ for the first segment.
- For each next segment $i \in \{2, \dots, K\}$, sample a travel time $\tilde{\theta}_i$ from $\pi_i^{t_i}$ with $t_i = \phi\left(\tilde{\vartheta}_j + \sum_{l=0}^{i} \tilde{\theta}_l\right)$ i.e., the time interval corresponding to the departure time on segment i.
- The sampled path travel time is then set to $\tilde{\Theta}_P^j = \sum_{i=1}^{K} \tilde{\theta}_i$.

5.4.1.3 MCS-based Approach Assuming Time-Dependence and Spatially Correlated Random Variables (MCS-TD-S)

Assuming that link travel times have time-varying distributions and are spatially correlated with one-another, the sampling approach needs to be adjusted so that segment travel times are sampled with correlation.

When considering stochastic dynamic networks, the estimation of dependencies between link travel times is an added source of difficulty. This application assumes that dependence is accounted for given the correlations and marginal travel time distributions, with the ability to sample conditionally from the marginal distributions. This variation of the MCS approach, abbreviated MCS-TD-S, assumes stationary correlations between travel times on different segments on the path. Namely, even though $\theta_i^{t_i}$ and $\theta_i^{t'_i}$ for some segment $i \in \{1, \dots, K\}$ and time periods $t_i, t_i' \in \mathcal{T}$ have different distributions, the covariance between $\theta_i^{t_i}$ and $\theta_j^{t_j}$ is defined so that

$$cov\left(\theta_i^{t_i}, \theta_j^{t_j}\right) = cov\left(\theta_i^{t_i'}, \theta_j^{t_j'}\right) \ \forall\, i, j \in \{1, \dots, K\}, i \neq j, \ \forall\, t_i, t_i', t_j, t_j' \in \mathcal{T}.$$

A number of techniques exist for sampling correlated random variables, but they may not be directly applicable to the case with time-varying distributions. An extension to the MCS-TD approach is presented here, abbreviated as MCS-TD-S to be used with stationary covariance structures.

Step 1: Estimate the covariance:

- For segments $i, j \in \{1, \dots, K\}, i \neq j$, select a sample of size S from the set of trajectories traversing both i and j. These are the joint samples for the segment travel times on i, j.
- Estimate the covariance between all joint link travel times on the pair (i, j).

Step 2: Obtain the initial sample:

For iterations $j \in \{1, \dots, M\}$:

- Sample a departure time $\tilde{\vartheta}_j$ is from the distribution ρ and find the corresponding time interval $\mathcal{t}_j = \phi(\tilde{\vartheta}_j)$.
- Sample a travel time $\tilde{\theta}_1$ for the first segment from $\pi_1^{t_1}$ where $t_1 = \mathcal{t}_j$
- For each next segment, $i \in \{2, \dots, K\}$, sample a travel time $\tilde{\theta}_i$ from $\pi_i^{t_i}$ where $t_i = \phi\left(\tilde{\vartheta}_j + \sum_{l=0}^{i} \tilde{\theta}_l\right)$.
- Save the vector of samples $\tilde{\theta}^j = \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_K\}$.

Step 3: Obtain a correlated sample:

For each sample $j \in \{1, \dots, M\}$:

- Multiply the $\tilde{\theta}^j = \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_K\}$ by the covariance Σ to obtain a new $\tilde{\theta}^{j'}$.

- Then set $\widetilde{\Theta}_P^j = \sum_{i=1}^{K} \tilde{\theta}_i^{j'}$.

This method for sampling correlated random variables is based on an approximate approach proposed by Lurie and Goldberg (1998), but adjusted so as to allow for time-varying random variables with a stationary covariance structure.

5.4.1.4 MCS-based Approach Assuming Time-Dependence and Spatio-Temporally Correlated Random Variables (MCS-TD-ST)

This section modifies the previously presented approach, to be used under the assumption that travel time covariances vary over time, along with the distributions themselves. Thus, the covariance between the travel times on two segments $i, j \in \{1, \dots, K\}$ is defined for each combination of time-bins (t_i, t_j) where $t_i, t_j \in \mathcal{T}$ are the time-bins associated with the traversal of segments i and j respectively, resulting in a total of $|\mathcal{T}| \times |\mathcal{T}|$ covariance values to be estimated for each pair of links and a total of $|\mathcal{T}|^2|A|^2$ where A is the set of links in the network. It should be noted that the covariance can vary over time with the travel time distributions themselves, or less frequently. The modified approach, abbreviated as MCS-TD-ST is as follows:

Step 1: Estimate the covariance:

- For segments $i, j \in \{1, \dots, K\}, i \neq j$, a sample of size S is selected from the set of trajectories traversing both i and j.
- The covariance is estimated for each time-bin combination (t_i, t_j) where $t_i, t_j \in \mathcal{T}$ by computing the covariance of link travel times that occurred in the corresponding time-bins, as a matrix Σ of size $|\mathcal{T}|^2|A|^2$ where $\Sigma[i, j, t_i, t_j] = cov(\theta_i^{t_i}, \theta_j^{t_j})$.

Step 2: Obtain the initial sample:

For iterations $j \in \{1, \dots, M\}$:

- Sample a departure time $\tilde{\vartheta}_j$ from ρ and find the corresponding interval $t_j = \phi(\tilde{\vartheta}_j)$.
- Sample a travel time $\tilde{\theta}_1$ for the first segment from $\pi_1^{t_1}$ where $t_1 = t_j$
- For each next segment, $i \in \{2, \dots, K\}$, sample a travel time $\tilde{\theta}_i$ from $\pi_i^{t_i}$ where $t_i = \phi\left(\tilde{\vartheta}_j + \sum_{l=0}^{i} \tilde{\theta}_l\right)$.
- Save the vector of samples $\tilde{\theta}^j = \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_K\}$ and their corresponding time bins $\tilde{t}^s = \{\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_N\}$.

Step 3: Obtain a correlated sample:

For each sample $j \in \{1, \dots, M\}$:

- Multiply the $\tilde{\theta}^j = \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_N\}$ by the covariance Σ^j, that $\Sigma[i,k] = cov\left(\theta_i^{\tilde{t}_i}, \theta_k^{\tilde{t}_k}\right)$ where $\tilde{t}_i, \tilde{t}_j$ are as saved in $\tilde{t}^s = \{\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_N\}$ to obtain a new $\tilde{\theta}^{j'}$.
- Set $\tilde{\Theta}_P^j = \sum_{i=1}^{K} \tilde{\theta}_i^{j'}$.

5.4.2 Metropolis-Hastings Simulation (MHS) Based Estimation Approaches

The approaches presented in this section are based on the Metropolis-Hastings (M-H) algorithm, a powerful Markov chain method to simulate multivariate distributions. A detailed introduction of the M-H can be found in the exposition paper by Chib and Greenberg (1995), where the authors provide a derivation of the method and guidance for its implementation. The exposition also demonstrates that a few different algorithms, most notably the Gibbs sampler, are special cases of the Metropolis-Hastings algorithm, originally pointed out by Gelman (1993). A brief introduction to the approach is provided below.

5.4.2.1 Metropolis-Hastings Sampling Based Estimation Assuming Stationary Spatially Correlated Random Variables (MHS-S)

The M-H algorithm is a Markov chain Monte Carlo (MCMC) method, based on the MCMC theory of generating samples from a target density $\pi(\cdot)$ using a transition kernel whose n^{th} iterate converges to $\pi(\cdot)$. The problem is to find the appropriate transition kernel $P(x,\cdot)$ that represents the move from a starting or current point x. The M-H algorithm utilizes a candidate generating density (often referred to as proposal distribution) denoted $q(x,y)$ such that $\int q(x,y)dy = 1$, so that is the process is at a point x, the density generates a value y from $q(x,y)$. However, for $\pi(\cdot)$ to be the invariant density of $P(x,\cdot)$, $q(x,y)$ should satisfy the reversibility condition $\pi(x)q(x,y) = \pi(y)q(y,x)$. The M-H algorithm allows for the utilization of a candidate generating density $q(x,y)$ without satisfying this condition but by correcting the number of moves made from x to y by introducing a probability of move $\alpha(x,y) < 1$, i.e., probability that the move is made. Then, transitions from x to y where $x \neq y$ are made so that $p_{MH} \equiv q(x,y)\alpha(x,y)$. Then $\alpha(x,y)$ is to be found so that $p_{MH}(x,y)$ satisfies the reversibility condition. The Gibbs sampler is a M-H algorithm adaptation with a special proposal distribution so that the acceptance probability is always equal to 1.

Further detail and derivation of the M-H algorithm can be found in the study by Chib and Greenberg (1995), but some relevant remarks are included here. The M-H algorithm is specified by the candidate-generating density $q(x,y)$. If a candidate sample is rejected, the step is not made, and the current value remains the next item in the sequence. The iterative procedure of the M-H algorithm can be summarized as follows: (1) given a position x^j, sample a proposal position y^j from the proposal or transition distribution $q(y;x^j)$, (2) accept this proposal with probability

$\min\left(1, \frac{p(y|D)}{p(x^j|D)} \frac{q(x^j;y)}{q(y;x^j)}\right)$, where D is the observed data. If this step is accepted, then $x^{j+1} = y$, otherwise $x^{j+1} = x^j$. A common parametrization of $q(y;x^j)$ is a multivariate Gaussian distribution centered around x^j.

As mentioned previously, draws are regarded as samples from the target density only after the chain has passed the transient stage and the effect of the starting value can be ignored. Therefore, the remaining question is how large of an initial sample should be discarded, which may depend on the actual starting point itself.

For the problem of sampling path travel time distributions via joint segment travel times, an issue of general affine invariance can be encountered in employing the standard M-H algorithm, occurring when distributions may have high aspect ratios due to variation in scale. Ensemble samplers have been proposed for general applications with affine invariance, due to Goodman and Weare (2010), where an ensemble of M-H walkers are used, with different starting points, in order to obtain a good estimate of the distribution. This study will employ Goodman and Weare's ensemble M-H sampler, implemented according to Foreman-Mackey et al. (2013).

An important caveat of the M-H algorithm that still holds for the ensemble sampler is that a proposal density needs to be specified and the entire walk of the M-H algorithm is based on the same underlying candidate-generating density. While this assumption may be appropriate for use with time-invariant random variables, this section presents a modification to be applied for time-varying random variables. The M-H algorithm was tested with two types of proposal functions (often referred to as moves, as they generate the next move in the procedure) – a Gaussian proposal function, originally recommended by Metropolis et al. (Metropolis et al., 1953) and a second type of proposal introduced by Goodman and Weare (Goodman and Weare, 2010), informally called the “stretch move”, which has been shown to outperform the standard M-H. Thus, these two types of proposal functions result in two variations of the MHS-S method abbreviated as MHS-S-S and MHS-S-G for the stretch and Gaussian move, respectively.

5.4.2.2 Metropolis-Hastings Sampling Based Estimation Assuming Time-Varying Random Variables with Time-Varying Spatial Correlation (MHS-TD)

The implementation presented above is suitable for the sampling of time-invariant random variables. A second variation for Metropolis-Hastings Sampling (MHS) based estimation with time-varying random variables and time-varying correlation structure MHS-TD, is presented here. The special case of time-varying random variables with stationary covariance is omitted as it is incompatible with the MHS implementation.

The MHS sampling approach was applied for the case of time-varying travel time distributions with time-varying spatial correlations via a mixture model. A mixture distribution is simply the probability distribution of a random variable derived from a collection of other random variables – in this case the time-interval specific cases for the time interval combinations resulting from the time of traversal at each link of the path. Then the distribution functions, both probability density and cumulative distribution, can be expressed as a convex combination (i.e., a weighted sum) of the other density and functions, respectively. Similar as above, two types of proposal functions were tested resulting in two variations of the MHS-RD method abbreviated as MHS-TD-S and MHS-TD-G for the stretch and Gaussian move, respectively.

For this approach, the time intervals for the traversal of the segments are used as the basis for the mixture components, as vectors of size K equal to the number of path segments. Then, each mixture component is the stationary MHS estimate for the given time intervals and the probabilities associated with each component i.e., the mixture weights are determined according to the application. Similar to the approach seen in the MCS-based methods, if the goal is to estimate the path travel time distribution accurately based on the network conditions the probabilities are determined empirically form the data. If the goal is to estimate the path travel time distribution to be experienced by a traveler, departure times can be seen as uniformly distributed and time-intervals can be determined accordingly.

5.4.3 Normal To Anything (NORTA) Distribution Estimation Approaches

The set of approaches presented in this section are based on the NORTA (NORmal To Anything) distribution model, originally introduced by Cario and Nelson (1997). The NORTA is a model for representing a multi-dimensional random vector $\boldsymbol{X}$ with arbitrary marginal distributions and any feasible correlation matrix using a transformation-oriented approach. Specifically, the model transforms a base standard multivariate normal vector $\boldsymbol{Z}$ to achieve the desired marginal distributions for the components of the input random vector $\boldsymbol{X}$ and an appropriately adjusted correlation matrix. Hence, $\boldsymbol{X}$ is referred to as having a NORTA (NORmal To Anything) distribution. Cario and Nelson's work is closely related to methods that transform a random vector $\boldsymbol{U}$ with uniformly distributed marginals, known as a *copula*. The authors describe the NORTA transformation as a two-step process, where the multivariate normal vector $\boldsymbol{Z}$ is first transformed into a multivariate uniform vector $\boldsymbol{U}$**,** which is then transformed into the desired input vector $\boldsymbol{X}$. The major contribution of the NORTA approach is that it is a general-purpose and easy-to-use tool, that is not based on approaches to mix distributions or exploit properties of a particular family of distributions. However, as can be expected, this is achieved at the expense of computational efficiency in fitting and random-variable generation, as the authors point out.

A complete introduction to the NORTA model can be found in the original work (Cario and Nelson, 1997) as well as the preceding publication (Cario and Nelson, 1996) that introduces many of the concepts for developing this model. This section presents summary of the NORTA model that should be sufficient for the implementation and modifications presented after.

The goal of the NORTA model is to define a random vector $\boldsymbol{X} = (X_1, X_2, \dots, X_K)'$ with specified properties, including marginal distributions $X_i \sim F_{X_i}, i = 1, 2, \dots, K$ where each F_{X_i} is an arbitrary cumulative distribution function (cdf) and a given $\mathrm{Corr}[\boldsymbol{X}] = \boldsymbol{\Sigma_X}$.

Then $\boldsymbol{X}$ is represented as a transformation of a K-dimensional standard multivariate normal (MVN) vector $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_K)'$ with correlation matrix $\boldsymbol{\Sigma_Z}$. Denoting the univariate standard normal cdf as $\boldsymbol{\Phi}$ and setting $F_X^{-1}(u) \equiv \inf\{x: F_X(x) \geq u\}$ as the inverse cdf, the NORTA vector $\boldsymbol{X} = \left(F_{X_1}^{-1}[\boldsymbol{\Phi}(\mathrm{Z}_1)], F_{X_2}^{-1}[\boldsymbol{\Phi}(\mathrm{Z}_2)], \dots, F_{X_K}^{-1}[\boldsymbol{\Phi}(\mathrm{Z}_\mathrm{K})]\right)'$. Thus, the transformation $F_{X_i}^{-1}[\boldsymbol{\Phi}(\cdot)]$ ensures that X_i has the marginal distribution F_{X_i} and the central problem in fitting the NORTA model is to select a $\boldsymbol{\Sigma_Z}$ that gives the desired $\boldsymbol{\Sigma_X}$ after the transformation. The approach to estimating the $\boldsymbol{\Sigma_Z}$ from a given $\boldsymbol{\Sigma_X}$ or observed data $\widetilde{\boldsymbol{X}}$ is not presented here as it is not the central problem in this study, and the details can be found in the original paper (Cario and Nelson, 1997).

5.4.3.1 NORTA Distribution Estimation Assuming Stationary Spatial Correlation of Time-Dependent Random Variables (NORTA-TD-S)

For this application, the NORTA estimation approach is adapted to sample jointly distributed time-dependent random variables with a stationary (i.e., time-invariant) correlation. Thus, for a path P composed of K consecutive segments $\{l_1, l_2, \dots, l_K\}$ travel time on segment $l_k \ \forall\ i \in \{1, 2, \dots, K\}$ and for a time interval $\mathcal{t} \in \mathcal{T}$ is a random variable $\theta_k^{\mathcal{t}}$ with a cdf $F_{\theta_k}^{\mathcal{t}}$. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)'$ and Θ_P again denote the path travel time i.e., $\Theta_P = \sum_{k=1}^{K} \theta_k$.

With the stationary covariance $\mathbf{\Sigma}_{\boldsymbol{\theta}}$ is invariant across the time intervals in $\mathcal{T}$, and thus a single corresponding $\mathbf{\Sigma}_{\boldsymbol{Z}}$ needs to be estimated and then samples from $\boldsymbol{Z}$ are to be transformed back into samples from $\boldsymbol{\theta}$. Assumptions and notation for the departure time and its distribution are as introduced previously in section 5.4.1.2. Sampling from the modified NORTA-TD distribution is then as follows:

For iterations $j \in \{1, \dots, M\}$:

- Sample a starting time $\tilde{\vartheta}_j$ is from the distribution ρ and find the corresponding time interval $t_j = \phi(\tilde{\vartheta}_j)$.
- Obtain a sample for $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{Z}})$.
- For the first segment, let $\tilde{\theta}_1 = \left(F_{\theta_1}^{t_1}\right)^{-1}[\mathbf{\Phi}(\mathrm{Z}_1)]$ wh ere $t_1 = t_j$.
- For each next segment $k \in \{2, \dots, K\}$, a travel time $\tilde{\theta}_k = \left(F_{\theta_k}^{t_k}\right)^{-1}[\mathbf{\Phi}(Z_k)]$ with $t_k = \phi\left(\tilde{\vartheta}_j + \sum_{i=0}^{k} \tilde{\theta}_i\right)$ i.e., the time interval corresponding to the departure time on segment i.
- Then, the sampled path travel time is $\tilde{\Theta}_P^j = \sum_{k=1}^{K} \tilde{\theta}_k$.

Therefore, while $\boldsymbol{Z}$ can be sampled jointly, the transformation from $Z_1, Z_2, \dots, Z_K$ to $\theta_1, \theta_2, \dots, \theta_K$ is done sequentially since the time interval t_k corresponding to the segment l_k is obtained based on the knowing the sampled values for θ_i for $i < k$.

5.4.3.2 NORTA Distribution Estimation Assuming Time-Varying Spatial Correlation of Random Variables (NORTA-TD-ST)

The second modification of the NORTA approach is one where the time-varying random variables also have a time-varying correlation. As introduced previously, such a correlation would require the computation of $|\mathcal{T}|^2$ values for each value in $\mathbf{\Sigma}_{\boldsymbol{X}}$. For a vector of K variables, NORTA needs to perform $\frac{K(K-1)}{2}$ operations (Cario and Nelson, 1997). Thus, for the time-varying correlation the number of operations would be $\frac{K(K-1)}{2}|\mathcal{T}|^2$ where $\mathcal{T}$ is the set of time intervals over which the distribution varies. However, since an important limitation of the NORTA method is the computational inefficiency of generating $\mathbf{\Sigma_Z}$, this modification with a time-varying correlation, is designed so as to estimate as few of the elements of $\mathbf{\Sigma_Z}$ as possible. Additionally, as previously seen in section 5.4.1.4, even if all elements of this matrix are computed only a portion of them are being utilized.

Then, sampling from the modified NORTA-TD-TS distribution was implemented in the form of a mixture distribution between the various cases of stationary distributions with their occurrence estimated according to the empirical data. Similar to the approach taken for the MHS estimation, the time intervals for the traversal of the segments are used as the basis for the mixture components, as vectors of size K equal to the number of path segments. Then, each mixture component is the stationary NORTA estimate for the given time intervals and the probabilities associated with each component i.e., the mixture weights are determined according to the application.

5.4.4 Lognormal Approximation Estimation Approaches

The second parametric estimation approach is one commonly used in the literature to approximate the sum of correlated non-negative random variables via a lognormal distribution (Mehta et al., 2007). This approach has been applied in studies considering stochastic networks, initially by Chen et al. (2018) for the case of time-invariant stochastic networks, and later adapted for the case of time-varying random variables (Chen et al., 2020) using the dynamic moment-matching method (DMM). This study utilizes a mixture model to implement a time-varying modification of the lognormal approximation, similar to the approaches taken for the MHS and NORTA estimation methods.

5.4.4.1 Lognormal Approximation for Time-Invariant Random Variables with Spatial Correlation

The lognormal approximation, according to the Fenton-Wilkinson approach (Fenton, 1960) approximates path travel time distributions by a lognormal distribution whose first origin moment M_P and second central moment D_P are the sum of the origin moments of the links and the sum of all of the elements of the covariance matrix of link travel times, respectively (Chen et al., 2018). For a path P composed of K consecutive segments $\{l_1, l_2, \dots, l_K\}$ with time-invariant mean of the logarithm of travel times $\{\mu_1, \mu_2, \dots, \mu_K\}$, standard deviations of the logarithm of travel times $\{\delta_1, \delta_2, \dots, \delta_K\}$ and correlation coefficient $\rho_{k,k'}$ for segments k and k', the moments can be calculated with the moment-matching method they can be found as follows:

$$M_P = \sum_{k=1}^{K} \exp(\mu_k + 0.5\delta_k^2),$$

$$D_P = \sum_{k=1}^{K} \exp(2\mu_k + \delta_k^2)(\exp(\delta_k^2) - 1)$$

$$+ \sum_{k,k' \in \{1,\dots,K\}} \rho_{ij} \sqrt{\exp(2\mu_k + \delta_k^2)(\exp(\delta_k^2) - 1) \exp\left(2\mu_{k'} + \delta_{k'}^2\right)\left(\exp\left(\delta_{k'}^2\right) - 1\right)}$$

Then the approximate lognormal distribution parameters for the path travel time are:

$$\mu_P = \ln\left(\frac{M_P^2}{\sqrt{M_P^2 + D_P}}\right), \delta_P = \sqrt{\ln\left(1 + \frac{D_P}{M_P^2}\right)}$$

Finally, the path travel time distribution is then approximated so that $\Theta_P \sim Lognormal(\mu_P, \delta_P^2)$, or equivalently $\ln(\Theta_P) \sim N(\mu_P, \delta_P^2)$.

5.4.4.2 Lognormal Approximation for Time-Dependent Random Variables

The approach for using the lognormal approximation for time-dependent random variables in this study was developed using a mixture model of the stationary lognormal approximations across all time-intervals combinations for the time of traversal for each path segment.

An approach for lognormal approximation for the case of time-varying random variables was introduced by Chen et al. (2020). However, the presented approach is not suitable for the estimation of path travel time distributions across a time period, but rather for a given single time of day. Due to this and in order to maintain the consistency with some of the other modifications based on mixture models, such as the MHS and NORTA presented in sections 5.4.2 and 5.4.3, respectively.

5.5 Numerical Experiments

Numerical experiments were performed to evaluate the performance of the path travel time estimation approaches. This section introduces the study sites and data are introduced in 5.5.1, followed by the experimental design in 5.5.2, the performance measures used in this study in 5.5.3, leading into the results and analysis in 5.5.4.

5.5.1 Study Sites and Data

The methods for estimation of path travel time distributions were intended to be used for an urban network with the availability of trajectory data. Hence, this study used simulated trajectory data on the network of Chicago, with 1578 nodes and 4805 links as shown in Figure 5-1.

Trajectories observed for departure times in the morning peak period from 6 AM to 9 AM were considered.

To test the performance of the presented methods, a set of 100 paths were selected from the network, covering 902 of the network links. The paths were selected so as to have a large number of vehicle traversals, i.e., trajectory observations, that would provide a reliable estimate of the path travel time distribution to be used as the ground truth.

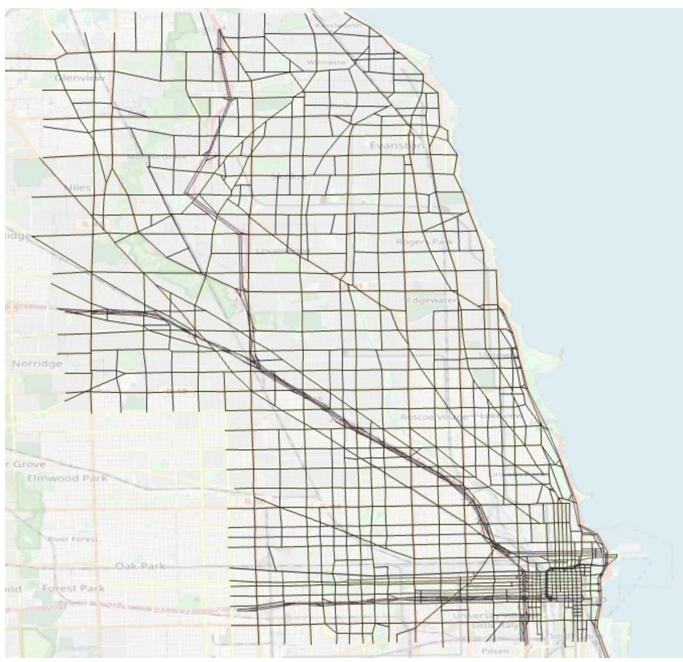

Figure 5-1. Large scale Chicago network

5.5.2 Experimental Design

The vehicle trajectory data for the morning peak period consisted of any vehicle trajectory observations traversing any portion of the paths considered. For the 100 selected paths, the travel time distributions estimated from empirical observations were used as ground truth. Link-based estimation was performed with the four Monte Carlo sampling approaches: MCS-I, MCS-TD, MCS-TD-S and MCS-TD-ST, the two Metropolis-Hastings sampling methods: MHS and MHS-

TD, three NORTA-based approaches NORTA-S, NORTA-TD-S and NORTA-TD-ST, and two variations of the lognormal approximation Lognormal and Lognormal-TD. The approaches with time-dependence were tested using a uniform time interval $\delta = 10\ min$. The number of samples M for each of the approaches was set to 1000.

5.5.3 Performance Measures

To evaluate the performance of the different approaches for the estimation of path travel time distributions, this chapter uses two sets of accuracy measures along with one measure of efficiency.

The first set of accuracy measures are statistical measures for the comparison of non-parametric distributions.

- The Kolmogorov-Smirnov (KS) statistic is a measure of closeness for two probability distributions and one of the most useful and general nonparametric methods for comparing two samples (Chakravarty et al., 1967; Daniel, 1990). The KS statistic takes the largest absolute difference between two distribution functions and can take values from 0 to 1, where a value closer to 0 signifies a better fit.
- The Epps-Singleton (ES) statistic (Epps and Singleton, 1986) is a test using the empirical characteristic function and based on a quadratic form in differences between respective components of empirical characteristic functions of two samples.
- The Wicoxon rank-sum test (often also referred to as the Mann-Whitney-Wilcoxon (MWW) test) is a nonparametric test with the null-hypothesis that the samples X and Y are drawn from the same distributions, with the alternative hypothesis that the values

of one sample are more likely to be larger than those of the other (Mann and Whitney, 1947; Wilcoxon, 1945).

The second set of accuracy measures are based on assessing the performance of the distribution estimation approaches in estimating select quantities related to the distribution, similarly to the approach taken by Kim and Mahmassani (2014b). Here, the mean absolute percentage error (MAPE) is calculated for each of six estimated quantities for each of the distributions: mean, coefficient of variation (CV), 25th percentile, median, 75th percentile and 90th percentile of the distribution. If x is the quantity to be estimated, and $\hat{x}$ is the estimate of x for a given method, then for n estimated values:

$$MAPE(\%) = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|.$$

In addition to the three statistical measures and the $MAPE$ for the six distribution quantities, the final measure to be considered is the computational run time for obtaining each estimate (or sample) of the distribution.

5.5.4 Results and Analysis

The numerical experiments results are summarized via the average distribution comparison statistics, the MAPE across six distribution quantities and mean computational run times across the 100 paths for each of the total of 13 estimation methods. Table 5-3 presents the results for the computational run times and the distribution comparison statistics KS, ES and MWW with their corresponding p-values, while the MAPE values for the distribution quantities are shown in Table 5-4.

Table 5-3. Computational run times and accuracy measures for the estimation approaches, including the KS, ES and MWW statistics and their corresponding p-values.

Method	Run time (s)	KS	KS-p	ES	ES-p	MWW	MWW-p
MCS-I	0.027	0.289	0.130	39.928	0.071	−1.295	0.071
MCS-TD	4.630	0.431	0.031	115.602	0.007	0.540	0.007
MCS-TD-S	9.431	0.493	0.015	447.273	0.010	−0.257	0.010
MCS-TD-ST	10.420	0.457	0.028	156.239	0.013	1.232	0.013
MHS-S-S	1942.210	0.416	0.009	79.999	0.018	−4.771	0.018
MHS-S-G	1724.065	0.722	0.000	198.151	0.000	0.684	0.000
MHS-TD-S	1158.337	0.266	0.018	130.732	0.021	2.882	0.021
MHS-TD-G	1158.337	0.263	0.016	124.040	0.013	2.620	0.013
NORTA-S	19.253	0.143	0.410	9.382	0.341	−0.593	0.341
NORTA-TD	19.818	0.255	0.131	26.409	0.109	−0.523	0.109
NORTA-TD-ST	2491.961	0.192	0.240	13.537	0.175	1.124	0.175
LOGN	0.314	0.295	0.075	38.016	0.055	−0.463	0.055
LOGN-TD	12.660	0.220	0.198	15.677	0.166	0.398	0.166

Table 5-4. MAPE (%) values for six distributions quantities for all approaches

Method	mean	CV	25th	median	75th	90th
MCS-I	22.57	333.34	39.96	15.90	20.47	28.42
MCS-TD	23.44	83.34	46.88	28.02	26.34	24.53
MCS-TD-S	33.25	182.01	49.42	41.98	40.05	36.62
MCS-TD-ST	24.34	75.28	55.19	40.56	32.57	28.61
MHS-S-S	11.57	63.67	43.31	24.43	23.12	11.30
MHS-S-G	19.02	93.89	85.65	39.93	40.56	24.27
MHS-TD-S	13.46	98.13	27.92	10.68	11.40	21.52
MHS-TD-G	12.36	89.13	26.55	10.62	9.63	20.15
NORTA-S	2.49	16.94	8.59	6.53	4.85	4.38
NORTA-TD	28.36	544.41	35.09	14.26	24.49	37.75
NORTA-TD-ST	10.26	180.60	26.64	7.65	7.91	10.58
LOGN	22.20	1139.32	56.65	12.82	17.57	24.92
LOGN-TD	8.63	128.78	36.75	8.26	7.54	11.16

The results in Table 5-3 show a significant variation in the computational run times between different approach types and their modifications, making it apparent that some approaches may be much more suitable for real-time or online applications, while others may only be relevant for offline applications. Specifically, the MHS approaches and the NORTA approach with time-varying correlations exhibit significantly higher computational run times.

The KS statistic is considered further in Figure 5-2, as the most widely used measure for distribution comparison and testing. It can be observed that the MCS approaches show a very high variation in KS statistic value, while the NORTA approaches are consistent it their low variation. It is interesting to observe the NORTA-S (stationary) approach which exhibits the lowest mean and very low variation for the KS statistic. Low KS statistic values indicate a better fit of the distributions.

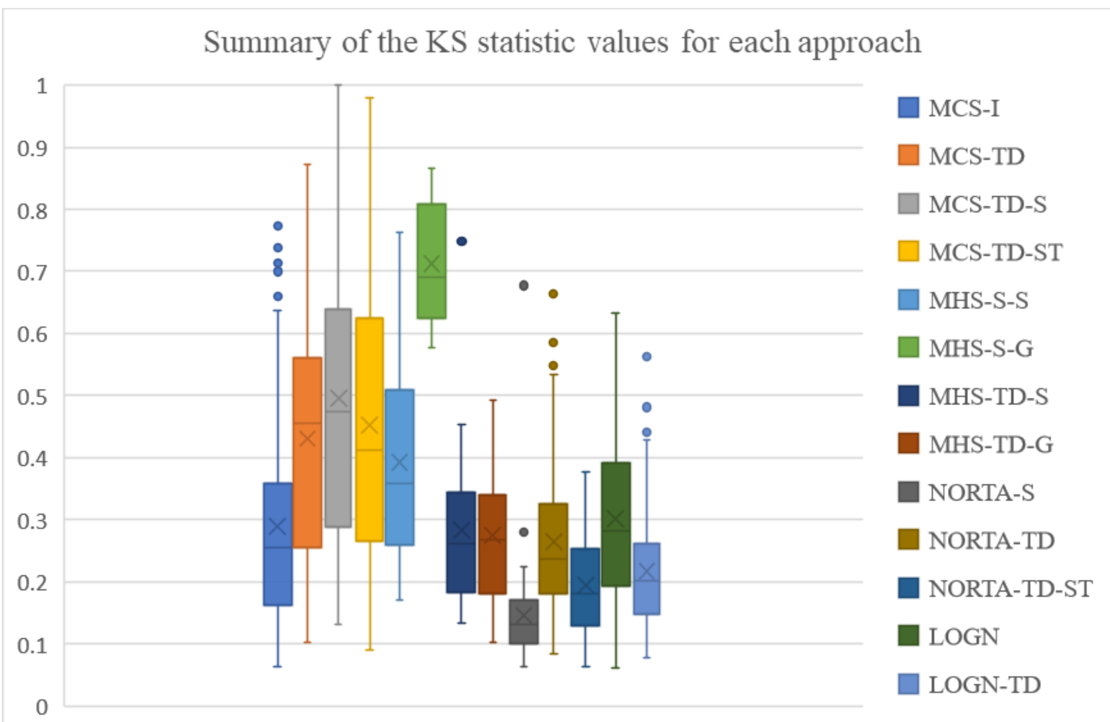

Figure 5-2. Box Plot of the KS statistic values for all estimation approaches.

To consider the other statistics and compare across them, the focus is on the corresponding p-values since the statistics themselves have varying scales and interpretation, but the p-values translate each statistic into the probability under the null hypothesis, which in all cases is a proxy for the estimated and ground truth observations being from the same distribution. The average p-values for the KS and ES statistics are shown in Figure 5-3, and the MWW p-values are omitted as they are equal to those for the ES statistic. High p-values are preferable as they indicate higher probability that the simulated sample is drawn from the same distribution as the ground truth. It can be observed that the NORTA-S approach results in especially high average p-values for both the KS and ES statistic.

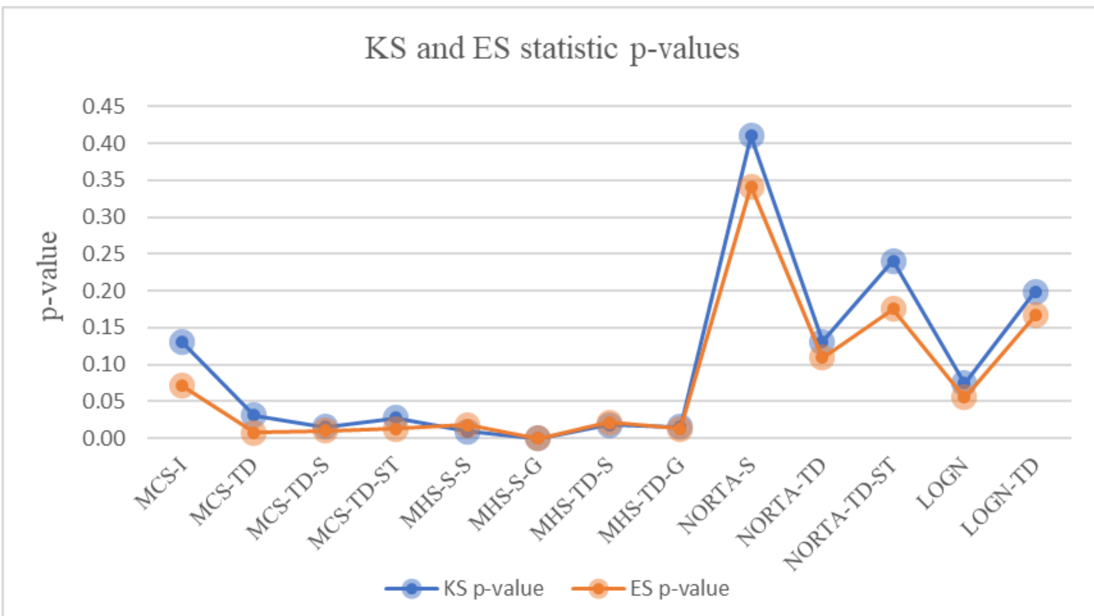

Figure 5-3. Average p-values for the KS and ES statistics across all approaches

The p-values are further analyzed via the summary of their variation for the ES statistic across all approaches, shown in Figure 5-4.

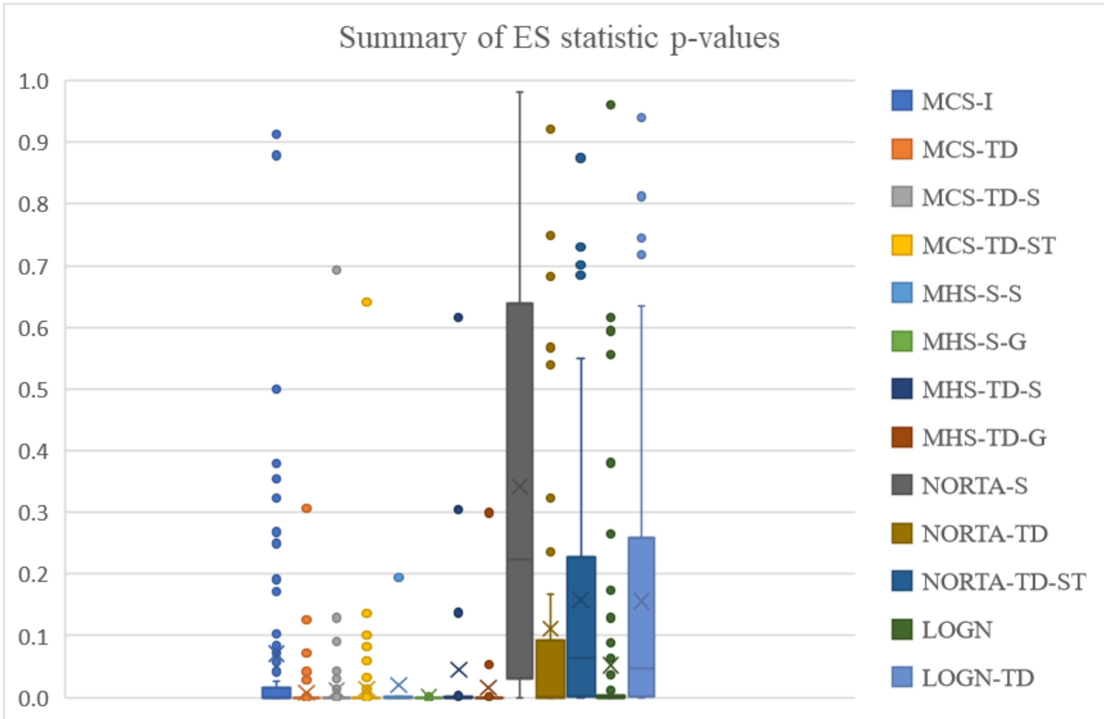

Figure 5-4. Box Plot of the ES statistic p-values values for all estimation approaches

While the NORTA-S approach had high average p-values, it can be observed that it also exhibits the highest variation in p-values with the 25th and 75th percentile at 0.031 and 0.637, respectively. These results make it difficult to conclude that the promising NORTA-S approach can be reliably used for accurate estimation of path travel time distributions. Most approaches, other than the NORTA and Lognormal approximation have small variation and undesirably low p-values. Some interesting approaches to consider based on the results in Figure 5-4, may be the NORTA-TD and Lognormal-TD approaches with lower mean p-values compared to NORTA-S but significantly higher than the other methods and with lower variation. This is further supported by considering the plot of cumulative distribution functions for the ES statistic p-values, shown in Figure 5-5 for the NORTA and Lognormal approaches. The plots show that for the p-value maximization problem, the cumulative distribution function for the p-values of the NORTA-S is stochastically dominated at the first order by that of the other four approaches. At any cumulative probability value, the NORTA-S approach has the lowest ES p-value among these approaches.

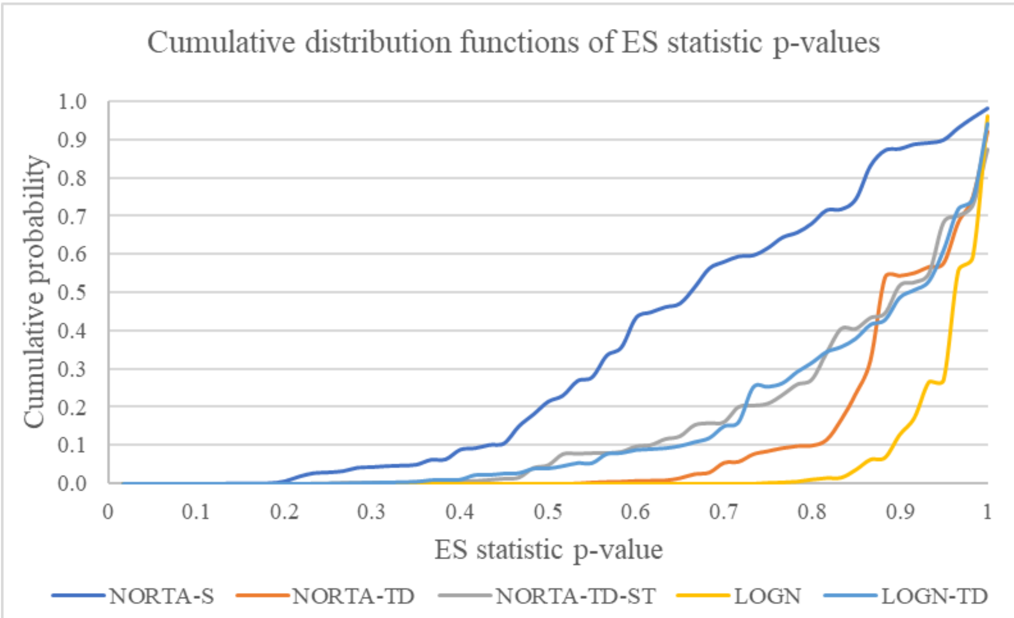

Figure 5-5. Cumulative distribution function of ES statistic p-values for five of the estimation approaches

The estimation approaches can be further evaluated via their performance in terms of estimating the distribution quantities: the mean, CV, 25th percentile, median, 75th and 90th percentile values. The MAPE values for these quantities for each approach are shown in Table 5-4.

These results show some variation in the estimation errors for the distribution values with the different approaches. The MAPE on the mean and CV values are additionally shown in Figure 5-6. With the mean error values, the NORTA-S approach shows the lowest error for the estimation of both the mean and CV values of the path travel time distributions. Interestingly, the NORTA-TD-ST approach shows better performance for the mean and CV MAPE values compared to the NORTA-TD approach, showing the benefit of using a non-stationary correlation structure. Similarly, the Lognormal-TD improves on the Lognormal approximation, and the MHS-TD-S and

MHS-TD-G approaches have lower error compared to the time-invariant versions MHS-S, except for the low MAPE on the mean with MHS-S-S.

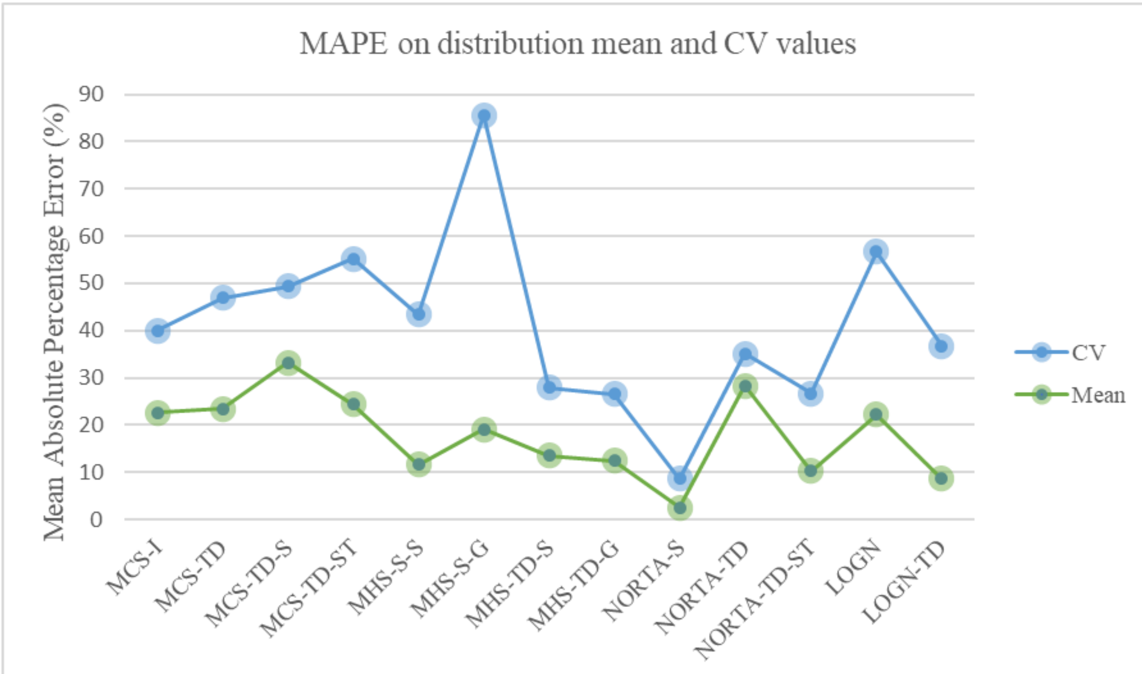

Figure 5-6. MAPE for the mean and CV of path travel time distributions for all approaches

The MAPE for the remaining distribution quantities, the 25th percentile, median (50th percentile), 75th and 90th percentile are presented in Figure 5-7. From the visualization, it can be seen that a few of the approaches show an imbalance in their ability to accurately estimate the right or left tail of the distribution. Specifically, the MHS-S-S shows low errors for the high percentiles, but high errors for the low percentiles, while the opposite is true for the MHS-TD approaches. For the NORTA-TD and Lognormal approximation, the error increases farther along in the right tail of the distribution. Only a few of the approaches are consistent in their estimation, namely the MCS-TD and MCS-TD-S with relatively high MAPE near 30% and 40% respectively for all four quantities, and NORTA-S, NORTA-TD-ST with errors averaging near 5% and 10%, respectively. Finally, the Lognormal-TD shows consistent errors near 10% for all four quantities.

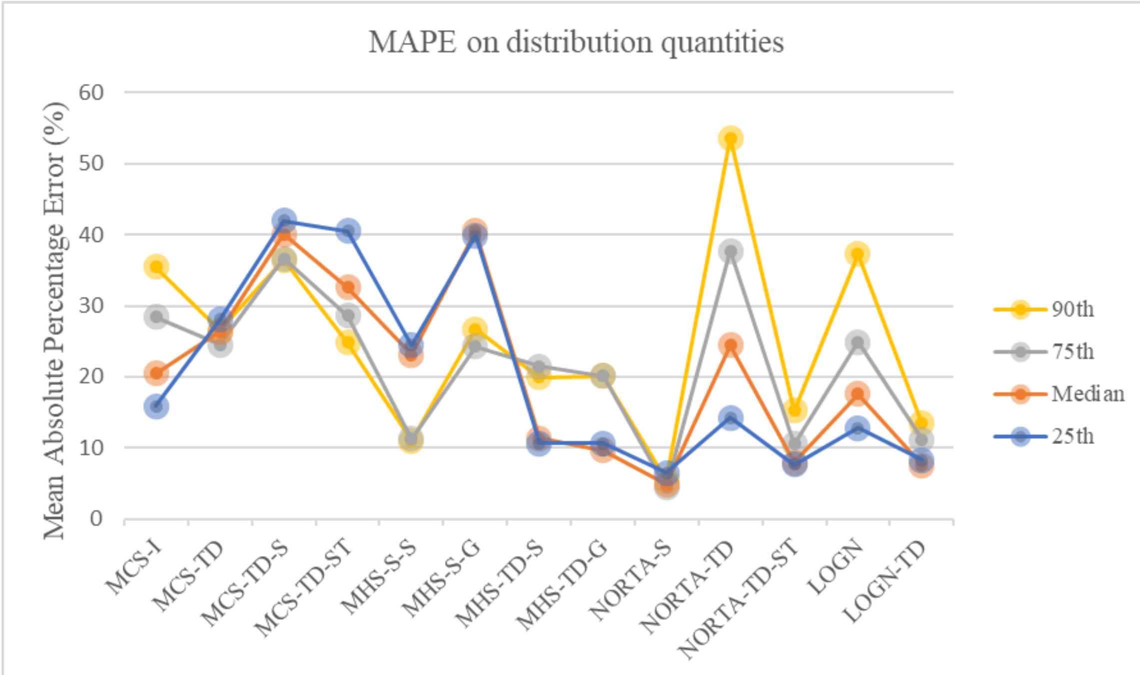

Figure 5-7. MAPE for the 25th percentile, median (50th percentile), 75th and 90th percentile of path travel time distributions

Combining these results from Figure 5-6 and Figure 5-7, with those from the previous figures Figure 5-2, Figure 5-3, and Figure 5-4, a few of the approaches can be discussed further. The NORTA-S approach achieves low mean error values, but be the least reliable approach among the thirteen, as its p-values vary significantly across the numerical tests. The modification NORTA-TD-ST performs better when it comes to the MAPE across distribution quantities, while also having relatively high p-values with significantly less variation compared to NORTA-S. However, while this approach is accurate across the different measures, it should be noted that it has the highest run times across all presented methods, since it requires the estimation of multiple covariances, which is the computationally demanding aspect with NORTA. There are a few considerations to implementing this approach with reduced computational times. Firstly, recent advances in implementing NORTA have introduced ways to improve its efficiency, including an approach by Niaki and Abbasi (2008) for generating NORTA correlation matrices using artificial neural networks and a recent graphical processing unit (GPU)-accelerated NORTA algorithm for high dimensional multivariate simulation by Li et. al. (2019). Secondly, the time varying NORTA correlation matrices can be pre-computed offline for any online applications, thus eliminating large run time requirement. The Lognormal approximation methods also warrant further discussion, especially the Lognormal-TD, which has accuracy close to that of NORTA-TD-ST. The Lognormal-TD was implemented as a mixture model and as such it has higher computational run times relative to the time invariant Lognormal.

5.6 Conclusions and Future Work

This chapter addresses the question of estimating path travel time distributions in stochastic time-varying networks with generalized correlation structures. Four classes of estimation approaches are presented, including Monte Carlo simulation and Metropolis-Hastings sampling as non-parametric approaches, as well Normal to Anything (NORTA) distribution and Lognormal distribution approximations as parametric estimation approaches. Based on the presented taxonomy for the estimation of path travel time distributions, variations of the standard approaches for each of the four classes are introduced, to allow for different types of assumptions, input categories and output requirements categories.

The numerical experiments in this study are performed on a set of paths on the Chicago network, calibrated using real-world data. The results and analysis compare the performance of the thirteen presented approaches across a number of accuracy measures and the computational run times. The most interesting approaches are the NORTA estimation method and its modification with time-varying correlations NORTA-TD-ST, each showing better performance across different measures.

Chapter 6 Reliable A Priori Path Finding in Stochastic Dynamic Networks[2]

6.1 Overview

This chapter considers the problem of finding time-dependent reliable least-time paths (RLTP), where joint time-varying link travel time distributions are unknown and path travel time distributions are to be estimated. The RLTP problem thereby connects the path travel time distribution estimation approaches with their application to reliability-based path finding in stochastic dynamic networks with spatio-temporal dependencies.

This chapter presents an approach for the RLTP problem for a priori paths via a path generation approach. The path-generation algorithm is a label-correcting (LC) algorithm in line with those developed by (Miller-Hooks and Mahmassani, 2003a, 2000b; Nie and Wu, 2009a, 2009b) with adjustments for the problem considered here. Specifically, given the methodological difficulties of applying dominance criteria at intermediate nodes in stochastic networks with dependent travel times, this chapter presents two sets of approximate dominance criteria. The path generation approach is applied with a novel path travel time estimation approach, utilizing a time-dependent NORmal To Anything (NORTA) sampling technique from Chapter 5. Finally, this chapter applies the proposed solution approach to the large-scale Chicago network using simulated vehicle-trajectory data, performs a priori path finding for six reliability-based least-time objective functions, and evaluates the performance of the approximate path dominance criteria with respect to an exact solution.

---

[2] This chapter builds on the work in (Filipovska and Mahmassani, 2020b) and parallels an article under review for presentation at the 2022 101st Annual Meeting of the Transportation Research Board.

6.2 Problem Definition and Methodological Difficulties

This section defines the problem and the notation for the a priori RLTP problem. Firstly, the STV network with generalized link travel time correlations is presented as the setting for the problem in section 6.2.1. Secondly, definitions and distinctions of a priori and adaptive routing in STV networks are presented in section 6.2.2, followed by a discussion of optimality, and some of the characteristics of the problem that are crucial for the solution methodology in the context of correlated STV networks.

6.2.1 Stochastic Time-Varying Network Modeling and Notation

Let the STV network be a directed graph $G(N, A, \mathcal{T})$, where N is the set of $|N| = n$ nodes and A is the set of $|A| = m$ links, and $\mathcal{T}$ is the set of time periods. The link travel times are assumed to be random variables jointly distributed across time. The travel time on each link (i, j) at time t is a continuous positive random variable, denoted Θ_{ij}^{t}, constrained by a minimum and maximum possible travel time value with a truncated distribution π_{ij}^{t}. Note that here capital Θ_{ij} is used to denote the random variable, while the values it takes are denoted with θ_{ij}. The dependencies between the link travel times are defined via link-pairwise covariances, which vary over time-period pairs, so that $cov(\Theta_{ij}^{t_1}, \Theta_{kl}^{t_2})$ is the covariance between the travel time on link (i, j) during time interval $t_1 \in \mathcal{T}$ and that on link (k, l) during time period $t_2 \in \mathcal{T}$. This study assumes that link travel time distributions vary across the time periods in the set $\mathcal{T}$ and can be considered static outside the peak period covered by $\mathcal{T}$. An alternative and stronger assumption would be that the set of time intervals $\mathcal{T}$ covers the entire necessary time period, i.e., any path of interest can be traversed within the duration of the entire time period covered by the set $\mathcal{T}$ with probability 1 for a range of chosen departure times. Since the travel times for each link have truncated distributions, the duration of the time period covered by $\mathcal{T}$ can be determined for the given network on the basis of the maximum travel times on its links.

6.2.1.1 Estimation of Path Travel Time Distributions for Path Finding

An important challenge in modeling an STV network is the estimation of its path travel time distributions. If the network link travel time distributions with their temporal variation and spatio-temporal correlations are modeled as outlined above, the path travel time distribution for a given path and departure time cannot be directly extracted from the link travel time distributions and needs to be estimated instead. The problem of path travel time distribution estimation was introduced in Chapter 5, and the approaches presented there are relevant and will be applied with the solution approach for this chapter. A few of the key aspects of the problem definition are presented here so as to remain consistent with the notation used in this chapter.

Suppose a path P is composed of L consecutive links $\{(j_0, j_1), (j_1, j_2), \dots, (j_{L-1}, j_L\} \in A$. To simplify the notation, let the travel time distribution for each link $(j_{k-1}, j_k)\ \forall\ k \in \{1, \dots, L\}$ be a time-varying random variable, so that Θ_k^t is a distinct random variable of the travel time on link (j_{k-1}, j_k) during time period $t \in \mathcal{T}$ with a distribution π_k^t truncated by a minimum and maximum travel time θ_k^{min} and θ_k^{max}. Let Θ_P denote the travel time random variable for path P with a corresponding density function Π_P and cumulative distribution (cdf) U_P, where for a given travel time T, $U_P(T_P) = P(\Theta_P \leq T)$.

The presented solution methodology assumes known marginal link travel time distributions, estimated time-varying correlations, and the ability to perform conditional sampling.

Thus, the solution approach can utilize the estimation methods presented by Filipovska et al. (2021), and this chapter applies the best performing approaches from Chapter 5.

6.2.1.2 Assumptions for Path Finding Approaches

Path finding in STV networks with dependencies is made further difficult by the assumption of first-in-first-out (FIFO) consistency. In deterministic networks with time-varying link travel times, a link is FIFO if it is modeled so that entering the link later must result in leaving it later. In STV networks the travel time on each link, and by extension the arrival time at its end node, is a random variable. The FIFO consistency assumption can be extended to stochastic dynamic network problems in different ways, often depending on the specific problem definition. Studies by Miller-Hooks and Mahmassani (2003a, 1998a) present a definition in which a later departure time leads to a later arrival time with probability equal to 1 for any link in the network. Specifically, for a link $(i,j) \in A$, with a travel time θ_{ij}^{t} at time t, $P\{s + \theta_{ij}^{s} \leq t + \theta_{ij}^{t}\} = 1 \; \forall \; s \leq t$. An alternative condition, as a definition for Stochastic FIFO, is presented in later studies by Nie and Wu (2009a, 2009b) where a probability density function π_{ij}^{t} is FIFO consistent if its CDF satisfies the following condition: $U_{ij}^{t_1}(T_b - t_1) \geq U_{ij}^{t_2}(T_b - t_2) \; \forall \; t_1 \leq t_2$, for a given arrival time T_b. This condition ensures earlier arrival times are more probable with earlier departure times. In a network with independent link travel time random variables, the FIFO-consistency of all individual links implies the FIFO-consistency of any given path. Allowing for correlations or interdependent link travel times, an equivalent stochastic FIFO-consistency of the network can be established by imposing the condition on all marginal and conditional link travel time distributions, which then ensures the condition is satisfied for the travel time distribution of any path as the sum of jointly distributed random variables (Nie and Wu, 2009b).

6.2.2 A Priori Path Finding and Optimality in STV Networks

The central problem in this chapter is finding optimal a priori paths under reliability-based least-time objectives. The definition of a priori path finding and its distinction from adaptive routing is initially described in Chapter 2, and the key characteristic of a priori path finding is that a solution consists of a single path optimizing a given objective function for an origin-destination pair and a specified departure time. Formally, the problem is to determine a path from origin node $r \in N$ to all destination nodes $j \in N \backslash \{r\}$ for each of a range of departure times t. Some simplifying assumptions in this study are that waiting at nodes and cyclic paths are not permitted. Specifically, this chapter extends the work of Filipovska and Mahmassani (2020b) to present an approach for the RLTP problem using a path generation approach with alternative path dominance criteria using approximations to reduce the required computational effort and find a good approximate solution to the problem.

Additionally, the solution approach in this chapter can be used for one or more objectives focused on finding reliable least-time strategies. In stochastic networks, travel time distributions can be compared along several different criteria (i.e., expected travel time, variance, travel time budget, α-confidence travel time, etc), and different paths may be preferable under each objective. Reliable least-time strategies can be generally defined around least-time objectives, and some examples from the literature include: least expected travel time (LET) (Miller-Hooks and Mahmassani, 2000a), least possible time (Miller-Hooks and Mahmassani, 1998a), least α-confidence travel time (Chen et al., 2018; Zeng et al., 2015), on-time arrival probability (Yang and Zhou, 2017), reliable shortest paths (Chen et al., 2020; Zhang et al., 2017).

Two objective types are considered in this study: Value at Risk (VaR) and Conditional Value at Risk (CVaR). A number of previous studies for path finding in stochastic networks use the least α-confidence travel time objective (Chen et al., 2018; Nie and Wu, 2009a; Zeng et al., 2015), which is an application of the Value at Risk (VaR) measure of risk, typically used for evaluating the risk of financial investments. VaR is evaluated for a given probability value $p = 1 - \alpha$, so that the probability of a loss greater than $\text{VaR}(p)$ is p, while the probability of a loss less than $\text{VaR}(p)$ is $1 - p = \alpha$ (Holton, 2012). Applied to travel time distributions, the $\text{VaR}(1 - \alpha)$ is equivalent to the α-confidence travel time. Therefore, finding the least α-confidence travel time objective is equivalent to minimizing $\text{VaR}(1 - \alpha)$. Thus least α-confidence travel time is equivalent to least VaR for probability $p = 1 - \alpha$. However, VaR has been criticized as a measure of risk, most importantly for not being a coherent risk measure due to its violation of the sub-additivity property (Dowd, 2007). The Conditional Value-at-Risk (CVaR) is a coherent risk measure, defined as the average of VaR values for $p \geq 1 - \alpha$, i.e., the expected value for the worst α-percentile cases (Rockafellar and Uryasev, 2000). Its application from evaluating investment portfolios to travel time distributions needs to consider that a positive return on investments is seen as gain, while should be seen as a loss when considering travel times. However, moving from the VaR-based travel time objective of least α-percentile travel time to the CVaR-based objective of least expected travel time above the α-percentile can be understood as follows: the VaR-based objective captures some level of risk by considering and comparing the travel times at the α-percentile, while a CVaR-based objective captures the expectation of how much greater the travel time can get in that α-percentile tail. This study will consider and compare both of these objective types with varying

α values. In applications, different α values can be used to represent different travelers' sensitivity to risk and reliability preferences that may vary due to trip purpose or activity type.

6.2.3 Characteristics of STV Networks with Generalized Correlations

For a complete problem definition, this section presents and discusses some of the characteristics of STV networks with generalized correlations, how they relate to the problem at hand and the resulting sources of difficulty for the solution approach. Firstly, the non-applicability of Bellman's principle is demonstrated, secondly, some aspects of path comparisons are addressed.

6.2.3.1 Non-applicability of Bellman's Principle

The non-applicability of Bellman's principle in the context of stochastic networks has been shown for different cases and dominance criteria in STV networks with correlated link travel times (Hall, 1986; Huang and Gao, 2018; Nie and Wu, 2009b; Prakash and Srinivasan, 2017). The relevant results are presented here for completeness. Objective-based non-applicability of Bellman's principle, for example with respect to expected value dominance or α-confidence least time paths, have been shown in the literature (Gao and Huang, 2012; Huang and Gao, 2018). While Bellman's principle with respect to first-order stochastic dominance can be maintained for a number of path finding problems in STV networks (Nie and Wu, 2009b), and even some appropriately defined STV networks with limited dependencies (Nie and Wu, 2009b), the principle does not hold for STV networks with generalized correlations.

First-order stochastic dominance (FSD) is a case of partial stochastic ordering where a path k_{OD} dominates path l_{OD} in the first order if $U_k(T_s) \geq U_l(T_s) \; \forall \; T_s \in [T_{min}, T_{max}]$; and $\exists$ at least one nonzero Lebesgue measure open interval $\Lambda \in [T_{min}, T_{max}]$ s.t. $U_k(T_s) > U_l(T_s) \; \forall \; T_s \in \Lambda$. By this definition, a path l_{OD} is FSD-admissible if $\exists$ no path k_{OD} such that k_{OD} dominates l_{OD} for all departure times. Additionally, an FSD-admissible path l_{OD} is FSD-optimal if $\exists$ a departure time t_0 and an nonzero Lebesgue measure open interval $\Lambda \in [T_{min}, T_{max}]$ s.t. $U_l(T_s) > U_k(T_s)\ \forall\ T_s \in \Lambda, \forall\ l \neq k$.

Example 1. Consider the illustrative example network in Figure 6-1 and the joint link, subpath and path travel time realizations shown in Table 6-1. Bellman's principle with respect to FSD would claim that sub-paths of FSD-admissible paths must also be FSD admissible. Let us consider the paths *O-a-b-D* and *O-a-e-b-D* with their sub-paths to node *b*, i.e., *O-a-b* and *O-a-e-b*. Comparing the full paths, we notice that *O-a-e-b-D* is FSD-admissible (i.e. not dominated by *O-a-b-D*). However, its sub-path *O-a-e-b* is not FSD-admissible as it is dominated by *O-a-b*. Therefore, by counterexample we have shown that an FSD-admissible path can have a sub-path that is not FSD-admissible.■

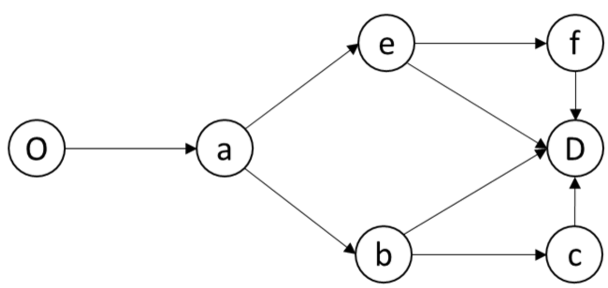

Figure 6-1. Example network for Example 1

Table 6-1. Possible joint link, sub-path, and path travel time realizations for Example 1

Realization	Link Travel Times							
	(O,a)	(a,b)	(a,e)	(e,b)	*O-a-b*	*O-a-e-b*	*O-a-b-D*	*O-a-e-b-D*
1	1	1	0.5	0.5	2	2	5	5
2	1	2	1	1	3	3	4.5	4.5
3	1	2	1.5	1.5	3	4	4	5
4	2	1	0.5	1.5	3	4	5	6
5	2	2	1	1	4	4	6	6
6	2	2	1	1.5	4	4.5	5.5	6

6.2.3.2 Path Comparisons

The problem of path comparisons is relevant and important for this problem, since path comparisons are at the foundation of the path finding approach. As indicated in the section above, first-order stochastic dominance (FSD) is a common criterion for path comparison in stochastic dynamic networks. Previous studies have shown that the set of a priori FSD non-dominated paths in STV networks will contain paths that may be inefficient in that, though non-dominated, they would never contribute to an adaptive routing strategy (Miller-Hooks and Mahmassani, 2003a). However, in STV networks with dependencies, the FSD comparison does not allow for the elimination of sub-paths at intermediate nodes due to the non-applicability of Bellman's principle. This chapter will present and compare the use of two variations of the FSD criterion for a priori path finding: firstly, the traditional FSD criterion with Bellman's principle, yielding approximate solutions, and secondly, an exact-to-approximate variation of the FSD criterion that makes use of the truncated travel time distributions in the network and the knowledge of the minimum and maximum possible travel times. The details of these approaches are presented as part of the solution methodology.

6.3 Solution Methodology

This section presents a solution approach for a priori path finding for the time-dependent reliable least-time path (RLTP) problem via a path generation algorithm. The algorithm is defined to utilize a variety of possible criteria for stochastic path comparison and dominance. Hence, the approach for stochastic path comparisons and dominance is described first in 6.3.1, followed by the time dependent RLTP solution algorithm in 6.3.2. With the dominance criteria presented here, the solution approach can be used for exact or approximate generation of eligible paths and can be applied for a priori path finding for a range or least-time reliability-based objectives, as presented in section 6.2.2. This section concludes with a note on the estimation of path travel time distributions in 6.3.3, an important aspect of the path finding and evaluation methodology.

6.3.1 Stochastic Path Comparisons and Dominance

The previous section introduces some of the methodological difficulties of the path finding problem in stochastic dynamic networks, including the non-applicability of Bellman's principle for path comparisons with correlated link travel times. Miller-Hooks and Mahmassani (1998b) present a general framework for determining a priori nondominated least time paths in stochastic time-varying networks. The authors present four different implementations with three different notions of dominance, under the assumption of independence of travel time random variables across links and across different time intervals. The three dominance criteria include deterministic pairwise dominance, stochastic pairwise dominance and expected value pairwise dominance. Deterministic pairwise dominance can be applied to the stochastic network with dependencies, with appropriate modeling of truncated link travel time distribution. However, stochastic pairwise dominance via the FSD criterion and expected value pairwise dominance cannot be applied for path generation in networks with correlations, as they cannot be applied at intermediate nodes, as described in section 6.2.3.

In this study, two types of dominance criteria are presented: an approximate adjusted FSD criterion that can be applied at intermediate nodes, and a deterministic dominance criterion that can be applied at intermediate nodes and relaxed to allow for approximate solutions.

Deterministic dominance: A path k is deterministically dominant over path l for a given departure time if, for all realizations of the network, the travel time along path k is always equal to or less than the travel time on path l. Let θ_k and θ_l denote the travel time random variables on the paths k and l with cumulative distribution functions (cdfs) U_k and U_l, respectively. If the travel times on the two paths are independent, comparing the two paths is analogous to comparing the two path travel time distributions independently and establishing that $P(\theta_l < \theta_k) = 0$. When using truncated travel time distributions, deterministic dominance can be established by simply comparing the minimum and maximum possible travel times, as is demonstrated by the range criterion and range algorithm by Miller-Hooks and Mahmassani (1998b). If $\theta_l^{max} \leq \theta_k^{min}$, it follows that $\theta_l < \theta_k$ for $\theta_l \in [\theta_l^{min}, \theta_l^{max})$ and $\theta_k \in (\theta_k^{min}, \theta_k^{max}]$, resulting in $P(\theta_l < \theta_k) = 0$.

This criterion can be used to allow for the generation of eligible paths and the elimination of paths at intermediate nodes when there is sufficient knowledge that they have no or low likelihood of being non-dominated or optimal. Two measures are considered here: a relaxed deterministic dominance (RDD) for stochastic comparison at varying risk levels and an adjusted first-order stochastic dominance (A-FSD) to be applied at intermediate nodes.

Assuming that the link travel times are truncated random variables with absolute time-invariant minimum and maximum travel time denoted θ_{ij}^{min} and θ_{ij}^{max} $\forall (i,j) \in A$. Then for a path k, $\theta_k^{min} = \sum_{ij \in k} \theta_{ij}^{min}$ and $\theta_k^{max} = \sum_{ij \in k} \theta_{ij}^{max}$ are bounds (though potentially loose bounds) on the path travel time. Then for a given origin-destination pair $i - s$, least time path finding can be performed with static deterministic costs set to θ_{ij}^{min}, resulting in the minimum possible travel time for any path from node i to destination s. Then for any node $i \in N \backslash \{s\}$, the absolute minimum travel time from i to s can be defined in this manner, denoted τ_i^{min}. The following proposition can be established:

Proposition 6-1: Given the definition of the STV network modeling and notation from Section 2.1. with truncated random travel times, consider a path k_{Oi} from origin node O to the intermediate node i with a travel time r.v. θ_k . Suppose the absolute minimum travel time from node i to destination D is τ_i^{min}. Then for any path l_{OD} from O to destination D via the sub-path k_{Oi}, the travel time random variable θ_l will be bounded as follows: $\theta_l \geq \theta_k + \tau_i^{min}$.

Proof 6-1. Let l_{OD} be path from O to D via k_{Oi} be decomposed into two sub-paths k_{Oi} and k'_{iD} from O to i and i to D, respectively. By definition, τ_i^{min} is a lower bound on travel times for all paths from i to D and as such $\theta_{k'} \geq \tau_i^{min}$. Then, it follows $\theta_l = \theta_k + \theta_{k'} \geq \theta_k + \tau_{k'}^{min}$.

6.3.1.1 Relaxed Deterministic Dominance (RDD) for Intermediate Nodes

As introduced above, for two paths l and k with their travel times starting at departure time t denoted θ_l^t and θ_k^t, deterministic path dominance is evaluated by testing if $P(\theta_l^t < \theta_k^t) = 0$. Then, from Proposition 6-1 follows the corollary stated below.

Corollary 6-1. Consider a path l' from O to D and its travel time for departure time t denoted $\theta_{l'}^{t_0}$. Let k be a path from O to intermediate node i, with travel time $\theta_{k_j}^t$ for departure time t, and l be a path from O to D via k with travel time θ_l^t. We know that $\theta_l^t \geq \theta_k^t + \tau_j^{min,D}$, then to evaluate $P(\theta_l^t \leq \theta_{l'}^t)$ an upper bound can be derived as follows, $P(\theta_l^t \leq \theta_{l'}^t) = P(\theta_k^t + \tau_j^{min,D} \leq \theta_l^t \leq \theta_{l'}^t) \leq P(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0})$.

The special case example for deterministic dominance is simpler. Namely, if $P(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}) = 0$ and $\theta_l^t \geq \theta_k^t + \tau_j^{min,D}$, then $P(\theta_l^t \leq \theta_{l'}^t) = 0$.

The relaxation of the deterministic dominance is introduced here as an approximate criterion. To improve the computational effort needed to produce for the generation of eligible paths, the deterministic criterion $P\left(\theta_l^t \leq \theta_{l'}^t\right) = 0$ is modified to allow for paths to be eliminated if they have a very low likelihood of being optimal $P\left(\theta_l^t \leq \theta_{l'}^t\right) \leq \epsilon$ where ϵ is some admissible risk to eliminating a potentially optimal path.

Then at an intermediate node i, from Corollary 6-1: if $P\left(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}\right) \leq \epsilon$, then $P\left(\theta_l^{t_0} \leq \theta_{l'}^{t_0}\right) \leq P\left(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}\right)$ implies $P\left(\theta_l^{t_0} \leq \theta_{l'}^{t_0}\right) \leq \epsilon$. Therefore, $P\left(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}\right) \leq \epsilon$ can be used as a criterion for early elimination at an intermediate node i by showing that even the best-case paths to the destination via subpath k will not meet the eligibility criterion. This probabilistic criterion based on the risk-level tolerance value ϵ is a heuristic criterion allowing for certain paths to be eliminated at intermediate nodes. The higher the risk tolerance value ϵ, the larger the number of eliminated paths (i.e., ineligible for the given risk tolerance), leading to reduced computational effort for the path finding, with the trade-off of potentially resulting in sub-optimal a priori path solutions.

6.3.1.2 Adjusted First-Order Stochastic Dominance (FSD) for Intermediate Nodes

Proposition 6-1 can also be used for the application of the FSD criterion at intermediate nodes in the network. However, it should be noted that while paths are computed from a single origin to all destinations, the eligibility of sub-paths at an intermediate node is determined for each destination. That is, a subpath k_{Oi} from O to intermediate node i will be tested for eligibility for all destination nodes $j \in N \backslash \{O\}$ and may be deemed eligible for some or all destination nodes. The first-order stochastic dominance can then be applied to subpaths from O to i with the absolute minimum lower bound for paths from i to D for each possible destination $D \in N \backslash \{O\}$. The criterion, including multiple departure times, is stated bellow.

Let k_{Oi} be a path from O to i with a travel time random variable θ_k^t for departure time t. For some $j \in N\backslash\{O, i\}$ construct the adjusted random variable $\theta_{k_j}^{t,min} = \theta_k^t + \tau_{i,j}^{min}$ with distribution function $F_{k_j}^{t,min}$. By Proposition 6-1, $\theta_{k_j}^{t,min} \leq \theta_l^t$ for any path l_{Oj} from O to j at departure time t via k_{Oi} (with distribution function F_l^t) and can be tested against any eligible path l' from O to j with travel time $\theta_{l'}^t$ for departure time t. If the distribution function of $\theta_{l'}^t$, denoted $F_{l'}^t$, stochastically dominates $F_{k_j}^{t,min}$, then the A-FSD criterion assumes it will stochastically dominate F_l^t. Namely, if $F_{l'}^t \succ_{FSD} F_{k_j}^{t,min}$, we say $F_{l'}^t \succ_{A-FSD} F_l^t$ for any path l_{Oj} from O to j at departure time t via k_{Oi}. Then all such l_{Oj} are A-FSD-dominated for departure time t and k_{Oi} is deemed ineligible for departure time t to destination node j.

6.3.2 Time-Dependent Reliable Least-Time Paths (RLTP) Algorithm

The solution method presented here considers he TD-RLTP problem where probability distributions can be considered in their continuous or temporally discrete form. The TD-RLTP algorithm is based on general path generation approach by Miller-Hooks and Mahmassani (Miller-Hooks and Mahmassani, 1998b), their LC algorithms for the LET problem (2000b) and the SPOTAR problem by Nie and Wu (2009a). This methodology directly builds on the study by Filipovska and Mahmsaani (2020b), but introduces new tests for path dominance and eligibility via the two path comparison tests presented in the previous section and also applies the methodology for multiple reliability-based least-time objective functions.

For a set of departure times within an established peak period, the procedure generates the eligible paths from a given origin node O to all possible destination nodes in the network or a specified subset of destination nodes. While the path search and path travel time distribution estimation are performed once, a path's eligibility is checked for each departure time and each destination node.

6.3.2.1 TD-RLTP Algorithm

Let the network and notation be defined according to the notation in section 6.2.1, which is here extended to incorporate the notation used in Chapter 5. Let the set $\mathcal{T} = \{\mathcal{t}_0, \mathcal{t}_1, \dots, \mathcal{t}_{L-1}\}$ be the set of time L intervals for the variation of travel time distributions where and $\mathcal{t}_l$ refers to the time interval $[\mathcal{t}_l, \mathcal{t}_{l+1})\ \forall\ l \in \{1, 2, \dots, L-1\}$. Let departure times at the origin be $\mathbb{t} \in \mathbb{T}$, so that the time period covered by $\mathcal{T}$, namely the interval $[\mathcal{t}_0, \mathcal{t}_L)$ must contain all of the departure times $\mathbb{t} \in \mathbb{T}$ and accommodate for the latest possible arrival times for all $\mathbb{t} \in \mathbb{T}$. In the estimation approaches in Chapter 5, departure times were seen as a random variable that should have the distribution of departure times realized in the network. However, departure times at the origin in the path finding problem are from a pre-specified set, and a path is to be found for each departure time. Let the function $\phi(\cdot)$ be defined so that any time t such that $\mathcal{t}_0 \le t < \mathcal{t}_L$ can be converted to the corresponding time interval $\mathcal{t}_l \in \mathcal{T}$ via the function $\phi(t) = \mathcal{t}$.

Path Generation Algorithm for Reliable Least-Time Paths (RLTP) Problem

Given:

The network $G(N, A, \mathcal{T})$, where $\mathcal{T}$ is the set of time-bins $\{\mathcal{t}_0, \mathcal{t}_1, \dots, \mathcal{t}_{L-1}\}$.
The function $\phi(\cdot) \in \mathcal{T}$.
An origin node O.
The set of departure times $\mathbb{t} \in \mathbb{T}$.
A set of destination nodes $\mathcal{D}$, which if unspecified is set to $\mathcal{D} = N \backslash \{O\}$.
The dominance criterion to be used (RDD with risk -tolerance parameter ϵ or A-FSD).
The value of the sample size S to be used for distribution simulation.

Find:

The path eligibility indicators $\Lambda_{id}^{k\mathbb{t}} = 0,\ i \in N, d \in \mathcal{D}, \mathbb{t} \in \mathbb{T},\ k \in \{1, 2, \dots, M\}$ indicating which paths k at intermediate node i are eligible to destination d for departure time $\mathbb{T}$ at the origin.

The travel time distributions for the set of eligible paths $k \in \{1, 2, \dots, M\}$ from the origin node O and departure times $\mathbb{t} \in \mathbb{T}$ to each of the destinations $d \in \mathcal{D}$, contained in the vector-label $U_d^{k\mathbb{t}}$.
The corresponding vector pointers $p_i^k, L_i^k\ \forall\ k \in \{1, 2, \dots, M\}, i \in N$ which can be used to trace back the path for each node.

Step 0: Initialization

Define M large enough to contain as many potentially eligible path identifiers as might be required.
Initiate the vector pointers $p_i^k = \infty, L_i^k = \infty\ \forall\ k \in \{1, 2, \dots, M\}, i \in N$, the node-path pair eligibility vector indicators $q_i^k(\mathbb{t}, d) = 0\ \forall\ k \in \{1, 2, \dots, M\}$ and the vector link labels $u_i^{k\mathbb{t}} = [None]_S$ of size S, $\forall\ i \in N, \mathbb{t} \in \mathbb{T}, d \in D$.
Define the vector label of path travel time distributions $U_d^{k\mathbb{t}} = [None]_S\ \forall\ d \in D, k \in \{1, 2, \dots, M\}, t_o \in T_0$.
Initiate a scan-eligible FIFO list SE = $\{\emptyset\}$.
For the origin node $O \in N$, define path 1_{OO} and let $L_O^1 = 0, p_O^1 = 0$.
Add the node-path ID pair to the SE list.

Step 0.1: Definition of absolute minimum travel time labels

Initialize the minimum extension labels as $\tau_{id}^{min} = \infty, \forall\ i \in N, d \in D$, and set $\tau_{id}^{min} = 0$ if $i = d$.
Perform a static, deterministic path search on the network $G(N, A)$ with link costs $\theta_{ij}^{min} \forall (i, j) \in A$.
Save the minimum travel times from each $i \in N$ to each $d \in D$ as τ_{id}^{min}.

Step 0.2: Initialization of distributions

Take the node-path-ID pair $(O - 1)$ from the SE list.
Find the set of outgoing links from the node O, i.e., $A_O = \{(O, j) \in A\}$.
<u>For</u> each departure time $\mathbb{t}$:
  Jointly obtain a sample $t_{Oj}^{\mathbb{t}}$ of size S for the random variables $\theta_{Oj}^{\mathbb{t}}$ distributed according to $\pi_{Oj}^{\mathbb{t}}$
  <u>For</u> each link $a = (O, j) \in A_o$:
    Save the sampled values in $u_j^{1\mathbb{t}} = t_{Oj}^{\mathbb{t}}$, maintaining the sample order. Update the pointers $L_j^1 = O, p_j^1 = 1$. Set $U_d^{1\mathbb{t}} = t_{Oj}^{\mathbb{t}}$.
    Set $q_j^1(\mathbb{t}, d) = 1\ \forall\ d \in D$.
    Add the node-path ID pair $(j - 1)$ to the SE list.
  <u>End for.</u>
<u>End for.</u>

Step 1: SE list scan

If the SE list is not empty, select the first node-path ID pair $(i - \mu)$ from the front of the queue in a FIFO manner. Go to Step 2.
Otherwise, if SE is empty, terminate.

Step 2: Conditional Sampling

Determine the subset $\mathbb{T}_\mu = \{\mathbb{t} \in \mathbb{T} | u_i^{\mu\mathbb{t}} \neq \emptyset\}$ i.e., departure times for which the path with identifier μ to node i is eligible for at least one destination node.
Trace back the path P_μ from node-path ID pair $i - \mu$.
Determine the set of next possible links $A_i = \{(i,j) \in A, j \notin P_\mu\}$:
Find their path ID as $\mu_j = \min\{k \leq M | L_j^k = \infty\}$.
For each departure time $\mathbb{t} \in \mathbb{T}_\mu$:
Determine the subset of destinations for which the path with identifier μ to node i is eligible, $\mathcal{D}_\mu = \{d \in \mathcal{D} | \Lambda_{jd}^{\mu_j\mathbb{t}}=1\}$
For each sample $s \in \{1, 2, \dots, S\}$:
Find the time intervals of departure for each link in P_μ via $t_{i'} = \phi\left(\mathbb{t} + \sum u_{i'}^{\mu\mathbb{t}}\right)$, extract the appropriate time-dependent joint link travel time distributions and covariance matrix using the exit bins, and jointly sample from $\pi_{a_j}^{t_j}$, where $a_j = (i,j) \in A_i$ conditional on the previous link travel times $u_{i'}^{\mu_{i'}\mathbb{t}}[s] \; \forall \; i' \in P_\mu$.
Save the samples into the temporary labels $u_j'^{\mathbb{t}}[s] \; \forall \; (i,j) \in A_j$.
End for.
End for.

Step 3: Path Comparisons

Call the **Path Comparison Procedure** to obtain the indicators $\Lambda_{\mathrm{j}}^{\mu_j}(\mathbb{t}, d)$, updated vector pointers $p_j^{\mu_j}, L_j^{\mu_j}$, vector link labels $u_j^{\mu_j\mathbb{t}}$ and $U_j^{\mu_j\mathbb{t}}$ if $j \in D$ and SE list.
Go to Step 1.

The path generation algorithm presented above terminates after having determined all eligible paths to all destination nodes from the origin, for all given departure times. Once it has been completed, the optimal a priori paths for each destination, departure time pair $d \in D, \mathbb{t} \in \mathbb{T}$ can be determined by computing the objective function value for the distribution saved in $U_d^{k\mathbb{t}} \; \forall \; k \in \{k \in \{1,2,\dots,M\} | \Lambda_{dd}^{k\mathbb{t}} = 1\}$, then selecting the k with the minimum objective function value.

It is evident that since this evaluation is performed subsequent to the eligible path generation procedure, the eligible paths are the same across all objectives. Thus, this algorithm is not objective-function specific, but its performance for different types of objectives may vary.

6.3.2.2 Path Eligibility Test

The path comparison procedure that updates the pointers, identifiers and labels is given below. This procedure performs a comparison of the newly identified path and determines its eligibility (i.e., whether it is dominated or not), according to the specified dominance criterion (RDD with $\epsilon \geq 0$ or A-FSD from section 6.3.1.

Path Comparisons Procedure

Given:

The set of relevant departure times $\mathbb{T}_\mu$ and destination nodes $\mathcal{D}_\mu$.
Last node and path identifier $(i - \mu)$. Next node and path identifier $(j - \mu)$.
The path P_μ from origin O to i with identifier μ.
The indicators Λ, vector pointers p, L, the vector link labels u and the vector label of path travel time distributions U if $j \in D$. The temporary labels u_j'.
The type of path dominance criterion: RDD (ϵ) or A-FSD.
Current SE list.

Find:

Indicators $\Lambda_j^{\mu_j}(\mathbb{t}, d)$, updated vector pointers $p_j^{\mu_j}, L_j^{\mu_j}$, the vector link labels $u_j^{\mu_j \mathbb{t}}$ and the vector label of path travel time distributions $U_j^{\mu_j \mathbb{t}}$ if $j \in D$. Updated SE list.

Procedure:

For each $\mathbb{t} \in \mathbb{T}_\mu$:
- Recover link travel time samples and determine the path travel time samples as the sum $\tau_{\mu_j}[s] = \left[\sum_{i' \in P_\mu} u_{i'}^{\mu_{i'} \mathbb{t}}[s]\right] + u_j'^{\mathbb{t}}[s] \; \forall \, s \in \{1, 2, \dots, S\}$. Set $z = 1$.
- For each $d \in \mathcal{D}^* = \mathcal{D}_\mu \backslash \{i\}$:
  - Determine the lower bound on the distribution of extensions of P_μ via j to d as $\tau^* = \tau_{\mu_j} + \tau_{jd}^{min}$.
  - Find the identifiers $K = \{k | \Lambda_{dd}^{k\mathbb{t}} = 1\}$ of all complete eligible paths to destination d for departure time $\mathbb{t}$.
  - For each path $k \in K$:
    - Set $c = 0$:
    - If the dominance criterion is RDD (ϵ):
      - Compute the probability $p = P\left(\tau^* \leq U_d^{k\mathbb{t}}\right)$.
      - If $p \geq \epsilon$, set $c = 1$.
    - Otherwise, if $\tau^* \succ_{FSD} U_d^{k\mathbb{t}}$ set $c = 1$.
    - If $c = 1$:
      - If $j = d$, set $U_d^{\mu' t_p} = \tau_{\mu_j}$.
      - Set $\Lambda_{jd}^{\mu_j \mathbb{t}} = 1$ and $z = z + 1$.

End for.
If $z \geq 1$:
Set $L_j^{\mu_j} = i, p_j^{\mu_j} = \mu$. Save $u_j^{\mu_j \mathbb{t}} = u_j'^{\mathbb{t}}$.
Add the node-path ID pair $(j - \mu')$ to the SE list.
End for.
End for.

6.3.3 A Note on the Estimation of Path Travel Time Distributions

An important problem in implementing path finding solutions in stochastic dynamic networks is the estimation of travel time distributions along paths or sub-paths with spatio-temporal dependencies. The approach presented here assumes known marginal link travel time distributions, time-dependent correlations between link travel times and the ability to conditionally sample based on that information. Thus, the algorithm can be applied for any case where those criteria are satisfied.

To make the approach applicable in general cases, not restricted to the distribution form for the marginal distributions or that link travel times have the same distribution form, utilizing the methods presented in previously in Chapter 5 is recommended. In this chapter, the Normal-to-anything (NORTA) approach time time-dependence and time-varying correlations is used when applying the path finding approach.

For the implementation of these estimation methods in path finding, especially for the cases with time-dependent correlation structure, it may be useful to pre-compute the distribution characteristics necessary for sampling. Specifically, the NORTA covariance structures can be used as an input to the solution approach to avoid extended computational run times associated with computing the time-varying correlation structure within path finding and potentially in an inefficient manner.

6.4 Numerical Experiments and Results

This section presents the numerical experiments designed to evaluate the performance of the TD-RLTP algorithm. The network and data used for the numerical experiments are presented in section 6.4.1, the design of the experiments is outlined in section 6.4.2, and the results are presented and analyzed in section 6.4.3.

6.4.1 Network and Data for Numerical Experiments

The numerical experiments for this study used the large-scale Chicago network of 1,578 nodes and 4,805 links, which was previously shown in Chapter 5. The data used for these experiments were obtained from simulations performed using a mesoscopic simulator, DYNASMART-P (Mahmassani et al., 2004). To obtain a data set with variability in link travel times, simulations were performed with varying demand levels and weather conditions, resulting in a total of 25 scenarios based on real-world observations. From the performed simulations, vehicle trajectory data were extracted for the morning peak period between 7:00 and 10:00 a.m. for the entire network. The numerical experiments used a randomly selected origin node and all destination nodes in the network, with a range of departure times in the early portion of the morning peak period. The departure times were considered every five minutes for the first 20 minutes of the peak period, i.e., at 7:00, 7:05, 7:10, 7:15, and 7:20 a.m.

6.4.2 Design of the Experiments

The numerical experiments were designed so as to evaluate the performance of the algorithm across a few key aspects that can be seen as parameters of the algorithm.

Firstly, the numerical experiments were designed to test the performance of the solution algorithm with the different path dominance criteria it was designed to use: the Adjusted First-Order Stochastic Dominance (A-FSD) criterion for intermediate nodes (section 6.3.1.2) and the Relaxed Deterministic Dominance (RDD) criterion for intermediate nodes (section 6.3.1.1). The RDD criterion includes an adjustable risk-tolerance parameter ϵ indicating the strength of dominance required for path dominance. The TD-RLTP algorithm was tested with the A-FSD criterion and with the RDD criterion with 6 different values for ϵ, $\epsilon \in \{0, 0.01, 0.05, 0.1, 0.15, 0.2\}$. It should be noted that the RDD criterion with $\epsilon = 0$ is equivalent to the deterministic dominance criterion introduced by Miller-Hooks and Mahmassani (Miller-Hooks and Mahmassani, 1998b), adjusted to be applied at intermediate nodes, which is a very strong criterion that is exact, even in the case with dependencies, but can be rather computationally expensive.

Secondly, the numerical experiments were designed to test the performance of the solution algorithm, along with the various dominance criteria, across different reliability-based least time objective functions. The algorithm and each of the dominance criteria were tested for a total 6 different objective functions of two types: Value at Risk (VaR) and Conditional Value at Risk (CVaR) based objectives, as described in section 6.2.2. Three values for α were considered for each of the objective types: $\alpha \in \{0.7, 0.8, 0.9\}$.

Thus, the performed numerical experiments performed the path generation for a randomly selected origin node O to all 1577 destination nodes $N \backslash \{O\}$, for 5 departure times, and using 7 different dominance criteria (i.e., the A-FSD and RDD with 6 values for ϵ), resulting in a total of 55,195 cases. Then the best a priori paths and their objective values were found according the 6 objective functions, for each departure time, destination, and dominance criterion, thus leading to a total of 331,170 resulting path solutions and objective function values.

As mentioned above, these numerical experiments apply the path finding approach with the time dependent NORTA approach for path travel time distribution estimation, described in the previous chapter. Additionally, the time dependent NORTA covariance structure was precomputed and used as an input to the solution approach to avoid extended computational run times and to allow for the evaluation of this approach independently of the computational effort required for the NORTA covariance computation.

Finally, to test for the importance of considering correlations, path finding was performed with path travel time distribution estimation correlation without correlation. For this portion of the experiments, path solutions were found by ignoring the presence of correlations for each of the 6 objective functions and all destination nodes. The objective function values for the optimal paths were then re-computed to account for correlations to be compared to the exact a priori path finding approach with the RDD ($\epsilon = 0$) criterion.

6.4.2.1 Research Questions and Performance Measures

In evaluating the results from the numerical experiments, the following research questions were considered:

- How does the accuracy of the path generation approaches change when considering the different dominance criteria for the TD-RLTP algorithm?
    - What is the effect on the number of incorrect paths, i.e., paths selected as optimal different from the exact solution?
    - What is the effect on the objective function value of selected path?
    - How do these values vary across the different objective functions?

- How does the computational effort change when considering the different dominance criteria for the TD-RLTP algorithm?
    - What is the effect on the computational run time?
    - What is the effect on the number of paths generated to each destination node?
- How do path finding solutions discounting correlations compare to those with correlations?
    - How many of the selected paths were suboptimal, i.e., different from the exact solution with correlations?
    - What is the effect on the objective function value of selected path?

Therefore, the performance measures considered here include the raw values for the computational run time in seconds, as well as the maximum and average number of paths per node demonstrating the computational effort for the TD-RLTP algorithm with different dominance criteria. Additionally, the percent change in computational run time relative to the exact solution case, i.e., using the deterministic dominance criterion by setting $\epsilon = 0$ in the RDD criterion. In terms of the accuracy for a priori path finding, the raw and average values for the objective functions are considered, and the performance measures include the percentage of incorrectly selected paths i.e., optimal paths different from the exact solution, and the Mean Absolute Percentage Error (MAPE) for the objective function value on those paths. These values are considered averaged across all cases, for each dominance criterion, as well as averages for each of the different objective functions, in order to test for trends or variations between the two objective types and their α-percentile values.

6.4.3 Results and Analysis of the Numerical Experiments

This section presents the summary of the results from the numerical experiments and their analysis in the corresponding sub-sections.

6.4.3.1 TD-RLTP Algorithm Accuracy Results

The results in Table 6-2 include the average objective value achieved by the optimal paths obtained with each dominance criterion and for each of the six different objective functions.

Table 6-2. Average objective function values for all objectives with different dominance criteria

Dominance Criterion		Objective Function					
		VaR, $\alpha = 0.7$	VaR, $\alpha = 0.8$	VaR, $\alpha = 0.9$	CVaR, $\alpha = 0.7$	CVaR, $\alpha = 0.8$	CVaR, $\alpha = 0.9$
RDD	$\epsilon = 0$	25.1370	27.8721	31.7347	32.0661	34.6180	38.8783
	$\epsilon = 0.01$	25.1370	27.8721	31.7347	32.0661	34.6180	38.8783
	$\epsilon = 0.05$	25.1370	27.8721	31.7347	32.0664	34.6197	38.8839
	$\epsilon = 0.1$	25.1370	27.8731	31.7348	32.0693	34.6394	38.9916
	$\epsilon = 0.15$	25.1371	27.8754	31.7488	32.0815	34.6945	39.2074
	$\epsilon = 0.2$	25.1371	27.8765	31.8208	32.1061	34.7480	39.3086
A-FSD		25.1370	27.8722	31.7347	32.0661	34.6180	38.8785

Averaged across the destination nodes and departure times, the objective function values exhibit little variation, showing that on average the impact of the dominance criterion is not significant. Therefore, Table 6-3 considers additional measures of performance relative to the exact case with the RDD $\epsilon = 0$ criterion.

The performance measures considered in Table 6-3 include: the percent of incorrect paths for each dominance criterion across all objective functions, i.e., the percent of optimal paths different from the exact solution, the MAPE for the objective function value for those paths and the average MAPE across all cases, i.e., including those where the MAPE is equal to zero and the optimal path is the same as the exact solution.

Table 6-3. Performance measures for all objectives with different dominance criteria

Dominance Criterion		Percent Incorrect Paths	Objective MAPE for Incorrect Paths	Average MAPE
RDD	ϵ = 0.01	0.0022%	0.09%	0.0000%
	ϵ = 0.05	0.0246%	8.67%	0.0021%
	ϵ = 0.1	0.1590%	12.55%	0.0200%
	ϵ = 0.15	0.4278%	14.92%	0.0638%
	ϵ = 0.2	0.8220%	15.63%	0.1285%
A-FSD		0.0269%	0.24%	0.0001%

These results show that the percent of incorrect paths and their objective value MAPE increase as the value of ϵ increases for RDD criterion but are relatively low for the A-FSD criterion. These results are also presented visually in Figure 6-2.

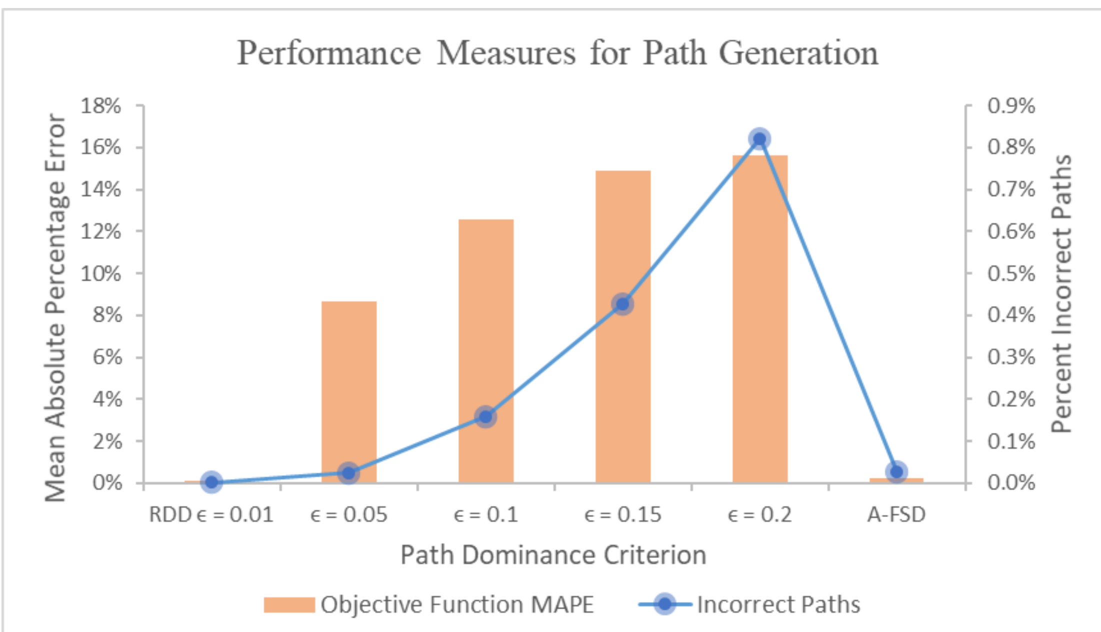

Figure 6-2. Performance measures for path generation with different dominance criteria

From Figure 6-2, it can be observed that both in terms of the number of incorrect paths and the MAPE of objective function values, the RDD with $\epsilon = 0.01$ has the lowest error relative to the exact solution with a total MAPE for the incorrect paths at 0.09% and the A-FSD criterion comes close with 0.24%. However, as the value of ϵ increases for the RDD criterion, the MAPE increases quickly, along with the percent of incorrect paths to maximum values of 15.63% and 0.822%, respectively, achieved at $\epsilon = 0.2$. Therefore, while the number of incorrectly chosen paths remains below 1%, the MAPE for those paths increases up to 15.63% with the increase of ϵ.

To consider whether and how the performance varies with the different objective functions, the percent of incorrect paths and their MAPE are also presented for each of the objective functions in Table 6-4 and Table 6-5, respectively. Several important observations can be made from these results. For both objective types, as the value of α increases the number of incorrect paths increases in most cases, both for the RDD and A-FSD dominance criteria, though this trend does not hold strictly.

Table 6-4. Percent incorrect paths for each objective with different dominance criteria

Dominance Criterion		Objective Function					
		VaR, $\alpha = 0.7$	VaR, $\alpha = 0.8$	VaR, $\alpha = 0.9$	CVaR, $\alpha = 0.7$	CVaR, $\alpha = 0.8$	CVaR, $\alpha = 0.9$
RDD	$\epsilon = 0.01$	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
	$\epsilon = 0.05$	0.00%	0.00%	0.00%	0.01%	0.03%	0.11%
	$\epsilon = 0.1$	0.00%	0.03%	0.01%	0.04%	0.34%	0.54%
	$\epsilon = 0.15$	0.01%	0.07%	0.16%	0.24%	0.87%	1.21%
	$\epsilon = 0.2$	0.03%	0.12%	0.59%	0.59%	1.40%	2.20%
A-FSD		0.00%	0.05%	0.00%	0.00%	0.01%	0.09%

Table 6-5. Objective function value MAPE of incorrect paths for each objective with different dominance criteria

Dominance Criterion		Objective Function					
		VaR, $\alpha = 0.7$	VaR, $\alpha = 0.8$	VaR, $\alpha = 0.9$	CVaR, $\alpha = 0.7$	CVaR, $\alpha = 0.8$	CVaR, $\alpha = 0.9$
RDD	$\epsilon = 0.01$	0.00%	0.00%	0.00%	0.00%	0.00%	0.53%
	$\epsilon = 0.05$	0.00%	0.00%	0.00%	12.52%	21.94%	17.58%
	$\epsilon = 0.1$	0.00%	5.57%	0.76%	30.97%	10.90%	27.13%
	$\epsilon = 0.15$	0.78%	8.98%	20.17%	14.16%	12.76%	32.68%
	$\epsilon = 0.2$	0.49%	6.95%	31.27%	14.34%	15.00%	25.72%
A-FSD		0.00%	0.43%	0.00%	0.00%	0.37%	0.67%

Considering the RDD criteria with $\epsilon > 0$, for lower α values for both the VaR and CVaR objectives, the percent of incorrect paths and MAPE become significant at higher ϵ values. Thus, the highest percent of incorrect paths occur for the Var and CVaR objectives at the highest α value considered here $\alpha = 0.9$ and the highest ϵ value, $\epsilon = 0.2$. Additional observations can be made by considering these results visually, as shown in Figure 6-3. It can be observed that the percent of incorrect paths is significantly higher when for the CVaR objective, for all considered values of α, and especially so when using the RDD criterion with larger ϵ values. The A-FSD criterion has a low percent of incorrect paths across all objectives and does not show a significant difference between the CVaR and VaR objectives as was the case for the RDD criteria, demonstrating that it may be robust with regards to the objective type and α value.

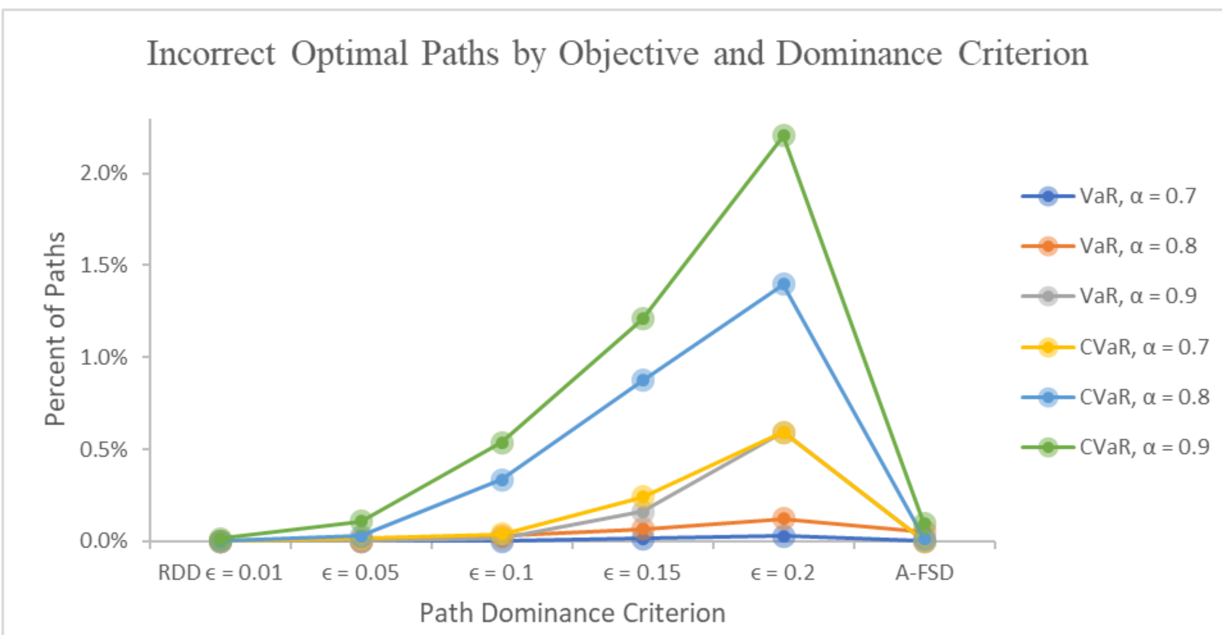

Figure 6-3. Percent incorrect paths for each objective and dominance criterion

6.4.3.2 TD-RLTP Algorithm Computational Effort Results

The computational effort results are shown in Table 6-6, including the computational run time, and the average and maximum number of generated eligible paths for each of the seven dominance criteria. The computational run time presented here is the total for all destination nodes and all departure times, while the number of eligible paths is considered for each case.

Table 6-6. Computational effort for path generation with different dominance criteria

Dominance Criterion		Computational Run Time	Average Number of Eligible Paths	Maximum Number of Eligible Paths
RDD	$\epsilon = 0$	4384.78	11.72	67
	$\epsilon = 0.01$	2433.28	6.49	57
	$\epsilon = 0.05$	1890.35	5.04	53
	$\epsilon = 0.1$	1572.11	4.20	37
	$\epsilon = 0.15$	1331.79	3.55	34
	$\epsilon = 0.2$	1132.02	3.03	30
A-FSD		947.14	2.53	19

These results are also depicted graphically in Figure 6-4, where it can be seen that all three values: the computational run time, and the average and maximum number of paths all decrease as the ϵ value increases for the RDD criterion, but the lowest values are achieved by the A-FSD criterion, which shows lower computational effort compared to all cases of $\epsilon \in \{0, 0.01, 0.05, 0.1, 0.15, 0.2\}$. Considering the computational run times relative to the exact case of RDD with $\epsilon = 0$, the run decreases from 44.5% going to RDD with $\epsilon = 0.01$ to 74.2% with $\epsilon = 0.2$. The largest time savings of 78.4% occur when using the A-FSD criterion.

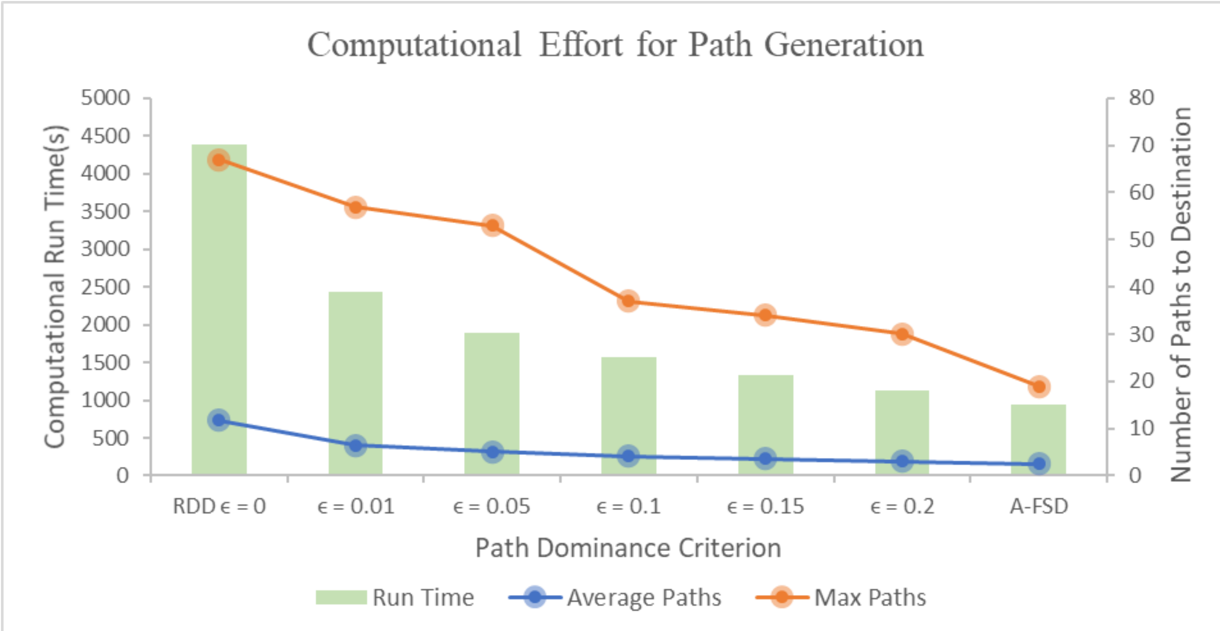

Figure 6-4. Computational effort for path generation with different dominance criteria

These results are especially interesting considering the accuracy results presented previously. The A-FSD criterion was shown to be one of the best performing in terms of error values relative to the exact solution, second only to the RDD criterion with $\epsilon = 0.01$. However, given that the computational run times for the RDD with $\epsilon = 0.01$ are over double those for the A-FSD criterion, the A-FSD criterion offers a good compromise in the trade-off between accuracy and computational effort.

6.4.3.3 Impact of correlations in TD-RLTP solutions

To evaluate the impact of correlations and the importance of accounting for correlations in path finding approaches, the path solutions without correlations are compared to the exact solution of RDD with $\epsilon = 0$. Similar to the results in section 6.4.3.1, the percent incorrect paths, the objective value MAPE for those incorrect paths and on average were considered, across all solutions and for each of the 6 objective functions. These results are shown in Table 6-7, where the VaR and CVaR objectives are abbreviated as V and C, respectively.

Table 6-7. Performance of Solutions without Correlations

	ALL Cases	Objective function					
		$V(\alpha = 0.7)$	$V(\alpha = 0.8)$	$V(\alpha = 0.9)$	$C(\alpha = 0.7)$	$C(\alpha = 0.8)$	$C(\alpha = 0.9)$
Percent Incorrect Paths	39.30%	25.98%	42.65%	44.46%	42.06%	43.33%	46.14%
Objective MAPE for Incorrect Paths	16.52%	14.47%	15.38%	17.54%	15.94%	16.99%	18.60%
Average MAPE	6.49%	3.76%	6.56%	7.80%	6.70%	7.36%	8.58%

These results show that across all cases the percent of incorrect paths was 39.3%, with objective MAPE for those paths on average at 16.52%. Furthermore, the percent incorrect paths and the MAPE for those paths vary with the chosen objective function, increasing with the value of α for both objective types, and showing higher values for the CVaR objectives relative to the VaR objectives at a given value for α. For the CVaR objective function with $\alpha = 0.9$, the percent of incorrect paths was highest at 46.14% and the MAPE for those paths on average at 18.6%. These values are significant, showing that anywhere from 25.98% to 46.14% of the solution paths were incorrect when discounting the correlations and with significantly large error values. These results demonstrate the importance of accounting for correlations in finding optimal a priori paths in stochastic dynamic networks, by quantifying the effect of discounting the presence of correlations. They further demonstrate that for applications with greater risk sensitivity, where the VaR and CVaR objective functions with larger values of α are used, the importance of accounting for correlations is greater and the effect of discounting correlations is even more significant.

6.5 Conclusion

This chapter presents an approach for solving the time-dependent reliable least-time path (RLTP) problem on STV networks with link travel time correlations via a path generation approach. The approach can be applied with a range of path dominance criteria, and this chapter presents an Adjusted First-Order Stochastic Dominance (A-FSD) criterion and a Relaxed Deterministic Dominance (RDD) criterion with an adjustable risk-tolerance level. The path generation approach with the A-FSD and RDD criteria is intended to be used for a priori path finding under reliability-based least-time objectives in stochastic networks. Two types of objectives are presented in this chapter, based on investment risk measures Value at Risk (VaR) and Conditional Value at Risk (CVaR) that can be applied with an adjustable confidence level α.

Numerical experiments were performed to evaluate the solution approach and the different dominance criteria. The dominance criteria were compared in terms of computational effort and error measures relative to an exact solution. The approach was also tested for the two types of objective functions, each applied with three confidence levels α. The numerical experiments show the applicability of the solution approach across different objective functions, the impact of the adjustable risk-tolerance level for the RDD criterion and the compromise achieved by the A-FSD criterion in terms of the trade-off between computational effort and accuracy of the solution. Future chapters will consider the application of this approach and the different dominance criteria for finding adaptive routing strategies.

Chapter 7 Trajectory-Adaptive Reliable Least-Time Routing Strategies[3]

7.1 Overview

This chapter focuses on the problem of finding optimal trajectory-adaptive routing strategies in stochastic dynamic networks with reliability-based objectives, or trajectory-adaptive reliable least-time strategies (TA-RLTS). The strategy finding problem in stochastic dynamic networks is approached under the assumption of complete spatio-temporal link travel time dependencies that can be modeled via joint time-varying travel time distributions with time-varying correlation structures. In stochastic dynamic networks with spatio-temporal correlations, the transportation network is modeled as a system in which travel times experienced in one part of the network at a future time can be dependent on travel times experiences in other parts of the network at earlier times. Information-adaptive routing problems are defined by two key factors that determine the availability of information and how it will affect a traveler's choices.

This chapter is concerned with a special case of partial information availability: the knowledge of the traveler's own trajectory while traveling, i.e., where they have been and at what time. A few previous studies have considered trajectory-adaptive routing problems (Huang and Gao, 2018; Opasanon and Miller-Hooks, 2006; Pretolani et al., 2009). Adaptivity to the traveler's own trajectory information is an appropriate assumption under decentralized routing systems which may have access to historical information for any a priori knowledge, but each user can only utilize the current trip information for en-route decisions.

[3] This chapter builds on article under review by Filipovska and Mahmassani.

In addition to the type of available information, the type of response a traveler can have to that information is also an important defining aspect of the problem. A traveler may choose to ignore any information they have access to, to react to it, or to be proactive (i.e., strategic) in response to the information. Reactive travelers consider the information as it arrives and make a new decision at each decision point (i.e., each intersection in the network). Proactive (or strategic) travelers make a plan, considering the availability of information at all later decision points, and follow a strategy that dictates their choice at each intersection based on information they will have received when they arrive there. This chapter is concerned with the latter, a traveler proactive to their own trajectory information. Thus, instead of searching for a path, the solution is a collection of paths, i.e., a routing strategy that the traveler will follow.

This chapter utilizes the representation for jointly distributed link travel times across the entire network as continuous random variables with time-varying distributions and correlation structures as presented in Chapter 4; it applies approaches for path travel time distribution estimation with generalized correlations as introduced in Chapter 5; and it utilizes the eligible path generation approach presented in Chapter 6 to introduce a generalized 2-stage optimal strategy finding approach.

The remainder of this chapter is organized as follows. Section 7.2 presents the problem statement and methodological difficulties, and includes specific definitions related to the trajectory-adaptive reliable least time strategy (TA-RLTS) problem in 7.2.3. The solution methodology is presented in section 7.3, with the first stage in 7.3.1 and the second in 7.3.2. Section 7.4 focuses on the numerical experiments with the experimental design in 7.4.1 and the results and analysis in 7.4.2. Conclusions and discussions on future work are presented in section 7.5.

7.2 Problem Statement and Methodological Difficulties

This section defines the problem for the trajectory adaptive reliable least-time strategy (TA-RLTS) problem and the methodological difficulties associated with it. The definition and notation of the stochastic dynamic network with generalized link travel times is the setting for this problem, as presented in section 6.2.1 and extended in section 7.2.1 for the specific problem considered here. Definitions and distinctions of types of adaptive routing in stochastic dynamic networks are presented in section 7.2.2. Specific definitions for the TA-RLTS problem are detailed in section 7.2.3 and the discussion of optimality and path comparisons is extended in section 7.2.4.

7.2.1 Stochastic Time-Varying Network Modeling and Notation

Let an STV network be a directed graph $G(N, A, \mathcal{T})$, where N is the set of $|N| = n$ nodes, A is the set of $|A| = m$ links, and $\mathcal{T}$ is the set of time periods, corresponding to the definitions and notation in Chapter 6. The link travel times are random variables jointly distributed across time, with Θ_{ij}^{t} denoting the travel time on each link (i, j) at time t - a continuous positive random variable with a truncated distribution π_{ij}^{t} constrained by a minimum and maximum possible value. The dependencies between the link travel times are defined via link-pairwise covariances that vary over time-period pairs, so that $cov(\Theta_{ij}^{t_1}, \Theta_{kl}^{t_2})$ is the covariance between the travel time on link (i, j) during time interval $t_1 \in \mathcal{T}$ and that on link (k, l) during time period $t_2 \in \mathcal{T}$.

The challenge of the estimation of path travel time distributions remains as introduced in Chapter 6. This problem setting, with link travel times modeled as time-varying random variables correlated across space and time, requires that each subsequent link's travel time distribution is for the corresponding time-interval t_i, and conditional on the realized travel times on each of the traversed link $\pi_i^{t_i}(\theta_i|\tau_j \forall j < i)$. Thus, the estimation of path travel time distributions is performed using approaches introduced in Chapter 5.

An important network modeling assumption for path finding is the first-in-first-out (FIFO) consistency assumption. Its importance for defining stochastic dynamic networks and path finding problem was introduced in Chapter 6 in the context of a priori path finding. In the context of adaptive routing, optimal strategies may involve cycles. Since the choice of route is based on the information available to the user, an optimal strategy may include a traveler exploring a route with the option of 'resetting' if new information reveals that a different choice at a previous node is sufficiently likely to yield improvements on the traveler's reliability-based objective. Further detail, formal definitions and examples can be found in studies by Polychronopoulos and Tsitsiklis (1996) and Provan (2003).

In this chapter, the solution space for the TA-RLTS problem is constrained to acyclic paths and strategies only. This chapter uses the definition of stochastic FIFO presented by Nie and Wu (2009b) and given previously in section 6.2.1, and the problem is defined under the assumption that cyclicity is precluded as a property of the user.

7.2.2 A priori and Adaptive Routing in STV Networks

The focus of this chapter is on the problem of determining trajectory-adaptive routing strategies and their associated travel time distributions to find the optimal routing strategy based on specified reliability-based optimality criteria. The distinction between a priori and adaptive routing is initially described in section 1.2. This section expands on that distinction via a small example.

In a priori path finding an entire route is selected before the departure at the origin node and no en-route deviations are permitted. Adaptive routing problems for proactive (i.e., strategic) travelers focus on determining a strategy composed of a set of paths with decisions to be made along subsequent nodes. Since link travel times are random variables, their actual travel times become known as they are realized (i.e., the link is traversed), and thus the departure time for each following link is also known only as a random variable.

Time-adaptive routing strategies are based on the idea that a better path could potentially be selected knowing the actual arrival time at an intermediate node. In the context of STV networks with correlated link travel times, in addition to knowing the arrival time, the revealed information of realized travel times on previous links also allows for conditional travel time distributions on any upcoming links given the previously traversed links' travel times.

In a time-adaptive context, strategies for which route to take, given the intermediate information of arrival time at each intermediate node, can be represented by an acyclic subnetwork (i.e. hyperpath) representation. A directed hypergraph model for a time-adaptive problem is given by Pretolani (2000), and later extended for the time-adaptive and history-adaptive multi-criterion routing in STV networks (Pretolani et al., 2009). The latter study uses a definition for history-adaptive routing equivalent to the definition for trajectory-adaptive routing by Huang and Gao (2018), and points out that in a history-adaptive strategy the successor of a node i at time t is not unique and is chosen based on the travel times experienced on previous links (also referred to as the arrival history).

This section demonstrates the effect of adaptive routing relative to a priori path finding and contrasts the different types of adaptive routing via examples. Consider the network $G(N, A)$

shown in Figure 7-1 and suppose the possible travel time realizations of the joint link travel time distribution are as given in Table 7-1, with a fixed departure time at the origin O, $t_0 = 0$.

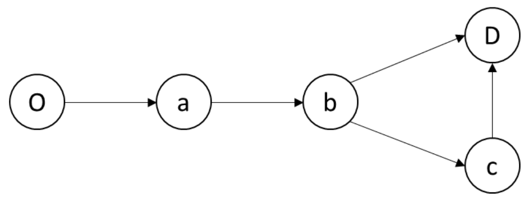

Figure 7-1. Example network 1

Consider the case of a traveler planning a trip from O to D, where one of two paths can be taken: *O-a-b-D* and *O-a-b-c-D*, for which the travel times for each realization of the network are as shown in the corresponding columns in Table 7-1. For simplicity, suppose that each of the given realizations have the same probability of occurrence in the network.

Table 7-1. Possible joint link and path travel time realizations for example network 1

	Link Travel Times					Path Travel Times	
Realization	(O, a)	(a, b)	(b, D)	(b, c)	(c, D)	*O-a-b-D*	*O-a-b-c-D*
1	1	1	3	1	1	5	4
2	1	2	1.5	1	1	4.5	5
3	1	2	1	1	1	4	5
4	2	1	2	0.5	1	5	4.5
5	2	2	2	2	1	6	7
6	2	2	1.5	1	2	5.5	7

Example 1. A Priori vs. Time-Adaptive Routing

In the a priori problem a traveler chooses one of the two possible paths before beginning their trip, based on the entire travel time distributions for the two paths. At the a priori level, the paths *O-a-b-D* and *O-a-b-c-D* have the discrete travel time distributions $\{4, 4.5, 5, 5, 5.5, 6\}$ and $\{4, 4.5, 5, 5, 7, 7\}$, respectively. As such, the distribution of the first path, *O-a-b-D*, dominates that of the second path, *O-a-b-c-D,* via first-order stochastic dominance, and has a lower expected value as well as α-confidence value for $\alpha > 0.7$.

In the time-adaptive problem, the traveler decides on a strategy on which path to travel based on the information of arrival time at the intermediate node where the decision is to be made (in this case node b), referred to as branching node. Therefore, for each of the three possible arrival times at node b, equal to 2, 3 or 4, the user chooses the better sub-path from b to D, based on their objective.

- Arriving at the branching node at time 2, the total path travel time experienced by the user can be 5 or 4 by selecting the sub-path *b-D* or the sub-path *b-c-D,* respectively. Thus, the traveler would select the extension sub-path *b-c-D* with the lower travel time.
- Arriving at the branching node at time 3 (corresponding to realizations 2, 3 and 4), the path travel time distributions that can be experienced by selecting the sub-paths *b-D* or *b-c-D* are {4, 4.5, 5} and {4.5, 5, 5}, respectively, and a user with the objective of minimum expected travel time would select the extension sub-path *b-c-D.*
- Finally, arriving at the branching node at time 4, the traveler again chooses between sub-paths *b-D* and *b-c-D* with travel times {5.5, 6} and {7, 7}, respectively, and would select the former of the two.

This example shows how the a priori solution differs from the time-adaptive solution and the quality of the two solutions can be compared. For the a priori solution, having selected path *O-a-b-D*, the travel time distribution of the solution path is {4, 4.5, 5, 5, 5.5, 6}, again assuming equal probability for each possible realized value. With the time-adaptive strategy the travel time distribution is {4, 4, 4.5, 5, 5.5, 6}, which dominates that of the path *O-a-b-D* via first-order stochastic dominance.

The solution to the a priori problem is a possible solution to the time-adaptive problem, i.e., equivalent to choosing the path *O-a-b-D* for all arrival times at the intermediate node. The solution space for the a priori problem is always a subset of the solution space for the time-adaptive problem. Thus, the optimal solution to the time-adaptive problem will be at least as good as that of the a priori problem, regardless of the objective.

Example 2. Time-adaptive vs. trajectory-adaptive routing

Extending the problem in Example 1, consider the trajectory-adaptive routing problem. Here the traveler has the information of their trajectory, so in addition to choosing the best path for each possible arrival time at the intermediate node, they make a choice for each trajectory (or history) with which they will have arrived at the intermediate node.

Node *b* can be reached at time 3 (corresponding to realizations 2, 3 and 4) with two different 'histories', namely having experienced travel times 1-2 or 2-1 on the previous links *(O, a) - (a, b)*. In the time-adaptive problem, arriving at node *b* at time 3 the traveler would choose node *D* – equivalent to choosing the path *O-a-b-D.* However, in the trajectory-adaptive case, the strategy would be summarized as follows:

- Arriving at node *b* with experienced ravel times 2 and 1 on links *(O, a) - (a, b)*, the conditional travel times on the sub-paths *b-D* and *b-c-D* would be 5 and 4.5 respectively, so the traveler would choose sub-path *b-c-D.*
- Arriving at node *b* with experienced travel times 1-2 on links *(O, a) - (a, b)*, the traveler would choose sub-path *b-D* with travel time distribution {1, 1.5} over the sub-path *b-c-D* with distribution {2, 2}.

This example shows that the solution for the time-adaptive problem is also a possible solution for the trajectory-adaptive problem. However, not being responsive to the trajectory information, the time-adaptive problem has a more restricted solution space, which again indicates that the optimal solution to the trajectory-adaptive problem will be at least as good as that of the time-adaptive problem, regardless of the objective.

7.2.3 Trajectory-Adaptive Reliable Least-Time Strategy Problem

This chapter considers the problem of finding reliable trajectory-adaptive routing strategies in stochastic dynamic networks with spatio-temporally correlated link travel times. The problem definition assumes that a traveler formulates a strategy for their trip to destination, with intermediate node decisions accounting for the potential future decisions given that choice. The strategy becomes realized as a decision is made at each branching node based on information revealed from the traveler's own trajectory.

The trajectory information H is defined as a series of consecutive node-time pairs that the traveler has experienced from the origin node i_0 at their departure time t_0, up to the current node i and time t: $H = \{(i_0, t_0), (i_1, t_1), \dots, (i, t)\}$. By this definition, equivalent to that by Huang and Gao (2018), the trajectory contains the information of the revealed travel times along the traversed links so that the observed travel time on link (i_{k-1}, i_k) departing at time t_{k-1} was $t_k - t_{k-1}$ for each node-time pair (i_k, t_k) for $k \neq 0$. Then, assuming dependence between link travel times across time, the travel time distributions on any future links on the traveler's potential route can be conditioned on the travel times experienced on links they traversed in the path to the current node. A definition for a trajectory-adaptive routing strategy, equivalent to that by Huang and Gao (2018), is presented here:

Definition 1. Trajectory-Adaptive Routing Strategy: A trajectory-adaptive routing strategy ς can be defined as a mapping from state to decision. The state is defined as the triplet $\{i, t, H\}$, where i is the current node, t is the current time – equivalent to the time of arrival at the current node, and H is the current trajectory information, as defined above. The action space at state $\{i, t, H\}$ is $\{j \in N : (i, j) \in A\}$, i.e., the set of nodes adjacent to i, and the decision for which node j to take next: $\varsigma: \{i, t, H\} \rightarrow j$.

The strategy at node i at time t, $\varsigma\big(i, t, H(i, t)\big)$ can be recursively defined as a combination of the next node j and the set of sub-policies exiting node j at possible arrival times with the corresponding resulting trajectory information H, $\{\varsigma_k(j, t_k), H(j, t_k)\}$. For a problem definition with a finite set of possible travel times on any given link, this recursive definition can be implemented exactly. The problem definition in this chapter requires the use of a simulation or estimation approach for the recursive implementation.

Definition 2. Routing Strategy Decision Node: A decision node for a trajectory-adaptive routing strategy ς is a node $i \in N$ such that $|\{j \in N : (i, j) \in A\}| > 1$, namely a node for which there is more than one possible next node that can be chosen.

Definition 3. Routing Strategy Branching Node: A branching node for a trajectory-adaptive routing strategy ς is a decision node $i \in N$ such that $\big|\{\varsigma(i, t, H) = j \;\forall\, \{i, t, H\}\}\big| > 1$, namely a decision node for which there are multiple unique sub-strategies across the different states it can take.

Based on the above definition, a branching node is formed only at those decision nodes that are the origin for more than one sub-strategy and not all nodes or all decision nodes along the trajectory-adaptive strategy are branching nodes. The decision nodes can be known by considering the network itself, but whether or not a decision node will be a branching node depends on the final strategy.

At the initial decision point, if the trip origin node is the first decision node, the user decides based on an evaluation of their entire strategy with respect to their reliability-based objective. The difference between proactive and reactive routing has previously been shown in other related studies, including those by Waller and Ziliaskopoulos (2002) and Gao and Huang (2012). What follows is a short example to illustrate the nature of proactive decision making and distinguish between the reactive traveler and the traveler who is proactive to information.

Given the previous definitions and the informal description of a routing strategy as a collection of paths, a few important definitions remain, to be illustrated via Example 3. Let a routing strategy path be defined as a path from origin $O = i_0$ to destination $D = i_{l+1}$ consisting of the consecutive links $\{(O, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_l, D)\}$ such that $\forall\, i_k \in \{i_0, i_1, .., i_l\}\ \exists\, t_k, H_k$ such that $\varsigma(i_k, t_k, H_k) = i_{k+1}$. Namely, for some arrival time and history combinations, there are sub-strategies from O to D that traverse the path $\{(O, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_l, D)\}$. Note that at the origin node, t_0 is the departure time at origin and $H_0 = \emptyset$ is an empty set, since there is no observed trajectory at the origin. Hence, the origin node cannot be a branching node, but it can be considered a decision node at which a choice is made between more than one strategy.

For any given routing strategy path, the existence of some t_k, H_k such that $\varsigma(i_k, t_k, H_k) = i_{k+1}$ is necessary but there may exist more than one such time and trajectory for each node i_k, so the general notation should be for a set $\mathbb{H}_k = \{\{t_k, H_k\} | \varsigma(i_k, t_k, H_k) = i_{k+1}\}$.

Definition 4. Path's Contribution to a Routing Strategy: For a path from $O = i_0$ to $D = i_{l+1}, \{(O, i_1), (i_1, i_2), \dots, (i_{l-1}, i_l), (i_l, D)\}$ and departure time t_0 with the corresponding sets $\mathbb{H}_k \neq$

$\emptyset\ \forall i_k \in \forall\ i_k \in \{i_0, i_1, .., i_l\}$, the path's contribution to the routing strategy is the probability at the final node, $p = \sum_{\{t_l, H_l\} \in \mathbb{H}_l} p(\{t_l, H_l\} | \mathbb{H}_{l-1}, \dots, \mathbb{H}_1, \mathbb{H}_0\}$. Informally, a path's contribution to a routing strategy is the likelihood that a path is selected if part of the routing strategy.

From this definition it follows that the sum of such probabilities for all routing strategy paths of a given strategy must be equal to 1, simply ensuring that for any realization of events in the network one and exactly one path must be selected. Thus let the set of paths on a strategy be denoted $\mathbb{K}$, where $\mathbb{K} = \left\{\{(O, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_l, D)\} \left| \left\{\{t_k, H_k\} \middle| \varsigma(i_k, t_k, H_k) = i_{k+1}\right\} \neq \emptyset\ \forall\ i_k \in \{i_0, i_1, .., i_l\}\right.\right\}$, i.e., the set of all paths that satisfy the definition for being routing strategy paths. For each path $\Bbbk \in \mathbb{K}$, let its contribution to the strategy, per definition 4 be denoted $p_{\Bbbk}$ and its path travel time random variable $\Theta_{\Bbbk}$ according to the notation in section 7.2.1.

<u>*Definition 5. Routing Strategy's Travel Time Distribution:*</u> A routing strategy's travel time cumulative distribution function (cdf) is determined as a mixture model, i.e., the cdf of the random variable $\Theta_S = \sum_{\Bbbk \in \mathbb{K}} p_{\Bbbk} \Theta_{\Bbbk}$. The strategy's cdf is simply a mixture of the cdfs of the paths it is composed of, with the contribution probability values serving as mixture weights.

Consider a network, expanded from that shown in Figure 7-1, as shown in Figure 7-2 and suppose the possible travel time realizations of the joint link travel time distribution are as given in Table 7-2, with a fixed departure time at the origin O, $t_0 = 0$. Suppose the traveler is planning a trip from O to D, based on the travel times for each realization of the network as shown in Table 7-2, assuming that each of the given realizations have the same probability of occurrence.

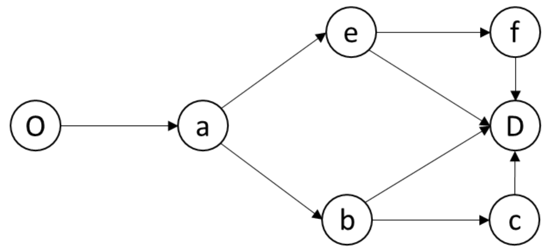

Figure 7-2. Example network 2

Table 7-2. Possible joint link travel time realizations for example network 2

Realization	Link Travel Times								
	(O,a)	(a,b)	(a,e)	(b,D)	(b,c)	(c,D)	(e,D)	(e,f)	(f,D)
1	1	1	0.5	3	1	1	0.5	0.5	1
2	1	2	1	1.5	1	1	1	1	1.5
3	1	2	1.5	1	1	1	0.5	1.5	1
4	2	1	0.5	2	0.5	1	1	1	1
5	2	2	1	2	2	1	0.5	1	1.5
6	2	2	1	1.5	1	2	1	1.5	1

Table 7-3. Possible joint link and path travel time realizations for example network 2

Realization	Link Travel Times						
	(O,a)	(a,b)	(a,e)	*O-a-b-D*	*O-a-b-c-D*	*O-a-e-D*	*O-a-e-f-D*
1	1	1	0.5	5	4	3	3.5
2	1	2	1	4.5	5	5	5.5
3	1	2	1.5	4	5	5	5.5
4	2	1	0.5	5	4.5	4.5	5
5	2	2	1	6	7	5.5	6.5
6	2	2	1	5.5	7	6	6.5

In this example network with 3 decision nodes *a, b, e,* the traveler could have a total of 4 different realized paths, and a decision made at node *a* is a choice for a preferred sub-strategy from that node, and similarly at nodes *b* and *e*.

Example 3. Trajectory-Adaptive routing for a Proactive Traveler

For simplicity suppose that the traveler wants to minimize their expected travel time from O to D. For conciseness, the path travel time distributions with the corresponding realizations from Table 7-2 and the travel times on links before decision points are shown in Table 7-3.

Consider the proactive traveler at the origin node, making a choice for what will be their decision at future node a, based on what they will have experienced on the link O-a. The proactive traveler anticipates the travel times on subsequent links (a-b and a-e) conditional on those experienced for link O-a and their decision at each subsequent node (b and e), based on those possible realizations. In this manner, they build the distribution for each sub-strategy and make a decision at the first decision node a. Note that in this example, the origin O is not a decision node.

If travel time on link O-a is $t_{Oa} = 1$, then:

Considering choosing b the traveler will choose paths

$$\begin{cases} O-a-b-c-D \text{ with expected travel time } 4 & \text{if } t_{ab}=1 \\ O-a-b-D \text{ with expected travel time } 4.25 & \text{if } t_{ab}=2 \end{cases}$$ for an overall expected travel time of 4.167.

Considering choosing e the traveler will choose paths

$$\begin{cases} O-a-e-D \text{ with travel time } 3 & \text{if } t_{ae}=0.5 \\ O-a-e-D \text{ with travel time } 5 & \text{if } t_{ae}=1 \\ O-a-e-D \text{ with travel time } 5 & \text{if } t_{ae}=1.5 \end{cases}$$ for an overall expected travel of 4.33.

Thus, if $t_{Oa} = 1$ the traveler would choose node b as the next node with a better expected travel time. This process can be repeated for the case when travel time on link O-a is $t_{Oa} = 2$. ■

Applying the definitions introduced in this section it can be noted that nodes a, b, e are decision nodes. At node e, the traveler chooses 2 next nodes for the two possible trajectories to a, thus a is a branching node. Node b is also a branching node for the same reason, but at decision node e the same decision is made for all cases, so this node is not a branching node.

Finally, given that the probability of arriving at a with $t_{Oa} = 1$ is 0.5, and *O-a-b-c-D* is chosen for 2 of the 3 realizations given $t_{Oa} = 1$, the contribution of *O-a-b-c-D* to the strategy is $0.5 \cdot \frac{2}{3} \approx 0.333$. For path *O-a-b-D,* the contribution to the strategy is $0.5 \cdot \frac{1}{3} \approx 0.111$. Finally, arriving at a with $t_{Oa} = 2$ with probability 0.5, the path *O-a-e-D* is selected for all realizations given $t_{Oa} = 2$, hence its contribution to the strategy is 0.5.

7.2.4 Optimality and Path Comparisons for Trajectory-Adaptive Routing Strategies

This section presents and discusses some important characteristics of stochastic dynamic networks with generalized correlations. These characteristics are addressed in relation to the problem at hand and provide support for the solution methodology presented in the following section. An important characteristic that applies to the problem in this chapter is the non-applicability of Bellman's principle, which is demonstrated in section 6.2.3.1 for the a priori path finding problem. Since a priori paths are special case solutions to the adaptive routing problem, the counterexample presented in the previous chapter holds and Bellman's principle remains non-applicable in this problem as well.

This section discusses optimality of trajectory-adaptive routing strategies as solutions of the TA-RLTP problem in 7.2.4.1, path comparisons for eligible paths in 7.2.4.2, and for trajectory-adaptive strategy building in 7.2.4.3.

7.2.4.1 Optimality of Reliable Trajectory-Adaptive Routing Strategies

Routing decisions in stochastic dynamic networks can be made based on a variety of reliability or distribution-based objective functions. Similar to the solution approach in Chapter 6, this chapter aims to develop an approach for trajectory-adaptive strategy finding that is not restricted to one single objective function. Example 3 shows that the traveler's routing strategy is determined based on the objective function and future decisions are also anticipated based on that objective. Thus, each strategy and sub-strategy must be built and evaluated for each objective function.

The presented methods, extended here for the case of trajectory-adaptive routing, can be used with a variety of possible least-time objectives, including but not limited to: least expected travel time (LET) (Miller-Hooks and Mahmassani, 2000a), least possible time (Miller-Hooks and Mahmassani, 1998a), least α-confidence travel time (Chen et al., 2018; Zeng et al., 2015), on-time arrival probability (Yang and Zhou, 2017), reliable shortest paths (Chen et al., 2020; Zhang et al., 2017). It should be noted that the proposed approach may perform differently with different types of objectives and is not expected to be applicable for certain types of objectives not centered on least time solutions, such as ones focused on variance.

This chapter considers two types of objectives, introduced for the a priori problem in Chapter 6, the minimum Value at Risk (VaR) and Conditional Value at Risk (CVaR) for varying α-percentile values. The details on those objective functions and their interpretation can be found in section 6.2.2.

7.2.4.2 Path Comparisons for Eligible Paths Generation

Solving the TA-RLTS problem in this chapter is based on the ability to perform path comparisons as a basis for the path generation and strategy building approach, which is evident from Example 3 on the trajectory-adaptive routing for a proactive traveler.

To address the issue of generating eligible paths to be part of a travelers' strategy as a combination of paths, eligible paths can be generated according to the dominance criteria and path generation approach presented in Chapter 6. The path generation approach was presented with two types of dominance criteria, a relaxed deterministic dominance (RDD) criterion with a variable risk-tolerance level that specifies the relaxation level, and an adjusted First-order Stochastic Dominance (A-FSD), both modified to be applied at intermediate nodes. The details on those path dominance criteria can be found in section 6.3.1.

7.2.4.3 Path Comparisons for Trajectory-Adaptive Strategy Building

Considering path comparisons for building a routing strategy, two important methodological difficulties must be addressed.

Firstly, at any branching node, the distributions of non-disjoint paths (i.e., paths that share links in the topological network that are traversed in the same time period) cannot be compared directly, since they are not independent. Discounting the dependence would result in comparing path travel times that cannot occur simultaneously in the network. This problem is addressed in more detail in a paper by Miller-Hooks and Mahmassani on path comparisons in STV networks (2003a). To account for the correlations the distributions of non-disjoint paths must be compared conditionally on the travel times on the shared links at the times that they are traversed.

Secondly, implementing such conditional distribution comparison is made more difficult when considering correlations. Given time-varying distributions and correlations, the conditional distributions should account for the time-interval in which future links will be traversed, which are themselves random variables, and thus there may be more than one possible correlation value that might be realized with different likelihoods. This is a problem of estimation of path travel time distributions, which is the focus of Chapter 5, thus this chapter referrers to the approaches presented in section 5.4.

Additionally, path comparisons at branching nodes may require different criteria to those used for a priori path comparisons. Previous studies have shown that the set of a priori FSD non-dominated paths in STV networks will contain paths that may be inefficient in that, though non-dominated, they would never contribute to an adaptive routing strategy (Miller-Hooks and Mahmassani, 2003a). However, in STV networks with dependencies, the opposite may be true. Namely, there may be paths that are a priori FSD-dominated but would still contribute to the adaptive strategy. Two examples are considered below.

Example 4. FSD Path Comparisons

Consider the example network 1, shown in Figure 7-1. The travel times, shown in Table 1, are for a single departure time $t_0 = 0$, and the FSD-admissible paths can be determined by comparing their travel time distribution. The two possible paths, *O-a-b-D* and *O-a-b-c-D,* have the travel time realizations with equal probabilities {4, 4.5, 5, 5, 5.5, 6} and {4, 4.5, 5, 5, 7, 7}, respectively. Comparing these two paths directly, it can be observed that for all possible total travel times $T_s \in \{4, 4.5, 5, 5.5, 6, 7\}$, the CDF for the path *O-a-b-D* is larger or equal to that of *O-a-b-c-D,* and thus the path *O-a-b-c-D* would be dominated by FSD and not FSD-admissible. However, Example 2 demonstrated that for two cases if the travel times on the initial links *O-a* and *a-b* are realized according to realizations 1 and 4, the selected path would be *O-a-b-c-D,* despite being FSD-dominated. ■

In Example 4, the FSD-dominated path would contribute to the final strategy for 2 of the 6 realizations of the network, equivalent to 33.3% of the cases, which points to a measure that can be used for evaluating the likelihood of a path to contribute the overall strategy. Namely, from the example network 1, it can be observed that the contribution of the path to the overall strategy would be limited by the likelihood with which its value is lower than that of another path comprising the same strategy. This comparison is equivalent to that of the relaxed deterministic dominance criterion with the ϵ value from Chapter 6. It should be noted that this probabilistic comparison of the two paths should be evaluated conditionally on the realized values on the shared links. The example continues, to illustrate the path comparison and likelihood computation.

<u>*Example 5. Probabilistic Path Comparisons*</u>

For this example, let paths *O-a-b-D* and *O-a-b-c-D* be called path 1 and 2, with path travel time random variables denoted $T_1$and T_2, respectively. Let each possible realization of these distributions be denoted $t_1^{(s)}$ and $t_1^{(s)}$ for the realization $s \in \{1, 2, \dots, 6\}$. Evaluating the travel time random variables on the two paths independently, $P(T_2 \leq T_1)$ would be as follows:

$$P(T_2 \leq T_1) = \sum_{k=1}^{6} \frac{1}{6} \cdot P\left(T_2 \leq t_1^{(k)}\right) = \frac{1}{6}\left(\frac{4}{6} + \frac{2}{6} + \frac{1}{6} + \frac{4}{6} + \frac{4}{6} + \frac{4}{6}\right) = \frac{19}{36} = 0.527.$$

However, conditional on the shared links' travel times, the correct comparison would be:

$$P(T_2 \leq T_1 | t_{O-a}, t_{a-b}) = \sum_{k=1}^{6} \frac{1}{6} \cdot P\left(T_2 \leq t_1^{(k)} | t_{O-a}, t_{a-b}\right) = \frac{1}{6}(1 + 0 + 0 + 1 + 0 + 0) = \frac{2}{6}$$
$$\approx 0.333. \blacksquare$$

In the case with two paths, the probability of lower travel time directly translates to the fraction of contribution to the strategy, since the value is based on the number of realizations. In cases with more than two paths, each such pairwise comparison would give an upper bound on the fraction of contribution to the strategy, with the lowest pairwise probability a given path can achieve being the tightest upper bound. To illustrate this principle, the following example is considered.

<u>*Example 6. Probabilistic Path Comparisons for Strategy Building*</u>

Consider the example network 2, in Figure 7-2, with the travel time distributions as given in Table 7-2 and Table 7-3. Let the paths *O-a-e-D* and *O-a-e-f-D* be called path 3 and 4 with travel time random variables denoted T_3 and T_4, respectively. The probability $P(T_2 \leq T_1) = 0.333$ as seen with Example 4, and it can be shown that $P(T_2 \leq T_3) = 0.5$ and $P(T_2 \leq T_4) = 0.5$. In forming the optimal strategy, path 2 would be at least as good as paths 3 and 4 in 50% of the cases, but at least as good as path 1 in 33.3% of the cases. Thus, there are cases where path 2 is chosen relative to paths 3 and 4, but path 1 can still be chosen in place of path 2. Thus, path 2 would be selected in at most 33.3% of the network realizations. ■

Therefore, the pairwise random variable comparison of path travel time distributions as introduced here can be used as the basis for eliminating paths that are not likely to contribute to the resulting strategy, i.e., likely inefficient, and is exactly the idea behind the relaxed deterministic dominance (RDD) criterion from the previous chapter. The probabilistic criterion in RDD is thus a proxy for the path's contribution to the strategy as an upper bound on the probability of the path's contribution to the strategy.

7.3 Solution Methodology

The proposed solution approach for the trajectory-adaptive reliable least-time strategies (TA-RLTS) problem in stochastic dynamic networks with generalized correlations is a 2-stage approach that can be used for finding exact and approximate solutions.

An important objective for this solution methodology is to unify several problems and solution types for path finding and routing in stochastic dynamic networks that have been treated separately in the literature. Firstly, the a priori path finding problem is unified with the trajectory adaptive routing problem via the shared solution approach for generating eligible paths for finding both a priori and adaptive solutions. Secondly, the solution methodology for both exact and approximate solutions unifies these, typically disparate, ways to approach the problem at hand.

The need for approximate solution methods has been demonstrated in a number of studies focused on adaptive routing in stochastic networks. However, while some distinct approximate strategy finding methods exist, the important gap of concern in this study is to provide an approach that can be used for both approximate and exact solutions, with an adjustable level of uncertainty that can be tuned for specific applications based on particular accuracy and efficiency requirements.

The definition for the TA-RLTS problem in this chapter, allowing for continuous random travel times requires a method that can approach the stochastic optimization problem with an infinite and uncountable event space, which cannot be solved by enumeration. Hence, the 2-stage solution approach is accompanied with simulation-based path travel time estimation approaches, as presented in Chapter 5.

Furthermore, the proposed solution methodology unifies different types of problems that can results from different objective or optimality definitions in stochastic dynamic networks, as presented in section 7.2.4.1. The solution methodology is not specific to a single objective function but is intended to be suitable for a range or reliability-based least-time objectives. With this characteristic, the approach is presented for finding the optimal strategy for multiple objectives in a single run of the algorithm.

Given a set of departure times and a specified origin node, Stage 1 of the solution approach is the generation of eligible paths to all destination nodes in the network according to the TD-RLTP approach from Chapter 6. Stage 2 uses the reduced network based on this set of eligible paths and builds the optimal trajectory-adaptive routing strategy with its travel time distribution. Section 7.3.1 summarizes the methodology for generating of eligible paths, the details of which can be found in Chapter 6. Section 7.3.2 presents the approach for finding the optimal routing strategies, i.e., the algorithm for Stage 2 of the proposed solution approach.

7.3.1 Eligible Path Generation

The solution approach for generating eligible non-dominated paths was presented as part of the solution for the RLTP problem in Chapter 6. Two types of dominance criteria were presented, adjusted from the general framework for determining a priori nondominated least time paths in stochastic time-varying networks by Miller-Hooks and Mahmassani (1998b). The approximate adjusted First-Order Stochastic Dominance (A-FSD) criterion and relaxed deterministic dominance criterion (RDD) are adjusted to be applied at intermediate nodes in the stochastic dynamic network with correlations, based on the stochastic dominance and deterministic dominance criteria by Miller-Hooks and Mahmassani (1998b).

The RDD criterion was modified with a relaxation parameter ϵ that specifies the allowable level of risk in eliminating a potentially viable path. In the TA-RLTS problem addressed here, this risk level corresponds to an upper bound on the likelihood of a path being part of the optimal least-time strategy, as shown via Examples 4 through 6. Some supporting explanations are included in this section.

In the TA-RLTS problem, only paths that share one or more of their initial links, from the origin to an intermediate node, could be part of the same strategy. A routing strategy constructed under a given objective function can be seen as a collection of a priori paths. Section 7.2.3 showed that the optimal strategy is always at least as good as any of the a priori paths it is composed of. Thus, the travel time of any a priori path can be seen as the upper bound on the travel time of the optimal strategy created using that path. Comparison of paths not on the same strategy is equivalent to the comparison of the upper bounds (on the objective) of the corresponding strategies and gives no indication of the likelihood of their usefulness for strategy building. Thus, in identifying eligible paths that will (or are likely to) contribute to a resulting strategy, only paths that share at least one initial link from the origin to an intermediate node are compared to one another.

Given this understanding, in generating eligible paths we will aim to define the likelihood that a path will be part of an optimal strategy. Conversely, if a path is to be eliminated from consideration for the optimal strategy (i.e., designated as ineligible), that positive likelihood can be seen as risk of loss on the objective of the strategy. Considering building a strategy using paths l and k, if $P(\theta_l < \theta_k) = 0$, path l can safely be eliminated since it will not contribute to a least-time strategy that would involve path k. Thus, based on the comparisons established in Examples 5 and 6 in section 7.2.4.3, the likelihood $P(\theta_k \leq \theta_l)$ can be identified as a measure of the path's eligibility. Conversely, in designating paths as ineligible $P(\theta_k \leq \theta_l)$ can be seen identified as a measure of risk for path *k* relative to path *l*.

If aiming to generate only paths that have some positive likelihood of being part of an optimal strategy and safely eliminate all paths with no such likelihood, the criterion $P(\theta_k \leq \theta_l) \leq 0$ may be used, where k and l are paths that share one or more of their initial links. The path generation with the criterion of eliminating all paths dominated according to this criterion is equivalent to the deterministic dominance criterion by Miller-Hooks and Mahmassani. To allow for approximate solutions in cases of large networks, this elimination criterion can be relaxed: instead of requiring that $P(\theta_k \leq \theta_l) \leq 0$ for a path to be non-eligible, a small value ϵ can be defined so that any path k is eliminated if there exists a path l (here, potentially on the same strategy) such that $P(\theta_k \leq \theta_l) \leq \epsilon$. Thus, the path k is eliminated with a small likelihood of contributing to the optimal strategy, which we refer to as the risk-level tolerance in generating eligible paths.

By definition, the criterion with an $\epsilon > 0$ will eliminate at least as many (and potentially more) paths as the deterministic dominance criterion and as such will result in solutions no better than the deterministic dominance criterion. However, by eliminating a larger number of paths, these approximate or relaxed dominance criteria can reduce the computational effort required for generating eligible paths and their resulting strategies. Following from the above and the results in examples 5 and 6, the following proposition is presented:

Proposition 7-1. For a pair of paths k_{rs} and l_{rs} for a specified origin-destination pair *r-s*, which share at least one link from the origin to an intermediate node, the probability $P\left(\theta_k^{t_0} \leq \theta_l^{t_0}\right)$ is an upper bound on the likelihood of path k contributing to the optimal joint strategy formed by k, l and other paths for the O-D pair $r - s$, here referred to as the risk of eliminating path k_{rs} for departure time t_0. The best (i.e., tightest) upper bound on the risk of eliminating path k_{rs} can be found as $\min_{l'_{rs}} \{P\left(\theta_k^{t_0} \leq \theta_{l'}^{t_0}\right)\}$ over all other paths l'_{rs} for the same strategy.

Proof 7-1. Suppose in comparing the paths k_{rs} and l_{rs} for a specified origin-destination pair *r-s*, which share at least one link from the origin to an intermediate node via joint realizations on the network, we find that $P\left(\theta_k^{t_0} \leq \theta_l^{t_0}\right) = p < 1$. This indicates that in $p < 1$ proportion of the possible realizations for the network, path k_{rs} has a shorter travel time. In building the strategy ς_1 that contains paths k and l, given the shared links from the origin to the destination between the two paths, suppose path k is selected for $q > p$ proportion of the possible realizations. Then for a $(q - p)$ portion of the realizations, travel time on path k is grater than that on path l but path k is selected for ς_1. Thus, there exists another strategy ς_2 in which for those $(q - p)$ realizations path

l with a lower travel time is selected, making strategy ς_2 FSD-dominant relative to ς_1, thus by contradiction showing that ς_1 cannot be the optimal strategy composed of paths k and l. From this, by induction we can see that for a larger number of paths, the contribution of k to the strategy is constrained by the minimum of the pairwise comparison probabilities.

The criterion established here allows for the comparison and elimination of paths via their full travel time distributions from the origin to destination. However, removing potentially inefficient paths at intermediate branching nodes is beneficial when the estimation of full path travel time distributions is computationally expensive, and even more so when working with larger networks. The above criterion is extended so as to evaluate the potential for useful paths at a given branching node. Using truncated link travel time distributions, the Proposition 6-1 and Corollary 6-1 from the previous chapter apply here.

From Corollary 6-1, it is established that if $P\left(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}\right) \leq \epsilon$, and $P\left(\theta_l^{t_0} \leq \theta_{l'}^{t_0}\right) \leq P\left(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}\right)$ then $P\left(\theta_l^{t_0} \leq \theta_{l'}^{t_0}\right) \leq \epsilon$. Therefore, following from the likelihood of including eligible paths discussed in Proposition 7-1 and its equivalent risk of designating ineligible paths, $P\left(\theta_{k_{ri}} + \tau_i^{min} \leq \theta_{l'}^{t_0}\right) \leq \epsilon$ can be used as a criterion for early elimination at an intermediate node i by showing that even the best case for its extensions will not meet the eligibility criterion. This probabilistic criterion based on the risk-level tolerance value ϵ is a heuristic criterion which allows for certain paths to be eliminated and not considered in the building of the final adaptive routing strategy, if their likelihood of contributing to the strategy is low enough. The higher the risk tolerance value ϵ, the larger the number of eliminated paths (here referred to as ineligible for the given risk tolerance), leading to reduced computational effort for the building of the optimal adaptive strategy, but with the trade-off of increased overall travel time value and objective function value of the resulting strategy.

Therefore, Stage 1 of the TA-RLTS solution approach is the TD-RLTP algorithm for path generation from section 6.3.2. The note on estimating path travel time distributions remains relevant here, and this chapter will also use the NORmal-To-Anything (NORTA) approach with time-dependence and time-varying distributions from Chapter 5. The path generation algorithm above terminates after having determined all eligible paths to all destination nodes from the origin, for all given departure times. The primary results needed from the Stage 1 procedure are the eligible path travel time distributions $U_d^{k\mathbb{t}}\ \forall\ d \in D, k \in \{1, 2, \ldots, M\}, a \in A, \mathbb{t} \in \mathbb{T}$. The sets $K_d = \{k\ |\Lambda_O^k(\mathbb{t}, d) = 1\} \forall\ d \in D, \mathbb{t} \in \mathbb{T}$ contain the identifiers for the eligible paths for each destination and departure time. are eligible, and each path can be traced back using the pointers $p_i^k, L_i^k\ \forall\ k \in \{1, 2, \ldots, M\}, i \in N$. The conditionally sampled link travel time distributions are also saved and can be further used in building the optimal strategy in Stage 2 of the solution approach.

7.3.2 Optimal Routing Strategy Finding

The second stage of the solution approach uses the generated eligible paths and their path travel time distributions from Stage 1 to find the optimal trajectory-adaptive strategies for a set of objectives. The strategy building can be performed for one or multiple objectives simultaneously, solving the routing strategy problem for heterogenous users with different reliability-based objectives.

The algorithm reads the set of eligible paths each destination to the origin node and at each decision node, the conditional distributions for the sub-strategy at that node are obtained, the objective value(s) are computed and the appropriate sub-strategy path for each sampled realization of the network is selected. Given that the reading of the network and the simulation are performed once, solving the optimal routing strategy for multiple objectives does not have a significant impact on the computational time and effort. However, since different objectives can be expected to result in different strategy distributions, strategy building with multiple criteria generates and holds a larger amount of data. The procedure for finding optimal routing strategies is presented below.

Continuing from Stage 1 the decision nodes in the network are based on the eligible paths from each node to the destination, not simply by the network structure as seen in the examples in the previous section.

Solution Algorithm, Stage 2: Procedure for Finding Optimal Routing Strategies

Given:

The network $G(N, A, \mathcal{T})$, where $\mathcal{T}$ is the set of time periods $\{t_0, t_1, \dots, t_{L-1}\}$. The function $\phi(\cdot) \in \mathcal{T}$.
The joint time-varying link travel time distributions with time-varying covariance structure.
The origin node O. The destination $d \in \mathcal{D}$, which if unspecified is set to $D = N \backslash \{O\}$.
The set of departure times $\mathbb{t} \in \mathbb{T}$.
The results from Stage 1: the travel time distributions for the set of eligible paths, for the selected dominance criterion, from O to each of the destinations $d \in \mathcal{D}$ and for each departure time $\mathbb{t} \in \mathbb{T}$ contained in the vector-labels $U_d^{k\mathbb{t}}$ and $u_i^{k\mathbb{t}}$, the corresponding vector pointers $p_i^k, L_i^k \ \forall\, k \in \{1, 2, \dots, M\}, i \in N$.
The set of objective functions to be used for determining the strategies, OBJ.

Find:

The travel time distributions of the optimal strategy $U_{ext}^{ik,\mathbb{t}}(obj)$ and the pointers $C_{ext}^{ik,\mathbb{t}}$ $\forall i \in N, \mathbb{t} \in \mathbb{T}, k \in \{1, 2 \dots, M\}\ s.t.\ L_i^k \neq \infty, obj \in OBJ$, for each $d \in \mathcal{D}$.

Step 0: Initialization

Step 0.1: Find all eligible paths.

For the destination node $d \in \mathcal{D}$, from the eligibility indicators $\Lambda^{k\mathbb{t}}\ \forall\, k \in \{1, 2, \dots, M\}, \mathbb{t} \in \mathbb{T}$, save the node-path ID pairs and links that each path traverses, using the pointers L and P as follows:

<u>For</u> each $\mathbb{t} \in \mathbb{T}$:

Define the set $P_n^{\mathbb{t}} = \{\}$ and $P_a^{\mathbb{t}} = \{\}$.

<u>For</u> each $k \in K = \{k \mid \Lambda_{dd}^{k\mathbb{t}} = 1\}$:

Initialize the lists $P_{nk} = \{\}$ and $P_{ak} = \{\}$.

Let $j = D, \mu = k$. Add $(j - \mu)$ to the list P_{nk}.

While $j \neq O$:

Let $j' = j$. Update $j = L_j^\mu$ and $\mu = p_j^\mu$.
Add $(j - \mu)$ to the list P_{nk}.
Add the link $a = (j, j')$ to the list P_{ak}.
Reverse the lists P_{nk} and P_{ak}. Add P_{nk} to $P_n^{\mathbb{t}}$ and P_{ak} to $P_a^{\mathbb{t}}$.
For each time $\mathbb{t} \in \mathbb{T}$:
Save the number of occurrences for each of the node-path ID pairs $(j - \mu)$ in any path in $P_n^{\mathbb{t}}$ (indicating the number of sub-paths to destination) as $N_{j-\mu}^{\mathbb{t}}$.

Step 0.2: Initialize pointers and labels

Initialize SE, a first-in-first-out (FIFO) queue of scan-eligible node-path ID pairs.
For each final pointer at the destination $(d - k) \mid \Lambda_{dd}^{k\mathbb{t}} = 1$, add $(d - k)$ to SE.
Initialize the list R = {} to contain all read nodes.
Initiate empty vector labels $U_{ext}^{ik,\mathbb{t}} = [None]_S$ and vector pointers $C_{ext}^{ik,\mathbb{t}} = [None]_S$ $\forall i \in N, k \in \{1, 2 \ldots, M\}, s.t. L_i^k \neq \infty, \forall t_p \in \mathbb{T}_0$ each of size S.

For each $t_p \in \mathbb{T}_0$:

Step 1: SE queue check

If the SE queue is not empty, take the node-path ID pair $(j - \mu)$ at the front of the queue, i.e., in a FIFO manner.
If the number of unique paths in P_{ak} is 1, go to Step 4.
Otherwise, find the previous node (j', μ') such that $j' = L_j^\mu$ and $\mu' = p_j^\mu$. Add (j', μ') to R.
If (j', μ') has $N_{j'-\mu'}^{\mathbb{t}}$ occurrences in R and $N_{j'-\mu'}^{\mathbb{t}} = 1$, add (j', μ') to the SE queue.
Else if (j', μ') has $N_{j'-\mu'}^{\mathbb{t}}$ occurrences in R and $N_{j'-\mu'}^{\mathbb{t}} > 1$, then (j', μ') is a decision node, go to Step 2.
Otherwise, go back to Step 1.
Otherwise, if SE queue is empty, go to Step 3.

Step 2: Decision node evaluation

Find the set of paths that contain the branching node-path ID pair $(j' - \mu')$ from $P_n^{\mathbb{t}}$ and $P_a^{\mathbb{t}}$ as $P_n^{\mathbb{t}}(j', \mu')$ and $P_a^{\mathbb{t}}(j', \mu')$.
Call the *Decision Node Evaluation Procedure* for node-path ID pair $(j' - \mu')$ to obtain the vector labels and pointers for the sub-strategy at the branching node, $U^{j'\mu'}$ and $C^{j'\mu'}$, respectively for each objective function obj. Let $U_{ext}^{j'\mu'}(obj) = U^{j'\mu'}(obj)$ and $C_{ext}^{j'\mu'}(obj) = C^{j'\mu'}(obj)$.
Replace the paths from $P_a^{t_p}(j, \mu)$ in $P_a^{t_p}$ with $P_s^{j\mu}$ (i.e., the shared portion of the path) and append the indicator ext at the end.
Add (j', μ') to the SE queue.
Go to Step 1.

Step 4: Origin node evaluation

The resulting $U_{ext}^{o,0}(obj)$ and $C_{ext}^{o,0}(obj)$ contain the strategy travel time distribution and point to the sub-strategy from the origin respectively, for each of the objective functions

obj. The objective function value for each of the strategies can be determined by evaluating the distribution for that objective $U_{ext}^{o,0}(obj)$.

The decision node evaluation procedure is given below. This procedure evaluates the possible strategy paths at the decision node and determines the travel time distribution of the conditional sub-strategy from that branching node.

Decision Node Evaluation Procedure

Given:

All of the items given for the Stage 2 Procedure for Finding Optimal Routing Strategies
The current branching node-path ID pair $(j-\mu)$, the set of paths that contain it from $P_n^{\mathfrak{t}}$ and $P_a^{\mathfrak{t}}$ as $P_n^{\mathfrak{t}}(j,\mu)$ and $P_a^{\mathfrak{t}}(j,\mu)$..

Find:

The vector labels $U^{j\mu}$ and pointers $C^{j\mu}$ for the sub-strategy at the branching node $(j-\mu)$

Step 0: Initialize

Find the shared links for the set of paths $P_a^{\mathfrak{t}}(j,\mu)$ as $P_s^{j\mu}$ and the node-path ID pairs from $P_n^{\mathfrak{t}}(j,\mu)$, from the origin to the branching node j.
Separately, save the sub-paths from node j to the destination d for each of the paths in set $P_a^{\mathfrak{t}}(j,\mu)$.
Retrieve the travel times on the shared links from the origin to node j from $u_a^{k\mathfrak{t}}$ and find the corresponding time intervals via the function $\phi(\cdot)$.
For each of the objective functions considered set $U^{obj} = [None]_S$

Step 1: Retrieve conditional distributions

<u>For</u> each extension sub-path $k' \in P_a^{\mathfrak{t}}(j,\mu)$:
- If the extension k' does not contain the indicator ext:
- <u>For</u> each sample $s \in \{1,2,\dots,S\}$:
  - Find the list of exit time intervals using ϕ, find the appropriate time-dependent joint link travel time distributions and covariance matrix using the exit bins.
  - <u>For</u> each link $a' \in k'$:
    - Sequentially sample from $\pi_{a'}^{e}$, where e is the time interval for exit time at $(j-\mu($, conditional on the previous link travel times $\tau_{a_c}^{\mu\mathfrak{t}}[s]$. Save the samples into the temporary labels $u_{temp}^{s}(a',s')\ \forall\ s' \in \{1,2,\dots,S'\}$.
    - Compute the sum $u_{k'}^{s} = \sum_{a'\in k'} u_{temp}^{s}(a',s')$
- Otherwise, if the extension k' contains the indicator ext:

Retrieve the link travel time distributions $u_a^z \forall a' \in k' if\ a \in A, s \in \{1, 2, \dots, S\}$ and $u_{ext} = U_{ext}^{j'z}$ for the final node j', where z is a place holder for the corresponding path identifier. Then let $u_{k'}^s = \sum_{a'} u_a^z(s) + u_{ext}(s)$.

Step 2: Evaluate branching node extensions

For each sample $s \in \{1, 2, \dots, S\}$:

For each objective function obj:

Determine the objective function value for the distribution of $O_{k'} = u_{k'}^s, \forall\ k' \in P_a^{\mathfrak{t}}(j, \mu)$.

Select $k^* = k' \in P_a^{\mathfrak{t}}(j, \mu)$ with the minimum or maximum (depending on the objective) value for the $O_{k'}$, let $U^{obj}(s) = u_{k^*}^s$ and $C^{obj}(s) = k^*$.

Set $U^{j\mu}(obj) = U^{obj}$ and $C^{j\mu}(obj) = C^{obj}$ ($\forall\ s \in \{1, 2, \dots, S\}$).

The result of the Stage 2 procedure is the final travel time distribution from the origin, based on each of the objectives obj, which can be directly retrieved from $U_{ext}^{o,0}(obj)$. The objective value for each of the distributions can be obtained with their corresponding objective functions. The vector pointers $C_{ext}^{o,0}(obj)$ point to the (conditional) sub-strategy from the origin for each of the objective functions obj.

7.4 Numerical Experiments

This section presents the numerical experiments designed to evaluate the performance of the solution methodology for the TA-RLTS problem in this chapter. The network and data used for the experiments are same as those for the experiments in Chapter 6, using the same 25 scenarios for the simulations, the 7:00 to 10:00 a.m. peak period, along with the same origin node and departure times. The design of this chapter's experiments is outlined in 7.4.1, and the results and analysis are presented in 7.4.2.

7.4.1 Design of the Numerical Experiments

The numerical experiments were designed to evaluate the performance of the solution approach, in terms of efficiency and accuracy, across the various dominance criteria for the path generation and different objective functions for selecting the optimal strategy.

Similar to the numerical experiments in the previous chapter, all 7 different dominance criteria were tested, the RDD with 6 ϵ values, $\epsilon \in \{0, 0.01, 0.05, 0.1, 0.15, 0.2\}$ and the A-FSD criterion. Again, the RDD with $\epsilon = 0$ is an exact criterion leading to exact solutions for the optimal strategy. The six objective functions are also defined corresponding to those in the previous chapter, the Value at Risk (VaR) and Conditional Value at Risk (CVaR), each with three values for the α-percentile tail, $\alpha \in \{0.7, 0.8, 0.9\}$.

The TA-RLTS problem was solved for a single randomly selected origin node to all destination nodes in the network, for the 5 departure times, with an optimal strategy determined for each dominance criterion and for each objective function, resulting in a total of 331,170 solutions.

As in the previous chapter, the path travel time distributions were estimated via the time-dependent NORTA approach with time-varying covariance structure, introduced in Chapter 5.

7.4.1.1 Research Questions and Performance Measures

The numerical experiments were designed to answer several research questions to gain an understanding of the TA-RLTS problem itself and the proposed solution approach.

Regarding the TA-RLTS problem and the solutions obtained in these numerical experiments, the experiments tested are intended to answer the following questions:

- How complex are the optimal routing strategies?

    - To be evaluated via the number of decision nodes and branching nodes in the optimal routing strategy, for each objective function and on average.
- How computationally expensive is it to obtain TA-RLTS solutions?
    - To be evaluated via the average and maximum computational run times.
- What is the effect of selecting a strategy compared to an a priori path solution?
    - To be evaluated via the objective values for optimal strategies compared to those of optimal a priori paths, for each objective function.

Regarding the performance of the solution approaches, the following questions are to be answered via the numerical experiments:

- How does the performance of the solution approach vary with the approximate dominance criteria?
    - To be evaluated relative to the base full strategy via the change in the number of branching nodes, the change in the objective function values, and the change in run times.
- How does the performance of the solution for the approximate criteria vary across the different objective functions?
    - To be evaluated via the change in the objective function values of optimal solutions for each objective function.

To answer these research questions, the applicable performance characteristics and measures include actual run times, numbers of decision and branching nodes for optimal strategies, and percent relative difference to measure change in objective values.

7.4.2 Results and Analysis of the Numerical Experiments

The numerical results are separated into two sections, according to the two types of research questions, evaluating TA-RLTS solutions relative to a priori paths in 7.4.2.1 and the performance of the TA-RLTS approach with its approximations in 7.4.2.2.

7.4.2.1 Evaluation of Trajectory-Adaptive Strategy Solutions

Considering the first two research questions, the results in Table 7-4 show the overall complexity of the TA-RLTS solutions via the average and maximum computational run times and the average and maximum numbers of branching nodes for each solution.

To clarify, a solution is considered a strategy obtained for a specific origin-destination pair, with a specified departure time, and for a specified objective function. However, a single run finds the solutions for all departure times and all objectives.

Table 7-4. Complexity of trajectory-adaptive strategy solutions: run times and branching nodes numbers

Dominance Criterion		Average Run Time (s)	Maximum Run Time (s)	Average Branching Nodes Number	Maximum Branching Nodes Number
RDD	$\epsilon = 0$	27.90	7.46	165.08	37
	$\epsilon = 0.01$	21.09	4.44	120.09	23
	$\epsilon = 0.05$	18.21	3.45	99.17	21
	$\epsilon = 0.1$	15.93	2.82	82.14	18
	$\epsilon = 0.15$	13.99	2.39	76.37	16
	$\epsilon = 0.2$	11.63	1.92	57.57	15
A-FSD		10.70	1.61	57.07	13

Table 7-4 shows that, on average, the run times decrease with the complexity of the strategy itself, i.e., its number of branching nodes, across the different dominance criteria used. This trend can be understood as follows: the stricter the dominance criterion, reducing the set of eligible paths to destination, the fewer the number of decision nodes, thus restricting the number of possible branching nodes and reducing the computational effort needed due to fewer evaluation points. These results are also shown in Figure 7-3.

Focusing on the solutions for the RDD criterion with $\epsilon = 0$, considered the base full strategy, on average run times for obtaining the strategy are 27.9 seconds, but can be as high as 165 seconds. On average, a trajectory-adaptive strategy had 7.46 branching nodes, and the maximum number of branching nodes can be up to 37. These values indicate that the strategies can contain a significant number of options compared to the a priori path. The average and maximum number of strategy paths for the full strategy solutions were 14.27 and 67 paths, respectively.

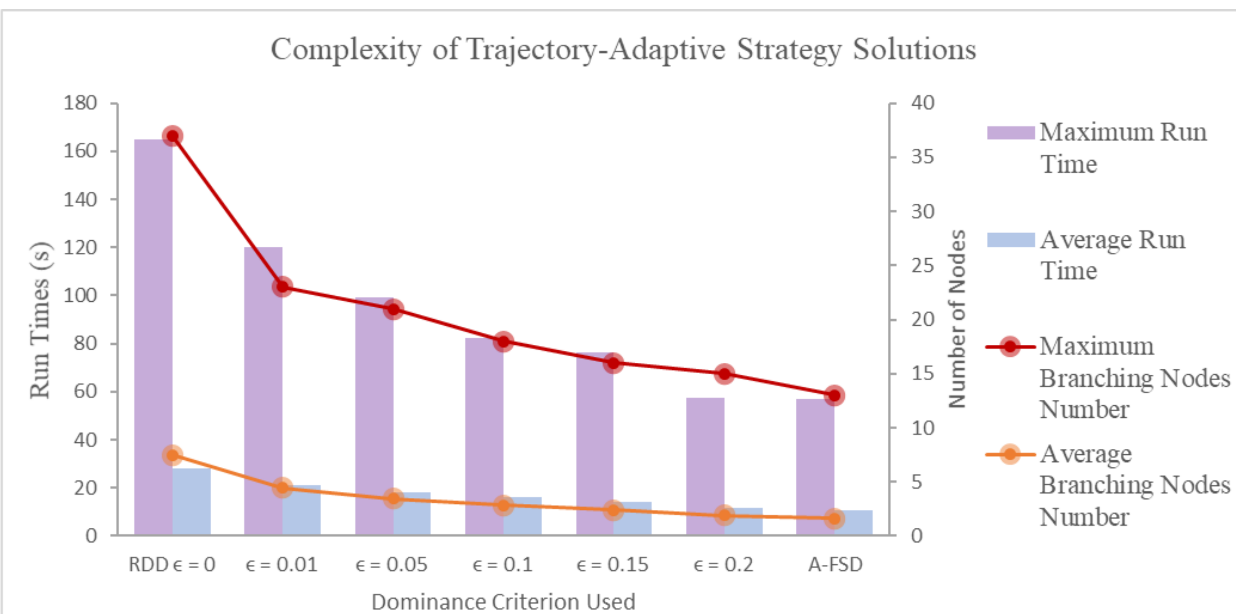

Figure 7-3. Overall complexity of trajectory-adaptive strategy solutions: run times and branching nodes numbers

Next, consider the effect of choosing a strategy, relative to an a priori solution in terms of the objective function values. The objective function values, on average across all cases and for each objective function (in minutes of travel time), for the full trajectory-adaptive strategy and the a priori path solutions are shown in Table 7-5, along with average and weighted average percent relative difference values.

Table 7-5. Objective values and differences for trajectory-adaptive strategy and a priori solution

	Average objective function value		Percent relative difference	
	Full TA Strategy	A Priori Path	Average	Weighted Average
All Cases	28.75	31.53	11.23%	8.51%
VaR, $\alpha = 0.7$	24.41	25.65	9.24%	4.86%
VaR, $\alpha = 0.8$	26.58	28.21	10.24%	5.76%
VaR, $\alpha = 0.9$	29.31	31.81	10.99%	7.88%
CVaR, $\alpha = 0.7$	28.82	32.02	11.41%	10.00%
CVaR, $\alpha = 0.8$	30.46	34.19	12.12%	10.93%
CVaR, $\alpha = 0.9$	32.95	37.28	13.40%	11.61%

Table 7-5 shows that on average, the percent relative difference in objective function value is 11.23%, meaning that the objective function value for the strategy is that much lower than for the a priori solution, and the weighted average that adjusts for the actual value of time savings is 8.51%. Furthermore, the percent relative difference, which can be interpreted as percent travel time savings on the objective function value, increase for the stronger, more risk-sensitive objectives. This trend can be better observed in the graphical representation of these results shown in Figure 7-4. As the α-percentile level increases, both for VaR- and CVaR-based objectives, the percent relative time savings increase. Similarly, the CVaR-based objectives have higher percent relative difference than the VaR-based objectives at equal values for α.

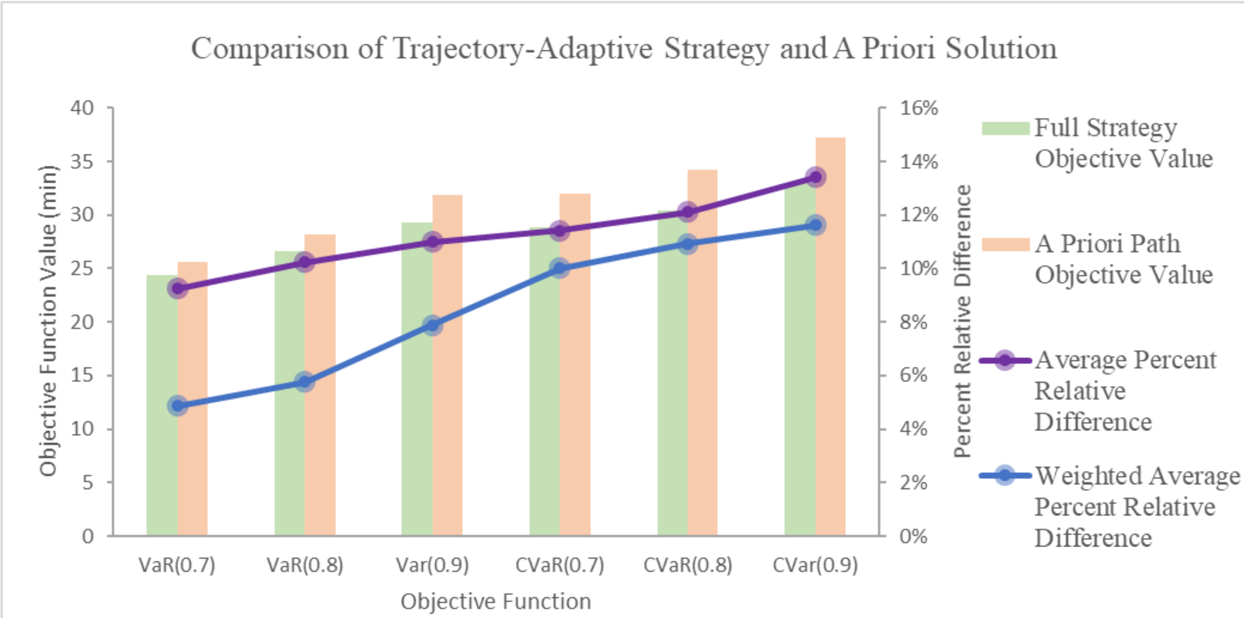

Figure 7-4. Objective values and differences for trajectory-adaptive strategy and a priori solution for all objective functions

The interpretation of these results is that for more risk-averse travelers or applications, where the objective is either with higher α values or a more risk sensitive CVaR-based objective, the percent relative difference between the two solutions is greater. Thus, for such travelers or applications, it may be more important to consider trajectory-adaptive solutions as they can lead to much more significant savings on the objective function, here observed on average at 13.4% for the $\text{CVaR}, \alpha = 0.9$ objective function. For less risk-sensitive travelers or applications, such as those with VaR-based objectives and $\alpha = 0.7$ or $\alpha = 0.8$, the time savings are much lower, but on average still at 9.24% and 10.24%, respectively.

7.4.2.2 Performance of the TA-RLTS Approach and Approximations

This section considers the performance of the solution approach with the approximate dominance criteria, relative to the full strategy for the exact criterion, i.e., RDD with $\epsilon = 0$. The run times and number of branching nodes, average and maximum values, for the different dominance criteria are shown in Table 7-4 and Figure 7-3 in the previous section, showing the downward trend in the solution complexity and computational effort as the ϵ increases for the RDD criterion and moving to the A-FSD criterion. In Table 7-6, these results are supplemented with the relative percent difference on the average run time, number of branches and objective function value for all cases, relative to the exact solution with the RDD with $\epsilon = 0$ criterion.

Table 7-6. Average percent relative difference in run times, number of branches and objective values for approximate solution cases

Approximate Criterion Used	Average Percent Relative Difference		
	Run Time	Number of Branches	Objective Value
RDD $\epsilon = 0.01$	−31.32%	−33.18%	3.36%
RDD $\epsilon = 0.05$	−41.32%	−45.72%	4.67%
RDD $\epsilon = 0.1$	−49.42%	−54.77%	5.39%
RDD $\epsilon = 0.15$	−55.09%	−61.14%	6.73%
RDD $\epsilon = 0.2$	−61.94%	−67.62%	7.85%
A-FSD	−65.60%	−71.49%	8.25%

Moving from the exact criterion to using the eligible paths generated via the approximate criteria can lead to significant time savings, up to 65.6% average decrease in the run time when using the A-FSD criterion, and from 31.32% to 61.94% run time savings when using the RDD with ϵ values from 0.01 to 0.2. This result is similar to that observed for the a priori solutions' run times in the previous chapter. The number of branches in the optimal strategy decreases with the run time. However, these savings in computational effort and complexity are accompanied with a strict increase in the percent relative difference of the objective function value, where the approximate approaches lead to an increase in the objective function value.

These trends can be better observed in Figure 7-5. An important observation from these results is that the solution based on the A-FSD criterion produces the highest increase in the objective function values relative to the exact solution.

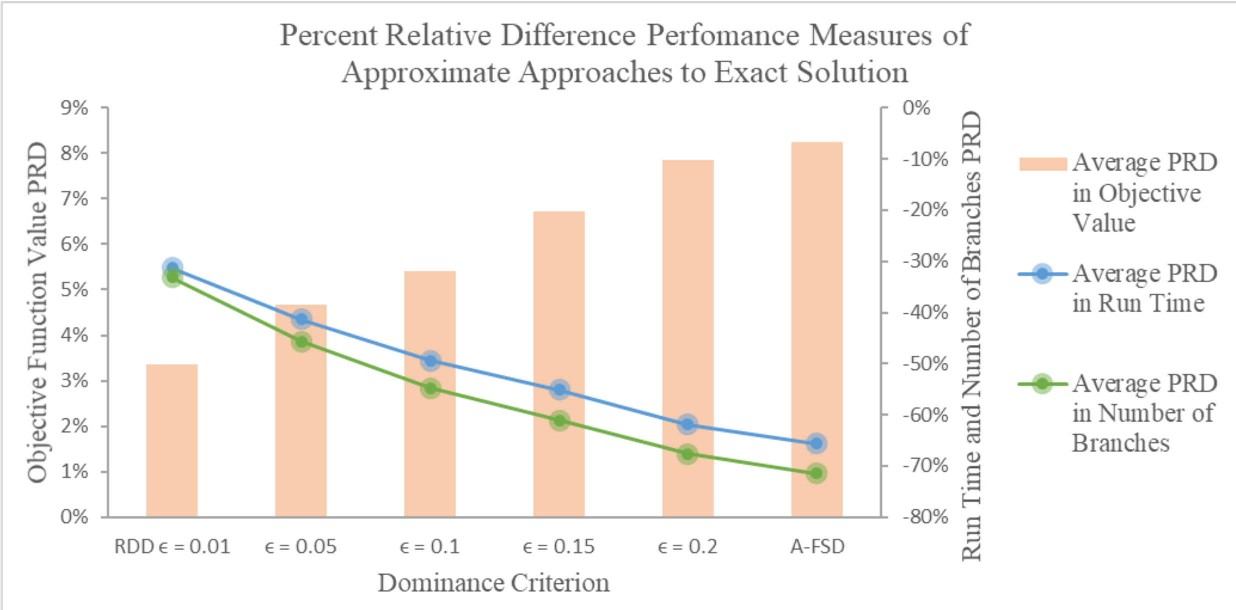

Figure 7-5. Average percent relative difference in run times, number of branches and objective values for approximate solution cases

This result is interesting, since the A-FSD was observed to be the second-best approximate criterion for the a priori solution. Of course, the important reason for this discrepancy is the fact that the FSD criterion (and subsequently the stronger A-FSD) criterion does not eliminate paths based on their likelihood to contribute to a strategy, but rather the path's potential to be a priori optimal. This was demonstrated via Example 4 in section 7.2.4. On the other hand, the RDD criterion, having been derived as a measure of the path's contribution to the strategy is seen to result in an increase in the objective function value that increases with the risk tolerance parameter ϵ.

Next, to consider how the performance varies across the objective functions, Table 7-7 shows the percent relative difference in objective value for the approximate approaches for each of the objective functions, where the VaR and CVaR objectives are abbreviated as V and C, respectively.

Table 7-7. Average percent relative difference objective values for approximate solution cases by objective function

Approximate Criterion Used	Average Percent Relative Difference					
	$\mathrm{V}(\alpha = 0.7)$	$\mathrm{V}(\alpha = 0.8)$	$\mathrm{V}(\alpha = 0.9)$	$\mathrm{C}(\alpha = 0.7)$	$\mathrm{V}(\alpha = 0.8)$	$\mathrm{V}(\alpha = 0.9)$
RDD $\epsilon = 0.01$	2.83%	3.19%	3.36%	3.36%	3.54%	3.88%
RDD $\epsilon = 0.05$	4.02%	4.48%	4.64%	4.69%	4.85%	5.34%
RDD $\epsilon = 0.1$	4.56%	5.09%	5.30%	5.43%	5.69%	6.30%
RDD $\epsilon = 0.15$	5.01%	5.89%	6.31%	6.87%	7.46%	8.82%
RDD $\epsilon = 0.2$	5.77%	6.84%	7.47%	7.98%	8.68%	10.32%
A-FSD	6.80%	7.89%	8.31%	8.25%	8.71%	9.56%

A similar trend is observed, where the stronger dominance criteria generally lead to a higher percent relative difference from the exact solution for all objectives. Additionally, the difference in objective value increases for the more risk-sensitive objectives. These trends are further visually presented in Figure 7-6. The results in Figure 7-6 show that the effect of using an approximate criterion is greater for some objective functions compared to others. Namely, for both the VaR- and CVaR-based objectives, increasing the α percentile level increases the relative difference in the objective value compared to the exact solution.

Therefore, for applications with more risk-sensitive objectives it may be more important to use the exact dominance criterion or determine which would be a good approximate criterion. This is important since for risk-sensitive objectives, such as CVaR with $\alpha = 0.9$, the increase in objective was observed to reach 10.32%, relative to the exact solution. On the other hand, for less risk-sensitive objectives approximate criteria may be a good choice. For example, for the VaR objective with $\alpha = 0.7$, the RDD crierion with $\epsilon = 0.01$ has the lowest increase in objective value of 2.83% and the highest effect on the objective was observed at 3.88%, for the run time savings of about 31% up to 65%, seen in Table 7-6.

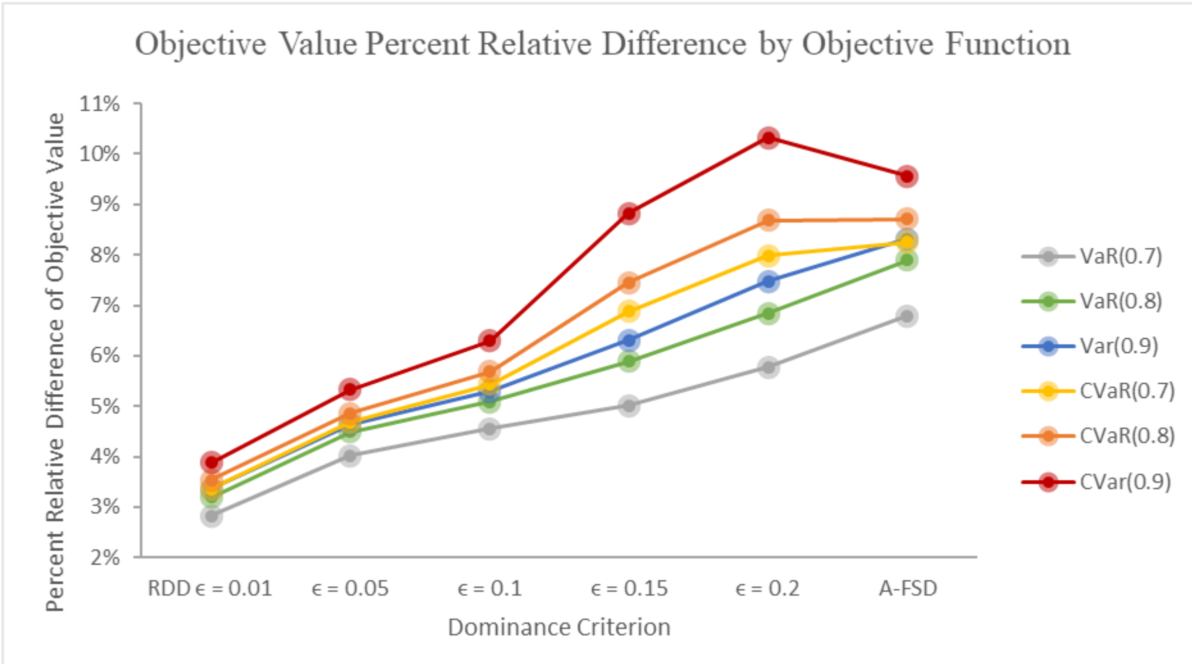

Figure 7-6. Average percent relative difference objective values for approximate solution cases by objective function

Another notable observation from Figure 7-6 is that the RDD criterion with sufficiently high ϵ value and for the more risk-sensitive objectives can be outperformed by the A-FSD criterion. In these results, this effect only occurs when $\epsilon = 0.2$, indicating that a 20% risk-tolerance level may be too high for certain objective types.

7.5 Conclusion and Future Work

This chapter focuses on the problem of finding optimal trajectory-adaptive reliable least-time strategies (TA-RLTS) in stochastic dynamic networks with generalized spatio-temporal correlations between link travel times. A two-stage path and strategy finding approach is presented, adjustable to admit different levels of risk or approximation into the solution via approximate dominance criteria for path generation. The solution method is suited for multiple reliability-based objectives in the least-time category.

Numerical experiments were performed on the network of Chicago to test the performance and applicability of the proposed solution algorithm and give insights into some of its characteristics. The experimental results showed the impact of a TA-RLTS solution compared to an a priori path solution. Additionally, the performance of the approach was tested with approximate dominance criteria and six different objective functions based on the Value at Risk (VaR) and Conditional Value at Risk (CVaR) risk measures. The numerical results demonstrate a trade-off between reduced computational effort (i.e., computational run time savings) and increased least-time based objectives (i.e., loss on the objective function). Additionally, the numerical experiments revealed that the effect of the approximation varies with the risk sensitivity of the objective function.

This study opens several questions that can be explored in future work. One assumption made in the problem definition is that cyclicity is precluded as a property of the user. Solving the problem with recourse can be considered in future work. Understanding the effect of allowing cyclic paths and the value of information available to or collected by the user are also worthy of further study. Additional computational tests may be needed in order to understand the performance of the solution algorithm and heuristic for different sizes and types of networks. Furthermore, there is some motivation of understanding how the effect of the risk-tolerance parameter may change in different networks and with different data sets.

Chapter 8 Information-Adaptive Routing in Connected Environments

8.1 Overview

Having considered the problems of a priori reliable least-time paths (RLTP) and trajectory-adaptive reliable least-time strategies (TA-RLTS) in stochastic dynamic networks, this chapter is focused on the problem of information-adaptive routing in the context of a connected environment.

Adaptive routing in stochastic dynamic networks was discussed in detail in Chapter 7, introducing adaptive routing problems as defined by two key characteristics of the problem: the availability of information and the traveler's response to that information. The problem of trajectory-adaptive routing strategies is concerned with a special case of partial information availability: the traveler's own trajectory, and specifically for the case of the proactive (i.e., strategic) traveler. The problem to be considered in this chapter approaches adaptive routing in the context of a connected environment and for a reactive traveler.

Information-adaptive routing in a connected environment allows for a more general definition of information availability. Information from connected vehicles traveling in the stochastic dynamic network may be available at varying levels, depending on the number of connected vehicles in the network, and in various parts across the network, depending on those vehicles' trajectories.

The problem defined in this chapter considers the use of such information that becomes available as time passes and connected vehicles make their way through the network with the response of a reactive traveler. In contrast to the proactive traveler's response to information, presented in the previous chapter, where the solution is a strategy consisting of a collection of paths, the reactive traveler receives information and adjusts the path to destination at each decision node. Thus, the solution for this type of problem consists of a path chosen at each decision node to the destination, based on the information available at the time of arrival at that node.

This chapter utilizes the representation for jointly distributed link travel times across the entire network as continuous random variables with time-varying distributions and correlation structures from Chapter 4; the approaches for path travel time distribution estimation from Chapter 5; and the eligible path generation approach from Chapter 6 to solve the information-adaptive reliable least-time routing (IA-RLTR) problem presented here.

The remainder of this chapter is organized as follows. The problem definition and its methodological difficulties are presented in section 8.2, including definitions of information availability in a connected environment in 8.2.2 and specifics on the IA-RLTR problem in 8.2.3. The solution methodology is presented in section 8.3, while section 8.4 focuses on the numerical experiments with the experimental design in 8.4.1 and the results and analysis in 8.4.2. Conclusions and discussions on future work are presented in section 8.5.

8.2 Problem Definition and Methodological Difficulties

The information-adaptive reliable least-time routing (IA-RLTR) problem and the methodological difficulties associated with it are presented in this section. First, an overview on the stochastic time-varying network and modeling is presented for completeness in 8.2.1, the details for which can be found in the corresponding sections in Chapter 6 and Chapter 7. Next, section 8.2.2 defines information availability in a connected environment and section 8.2.3 presents additional definitions and examples and defines the IA-RLTR problem at hand.

8.2.1 Stochastic Time-Varying Network Modeling and Notation

Similar to Chapter 6 and Chapter 7, this chapter defines an STV network $G(N, A, \mathcal{T})$ with N the set of $|N| = n$ nodes, A the set of $|A| = m$ links, and $\mathcal{T}$ the set of time periods. Link travel times are random variables jointly distributed across time, with travel time on link (i, j) at time t denoted Θ_{ij}^{t}. Here, the random variables are modeled to vary across the time periods in $\mathcal{T}$ but be constant within each single time period. Link travel times are modeled as continuous positive random variables with a truncated distributions π_{ij}^{t}, constrained by a minimum and maximum possible value. Dependencies between the link travel times are defined via link-pairwise covariances that vary over time-period pairs, so that $cov(\Theta_{ij}^{t_1}, \Theta_{kl}^{t_2})$ is the covariance between the travel time on link (i, j) during time period $t_1 \in \mathcal{T}$ and that on link (k, l) during time period $t_2 \in \mathcal{T}$.

Path travel time distributions are estimated according to the approaches in Chapter 5, and the assumptions from Chapter 6 and Chapter 7 carry over to this problem definition: only acyclic paths are considered and no waiting at nodes is permitted.

8.2.2 Information Availability in a Connected Environment

In this chapter, the stochastic dynamic network is assumed to be a connected environment. Elfar et al. (2018) define a connected environment as one where vehicles share their detailed trajectories through vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communications. Connected vehicles have the ability to share information on their location, speed, acceleration to other vehicles in the environment and receive the same types of information from other connected vehicles.

In a stochastic dynamic network with spatio-temporal dependencies, the travel times realized across the network and over time are highly interdependent. Thus, travel times that will be experienced along the path a particular traveler intends to traverse will depend on travel times experienced in other parts of the network and at earlier points in time. Modeling network link travel time dependencies allows for information regarding travel times across various parts of the network in the past to be used to update the knowledge of future travel time distributions on the links of interest to a given traveler or decision maker.

The previous chapter presents a definition of trajectory information H, as a series of consecutive node-time pairs the traveler has experienced from the origin node i_0 at departure time t_0 up to the current node i and time t: $H = \{(i_0, t_0), (i_1, t_1), \dots, (i, t)\}$. The trajectory information H also contains the revealed travel times along the traversed links.

This chapter assumes that the connected environment traveler has access to connected vehicle (CV) information from all CVs in the network, which traverse actual trajectories during the time of the traveler's trip. This definition assumes that a traveler does not have access to CV information prior to start of their trip and has access to trajectory information from all CVs that is received over time from the start of their trip. Thus, the traveler departing at origin i_0 and time t_0 has access to trajectory information H_v for each connected vehicle v, from the first node j_0 they reach at the time t_{j0} closest to t_0, up to the last node they reached j at the time t_j closest to current time t: $H_v^{t_0 t} = \{(j_0, t_{j0}), (j_1, t_{j1}), \dots, (j, t_j)\}$. Let $\mathcal{V}$ denote the set of all connected vehicles in the network, then the information available to the traveler would be $\mathcal{H}^{t_0 t} = \{H_v^{t_0 t} \ \forall\, v \in \mathcal{V}\}$.

It should be noted that even though $\mathcal{V}$ contains all the connected vehicles in the network, the definition of H_v ensures that each traveler has access only to the CV trajectory information from the traveler's start at origin at time t_0 to the current time t. Hence, any vehicles not actively traveling in the network during that time will deliver no information and any trajectory (or part of a trajectory) that a vehicle traversed prior to time t_0 will be excluded from $\mathcal{H}$.

This definition of information availability in a connected environment is a type of partial information availability that is more general than that from the previous chapter. In fact, access to the traveler's own trajectory information only, as used in Chapter 7, can be seen as a special case of this definition, where the traveler's own vehicle is the only one in the set $\mathcal{V}$. Other special cases of partial information availability can also be framed as special cases of this definition. For example, the information availability in a special neighborhood can be obtained by restricting H_v to contain information only for nodes i that are in the predefined spatial neighborhood. The full information case is also a special case of this definition, when the set $\mathcal{V}$ contains all vehicles in the network, i.e., in the case of a fully connected environment. This definition allows for this problem definition and its solution approach to be generalized for any level or type of information that can be specified under the definition for connected environment information.

8.2.3 Information-Adaptive Reliable Least-Time Routing Problem

This chapter considers the problem of finding information-adaptive reliable-least time routes (IA-RLTR) in stochastic dynamic networks with spatio-temporally correlated link travel times. The problem definition assumes the traveler has access to vehicle trajectory information from a connected environment, denoted $\mathcal{H}$, and defined in the previous section. The IA-RLTR assumes a reactive traveler that considers the information as it arrives and makes a new decision at each decision point. This problem definition differs significantly from that of the proactive traveler defined in the previous chapter. This section will define some of the key concepts for the problem definition and present short examples to illustrate those.

Definition 1. Information-Adaptive Routing Solution: An information-adaptive routing solution ρ can be defined as a mapping from state to decision. The state is defined as the triplet $\{i, t, \mathcal{H}\}$, where i is the current node, t is the current time – equivalent to the time of arrival at the current node, and $\mathcal{H}$ is the current information from the connected environment, as defined above. The action space at state $\{i, t, \mathcal{H}\}$ is $\{j \in N : (i, j) \in A\}$, i.e., the set of nodes adjacent to i, and the decision for which node j to take next: $\rho: \{i, t, \mathcal{H}\} \rightarrow j$.

The information-adaptive routing solution here is defined recursively, similarly to the trajectory-adaptive routing strategy. However, a there are some key differences in determining a strategy (for the proactive traveler) versus a path (for the reactive traveler). In determining which next node to select, i.e., the appropriate mapping $\rho(\{i, t, \mathcal{H}\}) = j$, the IA-RLTR problem selects a single path to the destination, rather than recursively considering the following decision nodes.

Definition 2. Decision Node: A decision node for an information-adaptive routing strategy ρ is a node $i \in N$ such that $|\{j \in N : (i, j) \in A\}| > 1$, namely a node for which there is more than one possible next node that can be chosen.

A key difference relative to the trajectory-adaptive strategy is that the IA-RLTR problem considers one decision node at a time and the next decision node is only known after the decision is made at the current node.

To illustrate these two differences, consider the example network 2 from the previous chapter, shown here in Figure 8-1 with the possible travel time realizations of the joint link travel time distribution given in Table 8-1 and Table 8-2. This example can be contrasted to Example 3 in the previous chapter.

Table 8-1. Possible joint link travel time realizations for example network 2

Realization	Link Travel Times								
	(O,a)	(a,b)	(a,e)	(b,D)	(b,c)	(c,D)	(e,D)	(e,f)	(f,D)
1	1	1	0.5	3	1	1	0.5	0.5	1
2	1	2	1	1.5	1	1	1	1	1.5
3	1	2	1.5	1	1	1	0.5	1.5	1
4	2	1	0.5	2	0.5	1	1	1	1
5	2	2	1	2	2	1	0.5	1	1.5
6	2	2	1	1.5	1	2	1	1.5	1

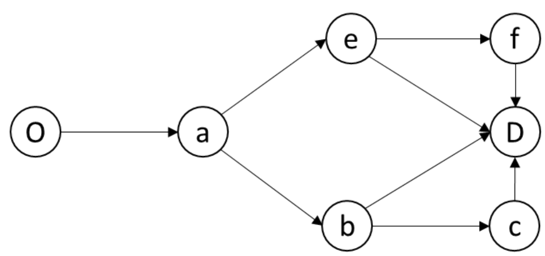

Figure 8-1. Example network 2

Table 8-2. Possible joint link and path travel time realizations for example network 2

Realization	Travel Times						
	(O,a)	(a,b)	(a,e)	*O-a-b-D*	*O-a-b-c-D*	*O-a-e-D*	*O-a-e-f-D*
1	1	1	0.5	5	4	3	3.5
2	1	2	1	4.5	5	5	5.5
3	1	2	1.5	4	5	5	5.5
4	2	1	0.5	5	4.5	4.5	5
5	2	2	1	6	7	5.5	6.5
6	2	2	1	5.5	7	6	6.5

Example 1. Information-Adaptive routing for a Proactive Traveler

Consider a reactive traveler who wants to minimize their expected travel time from O to D. The traveler starts at the origin with a chosen path to destination *O-a-e-D*, that minimizes the expected path travel time with no (additional) information, so their next node is *a*.

Suppose the traveler experiences travel time $t_{Oa} = 2$ on link *O-a* and also receives information from the CV environment that another vehicle traversed link *a-b* with travel time $t_{ab} = 2$. Then, at node a, conditional travel time distributions for all paths can be recomputed resulting in the following expected travel times: $E(t_{OabD}) = 5.75$, $E(t_{OabcD}) = 7$, $E(t_{OaeD}) = 5.75$ and $E(t_{OaefD}) = 6.5$ and the minimum expected travel time is 5.75 for both paths *O-a-b-D* and *O-a-e-D* and the traveler can select either of those. This process is repeated at the next decision node. ■

Thus, in contrast to the proactive traveler that makes a "plan" (i.e., chooses a strategy) to the destination node – considering the different possible outcomes and their likelihood, the reactive traveler simply makes a new choice at each decision node as it is encountered, and information is received.

Some additional terminology that will be used in relation to the IA-RLTR problem is presented here. The initial path is the path chosen at the origin and used to make the first decision for next node, when no information is available to the traveler, and will also be referred to as the a priori path. For each decision node i, the optimal and selected path according to the information available at that node will be referred to as the path at node i. A path will be considered to be 'in the solution' if it was the selected path at any decision node i. A change of path at node i takes place if the selected path at node i is different from the immediate previous selected path. The path realized with all intermediate node decisions, from the origin to the destination, is the a posteriori path and will also be referred to as the final path.

From these definitions, the existence of decision changes at intermediate nodes does not imply that the a posteriori node is different from the a priori node. In example network, the traveler could start with selected path *O-a-e-D,* then at node *a* decide on path *O-a-e-f-D*, then at node *e* select path *O-a-e-D* again. In that case, a change of path occurred at two intermediate nodes, but *O-a-e-D* is both the a priori and a posteriori path.

8.3 Solution Methodology

The proposed solution approach for the information-adaptive reliable least-time routing (IA-RLTR) problem in stochastic dynamic networks with connected environment information access is a 2-stage approach. Firstly, eligible a priori paths are generated using the approach for the RLTP problem in Chapter 6. Secondly, the information-adaptive path finding approach evaluates the paths at each decision node and determines the optimal path to destination based on the available information from the connected environment.

8.3.1 Eligible Paths Generation

The approach for generating eligible non-dominated a priori paths, presented as part of the solution for the RLTP problem in Chapter 6, is summarized here. The approach eliminates dominated paths at intermediate nodes based on two types of dominance criteria: an adjusted First-Order Stochastic Dominance (A-FSD) criterion and a relaxed deterministic dominance criterion (RDD), based on the stochastic dominance and deterministic dominance criteria by Miller-Hooks and Mahmassani (1998b). The RDD criterion was modified with a relaxation parameter ϵ that specifies the allowable level of risk in eliminating a potentially viable path.

8.3.2 Information-Adaptive Path Updating

The second stage of the solution approach uses the Stage 1 eligible paths and travel time distributions in a procedure for information-adaptive path updating. Unlike the previous algorithms, this portion of the solution approach is performed for each destination node, departure time, for a given objective function. The procedure presented below includes the updating of the joint link travel time distributions at each decision node, based on the conditional on the connected environment information delivered at the arrival time at that decision node. In this manner, the algorithm is self-contained. However, since the updating of the travel time distributions is conditional on the information up to the current time, the updating can be performed simultaneously for multiple O-D pairs, which would make the Stage 2 approach significantly more efficient.

Solution Algorithm, Stage 2: Procedure for Information-Adaptive Path Updating

Given:

The network $G(N, A, \mathcal{T})$. The function $\phi(\cdot) \in \mathcal{T}$.
The joint time-varying link travel time distributions with time-varying covariance structure.
The origin node O. Destination $d \in \mathcal{D}$. Departure time $\mathbb{t} \in \mathbb{T}$.
The results from Stage 1, for the selected dominance criterion: the path eligibility indicators $\Lambda_{id}^{k\mathbb{t}}$, the travel time distributions for the set of eligible paths and their links, $U_d^{k\mathbb{t}}$ and $u_i^{k\mathbb{t}}$ and the corresponding vector pointers p_i^k, L_i^k for $i \in N,\ \ k \in \{1, 2, \dots, M\}$.
The objective function f_{obj} for path selection.
The set of connected vehicles $\mathcal{V}$.

Find:

The set of decision nodes, $(i, \mu) \in \Delta_d^{\mathbb{t}}$ with their arrival times $t_{i,\mu}$.
The path to destination for each decision node $\mathcal{P}_i^{\mu} \forall\ (i, \mu) \in \Delta_d^{\mathbb{t}}$ and the corresponding updated travel time distributions $\mathcal{u}_i^{k\mathbb{t}}\ \forall\ i \in N,\ \ k \in \{1, 2, \dots, M\}$.

Step 0: Initialization

Step 0.1: Initialize pointers and labels

Initialize $\Delta_d^{\mathbb{t}} = \{(O, 1)\}$.
Set the arrival time at O, $t_{O,1} = \mathbb{t}$, and set it as the current time $t = t_{O,1}$
$\mathcal{u}_i^{k\mathbb{t}} = u_i^{k\mathbb{t}}\ \forall\ i \in N,\ \ k \in \{1, 2, \dots, M\}$.

Step 0.1: Retrieve all eligible paths.

For the destination node $d \in \mathcal{D}$, from the eligibility indicators $\Lambda^{k\mathbb{t}}\ \forall\ k \in \{1, 2, \dots, M\}$, save the node-path ID pairs and links for each path as P_{nk} and P_{ak} for $k \in K = \{k \mid \Lambda_{dd}^{k\mathbb{t}} = 1\}$ into sets $P_n^{\mathbb{t}}$ and $P_a^{\mathbb{t}}$, respectively.

Step 0.3: A priori path at origin

Find the optimal a priori path P^*, where

$$P^* = \underset{P_{nk} \in P_n^{\mathbb{t}}}{argmin} f_{obj}\left(\sum_{(i,\mu)\in P_{nk}} u_i^{\mu\mathbb{t}}\right)$$

Set $\mathcal{P}_O^1 = P^*$.
Set the current decision node $\delta = (O, 1)$.
Go to Step 1.

Step 1: Find next decision node

Find the path at current decision node δ, $P^* = \mathcal{P}(\delta)$.
Find the next node in $\delta' = P^*$.
Find the set of relevant paths $P_{\delta'} = \{P_{nk} \in P_n^{\mathbb{t}} \mid \delta' \in P_n^{\mathbb{t}}\}$
If the number of unique paths in P_{ak} is 1, go to Step 4.
Otherwise, find the next decision node as the farthest node δ that is in all paths in $P_{\delta'}$.
Set δ as the current decision node. Set $P_\delta = P_{\delta'}$.
Add δ to $\Delta_d^{\mathbb{t}}$
Go to Step 2.

Step 2: Distribution updating

Sample a travel time to δ via $u_i^{k\mathbb{t}}$, let the arrival time t_δ be set as the current time $t = t_\delta$.
Retrieve the CV trajectory information $\mathcal{H}^{\mathbb{t}t_\delta} = \{H_v^{\mathbb{t}t_\delta}\ \forall\ v \in \mathcal{V}\}$.
Find the set of links in the relevant paths as $A' \subset A, A' = \{(i,j) \in P\ \forall\ P \in P_\delta\}$.
Update the travel time distributions $\mathcal{u}_i^{k\mathbb{t}} = u_i^{k\mathbb{t}} | \mathcal{H}^{\mathbb{t}t_\delta}$.
Go to Step 3.

Step 3: Decision node evaluation

Find the optimal a priori path P^*, where

$$P^* = \underset{P \in P_\delta}{argmin} f_{obj}\left(\sum_{(i,\mu)\in P} \mathcal{u}_i^{\mu\mathbb{t}}\right)$$

Set $\mathcal{P}(\delta) = P^*$.
Go to Step 1.

Step 4: Termination

Return the set of decision nodes, $(i,\mu) \in \Delta_d^{\mathbb{t}}$ with their arrival times $t_{i,\mu}$.
The path to destination for each decision node $\mathcal{P}_i^{\mu} \forall\ (i,\mu) \in \Delta_d^{\mathbb{t}}$ and the corresponding updated link travel time distributions $\mathcal{u}_i^{k\mathbb{t}}\ \forall\ i \in N,\ \ k \in \{1, 2, \dots, M\}$.

At termination the procedure returns the decision nodes with the corresponding arrival times and the selected paths at each decision node. It should be noted that the updating of the future links travel times is done at each node with the new available information. Thus, the travel time distributions and objective values for the possible paths are becoming more certain at each next decision node. Thus, travel times or objective function values computed at earlier decision nodes cannot be directly compared to those computed at later nodes. The path updating procedure also returns the path selected at each decision node $\mathcal{P}_i^{\mu} \forall\ (i, \mu) \in \Delta_d^{\mathfrak{t}}$ and the final updated travel time distributions $u_i^{k\mathfrak{t}}\ \forall\ i \in N,\ \ k \in \{1, 2, \ldots, M\}$ with the most up to date information.

8.4 Numerical Experiments

This section presents the numerical experiments designed to evaluate the performance of the solution methodology for the TA-RLTS problem in this chapter. The network and data used for the experiments are same as those for the experiments in Chapter 7, using the same 25 scenarios for the simulations, the 7:00 to 10:00 a.m. peak period, along with the same origin node and departure times. This section describes the design of the numerical experiments in section 8.4.1, and the results and analysis are presented in section 8.4.2.

8.4.1 Design of Numerical Experiments

The numerical experiments were designed to answer a number of research questions regarding the IA-RLTR problem considered in this chapter and to evaluate the performance of the solution approach, in terms of efficiency and accuracy, across the various dominance criteria for the path generation and different objective functions for selecting the optimal path.

8.4.1.1 Research Questions and Performance Measures

This section outlines the research questions asked for these numerical experiments and the performance measures that can be used to answer them. Two sets of research questions are the basis of the design for these numerical experiments: the first set of questions presented below are concerned with the performance of the solution approach and approximations, while the second set of questions are regarding the IA-RLTR problem and the impact of information relative to the a priori RLTP problem.

Regarding the performance of the solution approaches, the following questions are addressed via the numerical experiments:

- How does the performance of the solution approach vary with the approximate dominance criteria?
  - To be evaluated relative to the exact solution case via number of changes in selected path, the change in the objective function values and run times.
- How does the performance of the solution for the approximate criteria vary across the different objective functions?
  - To be evaluated via the change in the objective function values of optimal solutions for each objective function.
- How does the effect of selecting the a posteriori versus a priori path vary for different approximate dominance criteria?
  - To be evaluated via the objective values for optimal a posteriori paths compared optimal a priori paths, with the different approximate dominance criteria.

Regarding the IA-RLTR problem and the solutions obtained in these numerical experiments, the experiments were designed to answer the following questions:

- What is the effect of connected vehicle information access?
  - To be evaluated via the number of path changes relative to the number of decision nodes for each solution, for different objective functions and different levels of information availability.

- What is the effect of selecting the a posteriori compared to an a priori path solution?
    - To be evaluated via the objective values for optimal a posteriori paths compared optimal a priori paths, for different objective functions and information availability levels.

To answer these research questions, the applicable performance characteristics and measures include actual run times, numbers of decision and branching nodes for optimal strategies, and percent relative difference to measure change in objective values.

8.4.1.2 Experimental Design

The experimental design for this study consists of two sets of numerical experiments corresponding to the two sets of research questions presented above.

To evaluate the performance of the solution algorithm with different dominance criteria and objectives, the percentage of CVs was fixed to $p = 20\%$ and the solution approach was tested with all 7 dominance criteria: the RDD with 6 ϵ values, $\epsilon \in \{0, 0.01, 0.05, 0.1, 0.15, 0.2\}$ and the A-FSD criterion, and for the six objective functions from the previous chapter were also used here: the Value at Risk (VaR) and Conditional Value at Risk (CVaR), each with three values for the α-percentile tail, $\alpha \in \{0.7, 0.8, 0.9\}$.

To evaluate the IA-RLTR problem and the impact of information relative to the RLTP problem, the solution approach was tested by setting the dominance criterion in Stage 1 of the solution approach to the RDD criterion with $\epsilon = 0$, which leads to exact path finding solutions. To test the impact of information level, the level of connectivity in the network was varied by varying the percentage of CVs in the network with $p \in \{10\%, 20\%, 30\%, 40\%, 50\%\}$, all to be compared to the base case of $p = 0\%$ which is equivalent to simply solving the RLTP problem from Chapter 6. To evaluate the effect with respect to the different objectives mentioned above.

Each of these tests were performed by solving the IA-RLTS problem for a single origin node and 5 departure times, same as that in Chapter 7, and for a set of 250 randomly selected destination nodes. Thus, 45,000 tests were performed for the first set of experiments and 52,500 tests for the second set of experiments, for a total of 97,500 tests.

8.4.1.3 Simulation of the Connected Environment

An important component of the numerical experiments for this chapter is the simulation of a connected environment to model the connected vehicle information access. To simulate the connected environment, one of the 25 simulated scenarios used for modeling the travel time distributions was selected as the 'realized scenario' and its data for the morning peak period from 7 to 10 a.m. were considered. Additionally, to test for the impact of information, the portion of connected vehicles (CVs) in the network were varied. These numerical experiments considered five cases, where CVs were assumed to make up $p \in \{10\%, 20\%, 30\%, 40\%, 50\%\}$ of all vehicles in the network. To ensure that the numerical results are more generalizable, these experiments used 5 different sets of samples for CV data. Namely, 5 different randomly selected sets of vehicles were used for each of the CV penetration levels from 10% to 50%.

As part of setting up the numerical experiments, the 5 sets of CVs in the network were pre-selected to ensure the same vehicles were delivering the same information across all numerical experiment cases. For each set vehicles were randomly selected with a few key rules. For consistency in the distribution of information over time, for each value of p, the corresponding percent of vehicles were selected from those starting their trip in each 30-minute window of the morning peak period. Additionally, since different vehicles may travel different trajectories and deliver varying quantities of data, the set of CVs was built so that the vehicles selected for $p = 10\%$ were a strict subset of those selected for $p = 20\%$, which were in turn a strict subset of those selected for $p = 30\%$ and so on.

A final note should be added regarding the design of experiments with different levels of connectivity. Since path travel time distributions are conditioned only on the available information for each value of p, the conditional travel time distributions may vary from one case to the next. Namely, the higher the level of connectivity (i.e., as p increases) the more information is available and the conditional travel time distributions become more accurate. Therefore, path solutions obtained with different levels of information cannot be directly compared with their travel time distributions at the given information level. In these experiments, the solutions with varying information levels were compared with the most updated conditional distributions at the highest information level.

8.4.2 Results and Analysis

The numerical results and their analysis are separated into two sections, corresponding to the two groups of research questions presented in the previous section. The results on the performance of the IA-RLTR approach with the approximate dominance criteria are presented in section 8.4.2.1, followed by results on the IA-RLTR solutions with varying levels of information and their comparison to a priori path solutions in section 8.4.2.2.

8.4.2.1 Performance of the IA-RLTR Solution Approach and Approximations

This section considers the performance of the solution approach with various approximate dominance criteria and across six different objective functions, for the case of CV penetration of $p = 20\%$ in the stochastic dynamic network.

The results presented here show average values, across all path solutions for the 250 O-D pairs with 5 departure times and for 6 different objective functions, i.e., a total of 3000 path solutions for each of the 7 dominance criteria. Table 8-3 shows some of the overall solution characteristics, including the run time (RT) in seconds, the percent updated paths, the number of decision nodes (DNs) per path, percent decision nodes (DNs) with change and run time (RT) per decision node (DN) in seconds.

Table 8-3. Overall solution characteristics with different dominance criteria

Dominance Criterion	RT (s)	Updated Paths (%)	DNs per Path	DNs with Change (%)	RT per DN (s)
RDD, $\epsilon = 0$	3284.47	56.00%	3.92	20.46%	697.23
RDD, $\epsilon = 0.01$	2110.71	52.33%	3.20	20.41%	546.07
RDD, $\epsilon = 0.05$	1619.52	47.33%	2.92	19.76%	459.00
RDD, $\epsilon = 0.1$	1367.64	46.17%	2.67	19.63%	423.79
RDD, $\epsilon = 0.15$	1311.62	44.17%	2.51	19.41%	422.00
RDD, $\epsilon = 0.2$	1040.88	43.33%	2.33	19.34%	368.38
A-FSD	1011.25	40.67%	2.26	18.14%	366.82

The run time and run time per decision node increase with the approximate solutions, as do the number of decision nodes per path. The number of decision nodes at which a change occurs also decreases slightly. These results can also be seen in Figure 8-2, where the impact of approximation on the computational effort is visualized.

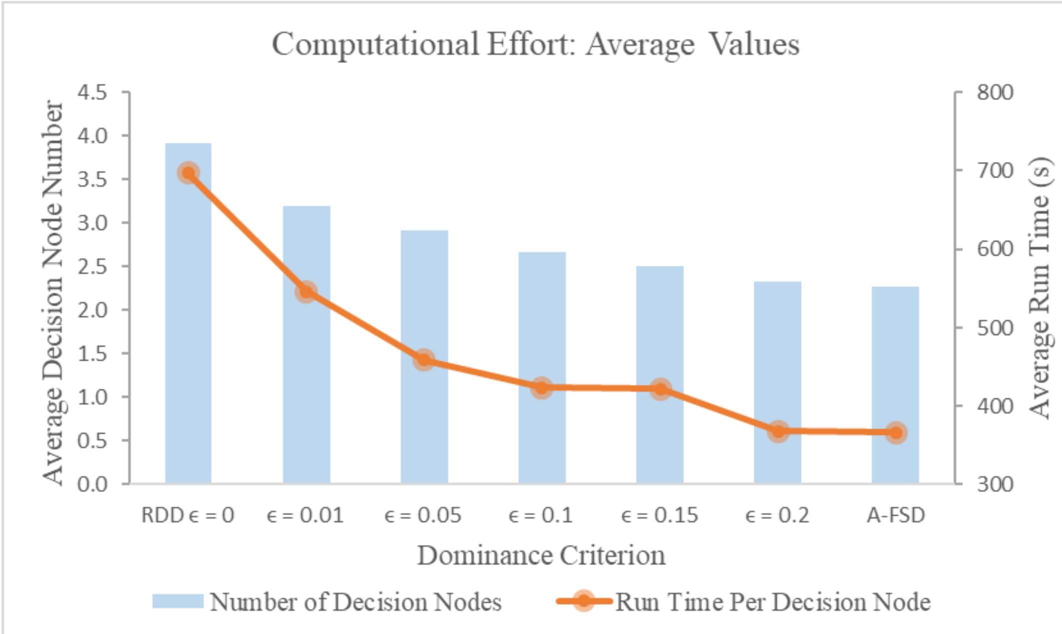

Figure 8-2. Average number of decision nodes and run time per decision node with different dominance criteria

Figure 8-2 shows that the average run time can decrease drastically by introducing approximations. Simply using the RDD criterion with $\epsilon = 0.01$ yields a 35% decrease in run time, while using the A-FSD criterion reduces the average run time by 69.2%. These time savings are in part associated with the number of decision nodes, but as the run time per decision node also decreases the time savings can also be due to the number of possible paths to be evaluated at each decision node.

This effect of using approximate solutions translates to the effectiveness of the obtained solution. Figure 8-3 shows the percent of updated paths and percent of decision nodes with path updates. Updated paths are considered paths in which a path change was made for at least one decision node, meaning that the availability of information led to the solution being updated from the a priori path. With the exact approach 56% of the paths were updated, and that number decreases down to 40.67% with the approximate solutions. This is an interesting result, showing that a large portion of the updated paths in the exact solution are still updated with approximation, and the run time savings close to 70% were achieved by reducing the updated paths by 15.33 percentage points. However, it is important to consider how these updates to the path translate into travel time improvements.

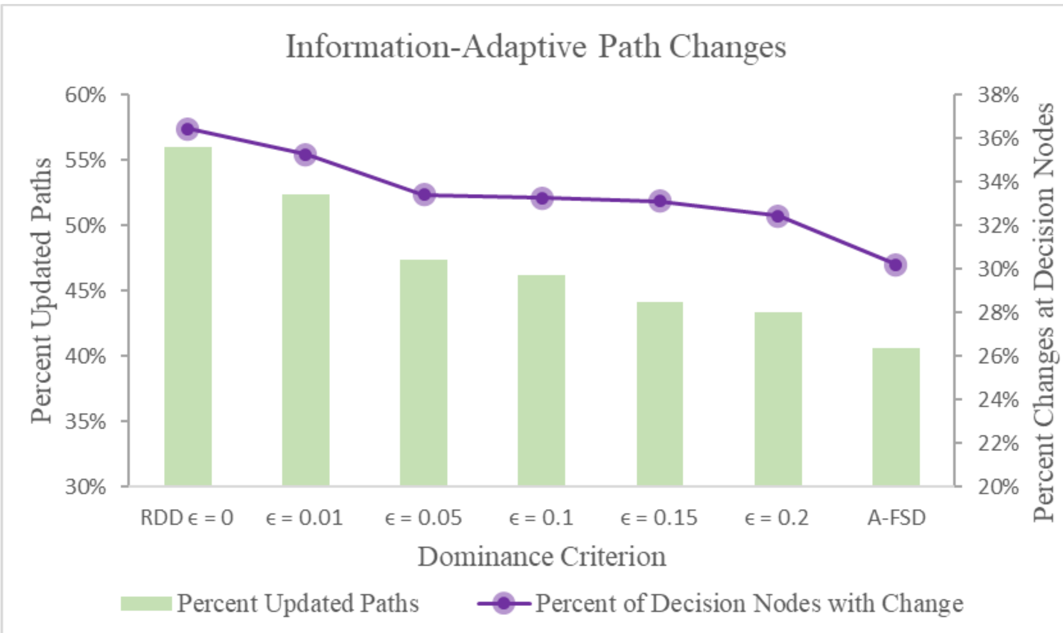

Figure 8-3. Percent updated paths and percent of decision nodes with path change with different dominance criteria

Overall results on travel time savings with path updating are presented in Table 8-4, including the average and maximum values of time savings in minutes and percent (i.e., percent relative difference) compared to the a priori solution.

The travel time savings across all paths and all objectives were on average 11.31% and up to 82.74% for the exact solution and decrease down to 6.48% on average and 76.99 % at maximum for the A-FSD criterion. The average values from Table 8-4 are also presented visually in Figure 8-4.

Table 8-4. Overall travel time savings with path updating for different dominance criteria

Dominance Criterion	Average Values		Maximum Values	
	Time Saved (min)	Percent Time Saved (%)	Time Saved (min)	Percent Time Saved (%)
RDD, $\epsilon = 0$	4.501	11.31%	109.24	82.74%
RDD, $\epsilon = 0.01$	3.958	10.04%	108.40	82.55%
RDD, $\epsilon = 0.05$	2.167	8.47%	71.47	82.76%
RDD, $\epsilon = 0.1$	1.993	7.79%	71.22	82.81%
RDD, $\epsilon = 0.15$	1.949	7.65%	71.34	82.92%
RDD, $\epsilon = 0.2$	1.745	7.04%	69.24	76.99%
A-FSD	1.602	6.48%	69.20	76.99%

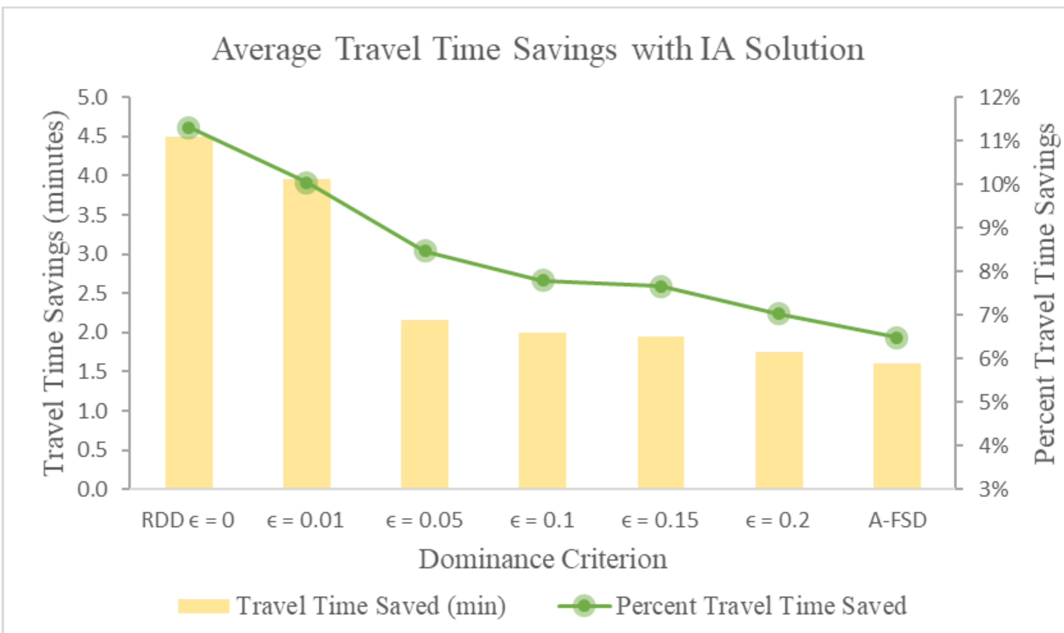

Figure 8-4. Average raw and percent travel time savings for updated paths

Looking at the average percent travel time savings, it is important to note the effect of small levels of approximation. Namely, considering the RDD criterion with $\epsilon = 0.01$, the average travel time savings reduce from 11.31% to 10.04%, i.e., close to 1 percentage point, and the maximum travel time savings reduce by less than 0.2 percentage point. However, this approximate solution was obtained with a 35% shorter run time relative to the exact solution, as was shown in Table 8-3 and Figure 8-2. Similarly, in terms of minutes of travel time savings, introducing a low level of risk, here $\epsilon = 0.01$ did not significantly impact the solution quality. To compare the performance of these approaches across different objective functions, the average travel time savings and percent travel time savings are shown in Table 8-5, where again the VaR and CVaR objective types are abbreviated as V and C, respectively.

Table 8-5. Travel time savings for different dominance criteria and objective functions

	Objective:	$V(\alpha = 0.7)$	$V(\alpha = 0.8)$	$V(\alpha = 0.9)$	$C(\alpha = 0.7)$	$C(\alpha = 0.8)$	$C(\alpha = 0.9)$
		Average Travel Time Saving (min)					
Dominance Criterion	RDD $\epsilon = 0.01$	3.68	6.14	4.37	4.22	4.38	4.36
	RDD $\epsilon = 0.05$	2.92	5.64	3.82	4.07	4.01	3.50
	RDD $\epsilon = 0.1$	1.69	3.16	1.72	2.11	2.09	2.30
	RDD $\epsilon = 0.15$	1.43	3.18	1.56	1.67	1.75	2.17
	RDD $\epsilon = 0.2$	1.39	3.21	1.49	1.90	1.82	2.20
	A-FSD	1.32	2.79	1.08	1.57	1.67	2.06
		Average Percent Travel Time Saving					
Dominance Criterion	RDD $\epsilon = 0.01$	10.45%	13.10%	11.16%	11.21%	11.16%	10.96%
	RDD $\epsilon = 0.05$	9.14%	11.75%	9.47%	10.36%	9.70%	9.94%
	RDD $\epsilon = 0.1$	7.88%	9.60%	7.32%	8.78%	8.60%	8.67%
	RDD $\epsilon = 0.15$	7.00%	8.98%	6.90%	7.93%	7.86%	8.22%
	RDD $\epsilon = 0.2$	6.62%	8.67%	6.66%	7.90%	7.83%	8.11%
	A-FSD	6.29%	8.29%	5.45%	7.17%	7.30%	7.68%

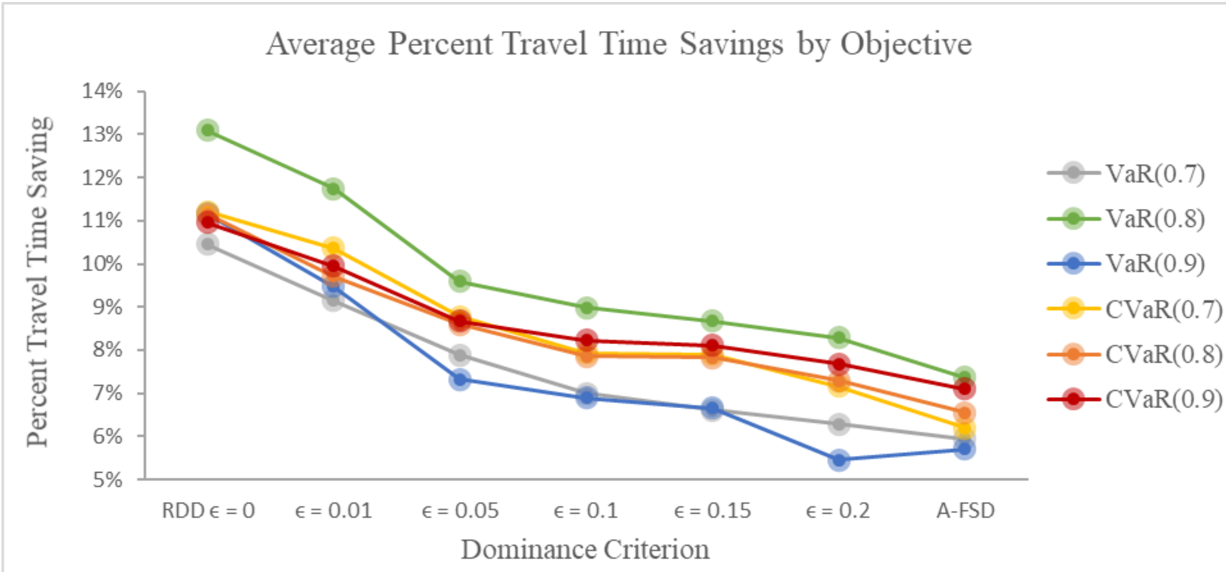

Figure 8-5. Average percent travel time saved for different dominance criteria and objective functions

It can be seen that there are no significant differences in the time savings for different objective functions, but the overall trend remains of decreasing time and percent travel time savings with the use of approximate dominance criteria. The average percent travel time savings for each objective function are shown in Figure 8-5, where it can be observed that the largest savings were achieved for the VaR objective with $\alpha = 0.8$, followed by the CVaR objectives, then the remaining VaR objectives. The trend of reducing percent travel time savings with the approximate dominance criteria can also be observed.

On the whole, these results demonstrate the potential of approximate dominance criteria in improving the computational efficiency of the solution approach and illustrate the trade-off between efficiency and solution quality. The results also show that using the RDD criterion with low values for the risk-tolerance level ϵ, here the lowest value tested being $\epsilon = 0.01$ can maintain the solution quality, relative to stronger approximations, while also yielding significant savings in computational effort.

8.4.2.2 Evaluation of Information-Adaptive Routing in a Connected Environment

This section presents the results on the IA-RLTR solutions with varying levels of information and considers how they compare to a priori solutions to assess the impact of information access. This set of experiments were performed for a single dominance criterion, the RDD with $\epsilon = 0$, which leads to exact solutions. It should be noted that these experiments were conducted separately from those in section 8.4.2.1 above, so their results are not compared directly.

In this portion of the numerical experiments, the effect of increasing information availability via increased penetration of connected vehicles in the connected environment is assessed via the percent of updated paths and their raw and percentage travel time savings in terms of the objective function value. These results are always compared relative to the original path solution, which is equivalent to the optimal a priori path from the RLTP problem. The general results, including the percent of updated paths, and average and maximum travel time savings are shown in Table 8-6.

Table 8-6. Impact of information-adaptive solution with different CV penetration levels

CV penetration (%)	Updated Paths (%)	Average Values		Maximum Values	
		Time Saved (min)	Percent Time Saved (%)	Time Saved (min)	Percent Time Saved (%)
10%	54.00%	2.68	7.63%	66.10	80.14%
20%	56.00%	3.29	8.21%	66.10	80.14%
30%	56.53%	3.30	8.96%	66.10	80.14%
40%	56.80%	3.50	9.44%	66.10	80.14%
50%	58.07%	4.16	11.28%	66.10	80.14%

The results in Table 8-6 show an increase in the number of updated paths as information availability increases with higher CV penetration level, from 54% to 58.07%. The average travel time savings, relative to the a priori solution, both in minutes and percent travel time increase with the increase of information, but the maximum travel time savings are constant across the 5 cases, indicating that they were achieved with the lowest level of information. Thus, the 10% CV penetration provided sufficient information for 54% of the paths to be updated, on average with 7.63% in travel time savings per updated path, and the subsequent increase in CV penetration by an additional 40 percentage points resulted in an increase in updated paths by only 4.07 percentage points and the percent travel time savings by 3.65 percentage points. Overall, this indicates that the average travel time savings across all considered paths, including those not updated at all, were 4.12% for the CV penetration level at 10% and 6.55% for the CV penetration level at 50%.

Considering the travel time savings for different objective function, shown in Table 8-7, the same general trend is observed, but no significant differences across the different objectives. From Table 8-7, the highest average travel time savings are achieved at the highest level of CV penetration for each objective. The difference in travel time savings from the 10% penetration level to 50% varies slightly across the different objectives, with the largest difference at 4.58 percentage points for the VaR ($\alpha = 0.8$) objective.

Table 8-7. Percent travel time savings for IA solutions for different objectives

	Objective:	V($\alpha = 0.7$)	V($\alpha = 0.8$)	V($\alpha = 0.9$)	C($\alpha = 0.7$)	C($\alpha = 0.8$)	C($\alpha = 0.9$)
% CV penetration	10%	4.25%	6.03%	7.48%	8.13%	9.91%	10.01%
	20%	4.93%	7.58%	7.79%	8.38%	10.42%	10.15%
	30%	7.06%	7.88%	8.05%	9.06%	11.11%	10.62%
	40%	7.16%	8.05%	9.21%	9.37%	11.77%	11.08%
	50%	8.16%	10.62%	11.99%	11.34%	13.34%	12.25%

These results show that the effect of access to information is important when moving from the a priori to a low level of information access, such as the 10% CV penetration level. However, the marginal improvements for increased CV penetration are relatively low. Of course, the impact of information access and CV penetration levels can vary due to a range of other factors not accounted for in these experiments, such as the distribution of CV trips across the network and over time. Such questions may provide interesting areas for future research.

8.5 Conclusion and Future Work

This chapter addresses the problem of finding optimal information-adaptive reliable least-time routes (IA-RLTR) in stochastic dynamic networks with spatio-temporal dependencies. The problem setting consists of a stochastic dynamic network that is also a connected dynamic environment. This chapter presents a general definition for information access in a connected vehicle environment and defines the IA-RLTR problem for a reactive traveler. A two-stage solution approach is presented, using the path generation algorithm for a priori paths from Chapter 6 and an information-adaptive path updating approach.

Numerical experiments on the large-scale Chicago network tested for the characteristics of the problem at hand and the performance of the solution approach. The experiments tested the performance of the approach with different dominance criteria and for different objective functions, and compared the effect of information access at different connected vehicle penetration levels. The experimental results show the trade-off between computational effort and solution quality in the IA-RLTR problem. Using approximate dominance criteria allows for reduced computational run times at the expense of finding the optimal solution and specifically in the travel time savings of the IA solution relative to a priori paths. However, the results show that controlling the level of approximation via the risk-level tolerance criterion ϵ allows for finding a balance between the two at low values for ϵ. Testing for different levels of information access by varying the level of CV penetration in the connected environment, the numerical results show the significant impact of information access at low CV penetration levels and small marginal increases as the CV penetration increases.

Several questions related to this study may be interesting for future research. Firstly, regarding the approximate solution approaches, future work may evaluate the impact of approximations and their performance in networks of different size or type. Secondly, in terms of the information availability and CV penetration levels, it may be interesting to investigate the effect of information based on the types of trips the CVs take and their distribution across space and time. Thirdly, a larger problem that can be considered in future work may be to combine the proactive (i.e., strategic) traveler approach from the TA-RLTS problem in the previous chapter with the reactive traveler approach from the problem in this chapter. In such a problem, a traveler could choose a strategy that is updated at each node based on incoming CV information. This problem would also allow for a comparison of the trajectory- and information-adaptive problems to evaluate the impact of the traveler's own trajectory versus CV information in achieving improvements in travel time.

Chapter 9 Concluding Remarks

9.1 Summary and Contributions

This dissertation focuses on addressing modeling and optimization problems in stochastic dynamic networks. The motivation for this topic arises both from the importance of reliability as a factor in evaluation and decision-making in transportation networks and from the need for a cohesive framework to approach optimization problems in this complex setting. One of the main objectives of this dissertation is to present a comprehensive and cohesive set of approaches for modeling, estimation, and optimization in stochastic dynamic networks.

The broad contributions of this dissertation include the characterization of stochastic dynamic networks in a data-driven and application-oriented manner, presenting approaches for the estimation of path travel time distributions, and defining and solving path finding problems for reliable least-time routing. The routing problems presented in this dissertation include the problem of a priori reliable least-time paths in Chapter 6, trajectory-adaptive reliable least-time strategies in Chapter 7 and information-adaptive reliable least-time routing in Chapter 8. The dissertation presents unifying solution approaches that can be used for solving all three of these problem types that also utilize the modeling and estimation approaches from the previous chapters. Additional and more specific contributions are presented in the overview sections for each chapter.

9.2 Applications

The problems studied in this dissertation have several important application areas. The characterization of stochastic transportation networks can be applied for performance measurement, performance monitoring and simulation modeling in the context of transportation policies, projects and applications concerned with the reliability performance of transportation systems. The estimation of path travel time distributions is a key aspect of assessing transportation networks from the user perspective, evaluating the transportation system by considering entire paths or trajectories and accounting for the importance of travel time reliability.

Reliable path finding problems have a host of applications, besides the most apparent problem of individual traveler routing. Reliability-based decision making is more important at larger scales, for applications such as freight or mobility service providers where lack of reliability translates to economic cost. Emerging transportation technology and service advancements may be even more concerned with reliable routing, such as the cases of electric vehicles, autonomous vehicles, ride-sourcing companies, etc. Emerging transportation data also call for making reliability-based decision-making adaptive to information. With the increased reliance on traveler information services and navigation systems, and the availability of trajectory data via geographical positioning systems (GPS) or connected vehicles (CVs), real-time information access is becoming more ubiquitous and the ability to use that information and respond to it is also becoming more important.

9.3 Future Research Areas

This dissertation, along with its results, analysis, and discussion, opens up several additional questions and considerations that provide a basis for future work. For the characterization of stochastic dynamic networks, more sophisticated approaches for community structures and change point detection can be applied, such as ones where change point detection can be performed on the community structure itself.

In terms of path finding and routing problems, solving the problem with recourse, and understanding the effect of allowing cyclic paths are worthy of further study. Additional computational tests may be needed in order to understand the performance of the solution algorithm and heuristic for different sizes and types of networks.

An interesting problem for future work may be to combine the proactive (i.e., strategic) traveler approach with the reactive traveler perspective to compare the two and potentially evaluate the impact of the traveler's own trajectory versus CV information. In terms of the information availability, future research may investigate the effect of information based on the type of trips for CVs and their distribution across space and time.

REFERENCES

Agrawal, R., Faloutsos, C., Swami, A., 1993. Efficient similarity search in sequence databases, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/3-540-57301-1_5

Aminikhanghahi, S., Cook, D.J., 2017. A survey of methods for time series change point detection. Knowl. Inf. Syst. https://doi.org/10.1007/s10115-016-0987-z

Arlot, S., Celisse, A., Harchaoui, Z., 2019. A Kernel Multiple Change-point Algorithm via Model Selection. J. Mach. Learn. Res.

Bai, J., 1995. Least absolute deviation estimation of a shift. Econom. Theory. https://doi.org/10.1017/S026646660000935X

Berndt, D., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. Work. Knowl. Knowl. Discov. Databases.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. https://doi.org/10.1088/1742-5468/2008/10/P10008

Cario, M.C., Nelson, B.L., 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Ind. Eng.

Cario, M.C., Nelson, B.L., 1996. Autoregressive to anything: Time-series input processes for simulation. Oper. Res. Lett. https://doi.org/10.1016/0167-6377(96)00017-X

Chakravarty, I.M., Roy, J.D., Laha, R.G., 1967. Handbook of methods of applied statistics.

Chen, B.Y., Lam, W.H.K., Sumalee, A., Li, Q., Shao, H., Fang, Z., 2013. Finding Reliable Shortest Paths in Road Networks Under Uncertainty. Networks Spat. Econ. https://doi.org/10.1007/s11067-012-9175-1

Chen, B.Y., Lam, W.H.K., Sumalee, A., Li, Z. lin, 2012. Reliable shortest path finding in stochastic networks with spatial correlated link travel times. Int. J. Geogr. Inf. Sci. https://doi.org/10.1080/13658816.2011.598133

Chen, B.Y., Li, Q., Lam, W.H.K., 2016. Finding the k reliable shortest paths under travel time uncertainty. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2016.09.013

Chen, M., Yu, G., Chen, P., Wang, Y., 2017. A copula-based approach for estimating the travel time reliability of urban arterial. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2017.06.007

Chen, P., Tong, R., Lu, G., Wang, Y., 2018. The α-reliable path problem in stochastic road networks with link correlations: A moment-matching-based path finding algorithm. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2018.05.022

Chen, P., Tong, R., Yu, B., Wang, Y., 2020. Reliable shortest path finding in stochastic time-dependent road network with spatial-temporal link correlations: A case study from Beijing. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2020.113192

Chen, X., Osorio, C., 2014. Analytical formulation of the trip travel time distribution. Transp. Res. Procedia 3, 366–373.

Cherifi, H., Palla, G., Szymanski, B.K., Lu, X., 2019. On community structure in complex networks: challenges and opportunities. Appl. Netw. Sci. https://doi.org/10.1007/s41109-019-0238-9

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Am. Stat. 49, 327–335.

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. https://doi.org/10.1103/PhysRevE.70.066111

Daniel, W.W., 1990. Kolmogorov–Smirnov one-sample test. Appl. nonparametric Stat. 2.

Dowd, K., 2007. Measuring market risk. John Wiley & Sons.

Eiger, A., Mirchandani, P.B., Soroush, H., 1985. Path Preferences and Optimal Paths in Probabilistic Networks. Transp. Sci. https://doi.org/10.1287/trsc.19.1.75

Eisele, W.L., Naik, B., Rilett, L.R., 2015. Estimating route travel time reliability from simultaneously collected link and route vehicle probe data and roadway sensor data. Int. J. Urban Sci. 19, 286–304.

Elfar, A., Talebpour, A., Mahmassani, H.S., 2018. Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment. Transp. Res. Rec. 2672, 185–195. https://doi.org/10.1177/0361198118795010

Epps, T.W., Singleton, K.J., 1986. An Omnibus Test for the Two- Sample Problem Using the Empirical Characteristic Function. J. Stat. Comput. Simul.

https://doi.org/10.1080/00949658608810963

Fakhrmoosavi, F., Zockaie, A., Abdelghany, K., Hashemi, H., 2019. An iterative learning approach for network contraction: Path finding problem in stochastic time-varying networks. Comput. Civ. Infrastruct. Eng. 34, 859–876. https://doi.org/10.1111/MICE.12460

Fakhrmoosavi, F., Zockaie, A., Abdelghany, K., Hashemi, H., 2018. Decomposition of Stochastic Time-Varying Networks for the Path-Finding Problem Considering Travel Time Correlations and Heterogeneity of Users, in: Transportation Research Board 97th Annual Meeting.

Fan, Y.Y., Kalaba, R.E., Moore, J.E., 2005. Shortest paths in stochastic networks with correlated link costs. Comput. Math. with Appl. https://doi.org/10.1016/j.camwa.2004.07.028

Fenton, L.F., 1960. The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems. IRE Trans. Commun. Syst. https://doi.org/10.1109/TCOM.1960.1097606

Filipovska, M., Mahmassani, H.S., 2021. Computation and Estimation of Path Travel Time Variability with Sparse Vehicle Trajectory Data, in: International Symposium on Transportation Data and Modelling.

Filipovska, M., Mahmassani, H.S., 2020a. Traffic flow breakdown prediction using machine learning approaches. Transp. Res. Rec. 2674, 560–570.

Filipovska, M., Mahmassani, H.S., 2020b. Reliable Least-Time Path Estimation and Computation in Stochastic Time-Varying Networks with Spatio-Temporal Dependencies, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020. https://doi.org/10.1109/ITSC45102.2020.9294650

Filipovska, M., Mahmassani, H.S., 2020c. Reliable least-time path estimation and computation in stochastic time-varying networks with spatio-temporal dependencies, in: 99th Annual Meeting of the Transportation Research Board.

Filipovska, M., Mahmassani, H.S., Mittal, A., 2021. Estimation of Path Travel Time Distributions in Stochastic Time-Varying Networks with Correlations.

Filipovska, M., Mahmassani, H.S., Mittal, A., 2019. Prediction and Mitigation of Flow Breakdown Occurrence for Weather Affected Networks: Case Study of Chicago, Illinois. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/0361198119851730

Foreman-Mackey, D., Hogg, D.W., Lang, D., Goodman, J., 2013. emcee: the MCMC hammer. Publ. Astron. Soc. Pacific 125, 306.

Fosgerau, M., Fukuda, D., 2012. Valuing travel time variability: Characteristics of the travel time distribution on an urban road. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2012.02.008

Frank, H., 1969. Shortest Paths in Probabilistic Graphs. Oper. Res. https://doi.org/10.1287/opre.17.4.583

Fryzlewicz, P., 2007. Unbalanced Haar technique for nonparametric function estimation. J. Am. Stat. Assoc. https://doi.org/10.1198/016214507000000860

Gao, S., 2004. Optimal adaptive routing and traffic assignment in stochastic time-dependent networks. October.

Gao, S., Chabini, I., 2006. Optimal routing policy problems in stochastic time-dependent networks. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2005.02.001

Gao, S., Huang, H., 2012. Real-time traveler information for optimal adaptive routing in stochastic time-dependent networks. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2011.09.007

Gelman, A., 1993. Iterative and non-iterative simulation algorithms. Comput. Sci. Stat.

Goodman, J., Weare, J., 2010. Ensemble samplers with affine invariance. Commun. Appl. Math. Comput. Sci. 5, 65–80.

Hall, R.W., 1986. FASTEST PATH THROUGH A NETWORK WITH RANDOM TIME-DEPENDENT TRAVEL TIMES. Transp. Sci. https://doi.org/10.1287/trsc.20.3.182

Holton, G.A., 2012. Value-at-Risk: Theory and Practice, e-book 2014. URL http//value-at-risk. net.

Huang, H., Gao, S., 2018. Trajectory-adaptive routing in dynamic networks with dependent random link travel times. Transp. Sci. https://doi.org/10.1287/trsc.2016.0691

Huang, H., Gao, S., 2012. Optimal paths in dynamic networks with dependent random link travel times. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2012.01.005

Hunter, T., Hofleitner, A., Reilly, J., Krichene, W., Thai, J., Kouvelas, A., Abbeel, P., Bayen, A., 2013. Arriving on time: estimating travel time distributions on large-scale road networks. arXiv Prepr. arXiv1302.6617.

Isukapati, I., List, G., Williams, B., Karr, A., 2013. Synthesizing route travel time distributions from segment travel time distributions. Transp. Res. Rec. https://doi.org/10.3141/2396-09

Ji, Z., Kim, Y.S., Chen, A., 2011. Multi-objective α-reliable path finding in stochastic networks with correlated link costs: A simulation-based multi-objective genetic algorithm approach (SMOGA). Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2010.07.064

Karrer, B., Levina, E., Newman, M.E.J., 2008. Robustness of community structure in networks. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. https://doi.org/10.1103/PhysRevE.77.046119

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S., 2001a. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowl. Inf. Syst. https://doi.org/10.1007/pl00011669

Keogh, E., Chu, S., Hart, D., Pazzani, M., 2001b. An online algorithm for segmenting time series. Proc. - IEEE Int. Conf. Data Mining, ICDM. https://doi.org/10.1109/icdm.2001.989531

Keogh, E., Ratanamahatana, C.A., 2005. Exact indexing of dynamic time warping. Knowl. Inf. Syst. https://doi.org/10.1007/s10115-004-0154-9

Keogh, E.J., Pazzani, M.J., 2000. Scaling up dynamic time warping for datamining applications, in: Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/347090.347153

Kim, J., Mahmassani, H.S., 2015. Compound Gamma representation for modeling travel time variability in a traffic network. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2015.06.011

Kim, J., Mahmassani, H.S., 2014a. A finite mixture model of vehicle-to-vehicle and day-to-day variability of traffic network travel times. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2014.05.011

Kim, J., Mahmassani, H.S., 2014b. How many runs? Analytical method for optimal scenario sampling to estimate travel time variance in traffic networks. Transp. Res. Rec. https://doi.org/10.3141/2467-06

Last, M., Kandel, A., Bunke, H., 2004. Data mining in time series databases, Series in Machine Perception and Artificial Intelligence 57.

Li, X., Schissler, A.G., Wu, R., Barford, L., Harris, F.C., 2019. A graphical processing unit accelerated normal to anything algorithm for high dimensional multivariate simulation, in: Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-3-030-14070-

0_46

Lung-Yut-Fong, A., Lévy-Leduc, C., Cappé, O., 2015. Homogeneity and change-point detection tests for multivariate data using rank statistics. J. la Société Française Stat. 156, 133–162.

Lurie, P.M., Goldberg, M.S., 1998. An Approximate Method for Sampling Correlated Random Variables from Partially-Specified Distributions. Manage. Sci. https://doi.org/10.1287/mnsc.44.2.203

Ma, Z., Koutsopoulos, H.N., Ferreira, L., Mesbah, M., 2017. Estimation of trip travel time distribution using a generalized Markov chain approach. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2016.11.008

Mahmassani, H.S., Hou, T., Dong, J., 2012. Characterizing Travel Time Variability in Vehicular Traffic Networks: Deriving a Robust Relation for Reliability Analysis. Transp. Res. Rec. 2315, 141–152. https://doi.org/10.3141/2315-15

Mahmassani, H.S., Sbayti, H., Zhou, X., 2004. Dynasmart-p: Intelligent transportation network planning tool: Version 1.0 users guide. Coll. Park. MD Maryl. Transp. Initiat. Univ. Maryl.

Mann, H.B., Whitney, D.R., 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann. Math. Stat. https://doi.org/10.1214/aoms/1177730491

Mehta, N.B., Wu, J., Molisch, A.F., Zhang, J., 2007. Approximating a sum of random variables with a lognormal. IEEE Trans. Wirel. Commun. https://doi.org/10.1109/TWC.2007.051000

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. https://doi.org/10.1063/1.1699114

Miller-Hooks, E., 2001a. Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. Networks. https://doi.org/10.1002/1097-0037(200101)37:1<35::AID-NET4>3.0.CO;2-G

Miller-Hooks, E., 2001b. Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. Networks. https://doi.org/10.1002/1097-0037(200101)37:1<35::AID-NET4>3.0.CO;2-G

Miller-Hooks, E., 1999. Optimal routing in time-varying, stochastic networks: Algorithms and implementations.

Miller-Hooks, E., Mahmassani, H., 2003a. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. Eur. J. Oper. Res. https://doi.org/10.1016/S0377-2217(02)00231-X

Miller-Hooks, E., Mahmassani, H.S., 2003b. Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks. Transp. Sci. 34, 198–215. https://doi.org/10.1287/trsc.34.2.198.12304

Miller-Hooks, E., Mahmassani, H.S., 2000a. Least expected time paths in stochastic, time-varying transportation networks. Transp. Sci. 34, 198–215. https://doi.org/10.1287/trsc.34.2.198.12304

Miller-Hooks, E., Mahmassani, H.S., 2000b. Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks. Transp. Sci. 34, 198–215. https://doi.org/10.1287/trsc.34.2.198.12304

Miller-Hooks, E., Mahmassani, H.S., 1998a. Least possible time paths in stochastic, time-varying networks. Comput. Oper. Res. https://doi.org/10.1016/S0305-0548(98)00027-6

Miller-Hooks, E., Mahmassani, H.S., 1998b. On the generation of nondominated paths in stochastic, time-varying networks, in: Proceedings of TRISTAN III (Triennial Symposium on Transportation Analysis), San Juan.

Mirchandani, P.B., 1976. Shortest distance and reliability of probabilistic networks. Comput. Oper. Res. https://doi.org/10.1016/0305-0548(76)90017-4

Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.P., 2010. Community structure in time-dependent, multiscale, and multiplex networks. Science (80-.). https://doi.org/10.1126/science.1184819

Newman, M., 2010. Networks: An Introduction, Networks: An Introduction. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Newman, M.E.J., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.0601602103

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. https://doi.org/10.1103/PhysRevE.69.026113

Niaki, S., Abbasi, B., 2008. Generating Correlation Matrices for Normal Random Vectors in

NORTA Algorithm Using Artificial Neural Networks. J. Uncertain Syst.

Nie, Y., Wu, X., 2009a. Shortest path problem considering on-time arrival probability. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2009.01.008

Nie, Y., Wu, X., 2009b. Reliable a Priori Shortest Path Problem with Limited Spatial and Temporal Dependencies, in: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Transportation and Traffic Theory 2009: Golden Jubilee: Papers Selected for Presentation at ISTTT18, a Peer Reviewed Series since 1959. Springer US, Boston, MA, pp. 169–195. https://doi.org/10.1007/978-1-4419-0820-9_9

Nielsen, L.R., Andersen, K.A., Pretolani, D., 2014. Ranking paths in stochastic time-dependent networks. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2013.10.022

Opasanon, S., Miller-Hooks, E., 2006. Multicriteria adaptive paths in stochastic, time-varying networks. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2004.12.003

Perng, C.S., Wang, H., Zhang, S.R., Parker, D.S., 2000. Landmarks: A new model for similarity-based pattern querying in time series databases. Proc. - Int. Conf. Data Eng. https://doi.org/10.1109/icde.2000.839385

Polychronopoulos, G.H., Tsitsiklis, J.N., 1996. Stochastic shortest path problems with recourse. Networks. https://doi.org/10.1002/(SICI)1097-0037(199603)27:2<133::AID-NET5>3.0.CO;2-L

Prakash, A., Srinivasan, K.K., 2014. Sample-based algorithm to determine minimum robust cost path with correlated link travel times. Transp. Res. Rec. https://doi.org/10.3141/2467-12

Prakash, A.A., Seshadri, R., Srinivasan, K.K., 2018. A consistent reliability-based user-equilibrium problem with risk-averse users and endogenous travel time correlations: Formulation and solution algorithm. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2018.06.003

Prakash, A.A., Srinivasan, K.K., 2018. Pruning algorithms to determine reliable paths on networks with random and correlated link travel times. Transp. Sci. https://doi.org/10.1287/trsc.2015.0668

Prakash, A.A., Srinivasan, K.K., 2017. Finding the Most Reliable Strategy on Stochastic and Time-Dependent Transportation Networks: A Hypergraph Based Formulation. Networks Spat. Econ. https://doi.org/10.1007/s11067-017-9345-2

Prakash, A.A., Srinivasan, K.K., 2015. Optimal reliability path problem on networks with random and correlated link travel times bilevel formulation and solution procedure. Transp. Res. Rec. https://doi.org/10.3141/2497-06

Pretolani, D., 2000. A directed hypergraph model for random time dependent shortest paths. Eur. J. Oper. Res. https://doi.org/10.1016/S0377-2217(99)00259-3

Pretolani, D., Nielsen, L.R., Andersen, K.A., Ehrgott, M., 2009. Time-adaptive and history-adaptive multicriterion routing in stochastic, time-dependent networks. Oper. Res. Lett. https://doi.org/10.1016/j.orl.2009.02.001

Provan, J.S., 2003. A Polynomial-Time Algorithm to Find Shortest Paths with Recourse. Networks. https://doi.org/10.1002/net.10063

Pu, W., 2011. Analytic Relationships between Travel Time Reliability Measures. Transp. Res. Rec. 2254, 122–130. https://doi.org/10.3141/2254-13

Rahmani, M., Jenelius, E., Koutsopoulos, H.N., 2013. Route travel time estimation using low-frequency floating car data, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. https://doi.org/10.1109/ITSC.2013.6728569

Ramezani, M., Geroliminis, N., 2012. On the estimation of arterial route travel time distribution with Markov chains. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2012.08.004

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. J. risk 2, 21–42.

Sigal, C.E., Pritsker, A.A.B., Solberg, J.J., 1980. The Stochastic Shortest Route Problem. Oper. Res. https://doi.org/10.1287/opre.28.5.1122

Srinivasan, K.K., Prakash, A.A., Seshadri, R., 2014. Finding most reliable paths on networks with correlated and shifted log-normal travel times. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2013.10.011

Truong, C., Oudre, L., Vayatis, N., 2020. Selective review of offline change point detection methods. Signal Processing. https://doi.org/10.1016/j.sigpro.2019.107299

Waller, S.T., Ziliaskopoulos, A.K., 2002. On the Online Shortest Path Problem with Limited Arc Cost Dependencies. Networks. https://doi.org/10.1002/net.10049

Westgate, B.S., Woodard, D.B., Matteson, D.S., Henderson, S.G., 2013. Travel time estimation for ambulances using Bayesian data augmentation. Ann. Appl. Stat. https://doi.org/10.1214/13-AOAS626

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biometrics Bull. https://doi.org/10.2307/3001968

Wu, X., Nie, Y., 2009. Implementation Issues for the Reliable a Priori Shortest Path Problem. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.3141/2091-06

Xing, T., Zhou, X., 2013. Reformulation and solution algorithms for absolute and percentile robust shortest path problems. IEEE Trans. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2013.2250966

Xing, T., Zhou, X., 2011. Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2011.06.004

Yang, L., Zhou, X., 2017. Optimizing on-time arrival probability and percentile travel time for elementary path finding in time-dependent transportation networks: Linear mixed integer programming reformulations. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2016.11.012

Yang, L., Zhou, X., 2014. Constraint reformulation and a Lagrangian relaxation-based solution algorithm for a least expected time path problem. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2013.10.012

Yelchuru, B., Raj Kamalanathsharma, Mahmassani, H.S., Hong, Z., Mittal, A., Xu, X., 2017. Analysis, modeling, and simulation (ams) testbed framework for dynamic mobility applications (dma) and active transportation and demand management (atdm) programs. US Department of Transportation.

Yi, B, K., Faloutsos, C., 2000. Fast Time Sequence Indexing for Arbitrary Lp norms. Proc. 26st Int. Conf. VLDB'00.

Yildirimoglu, M., Geroliminis, N., 2013. Experienced travel time prediction for congested freeways. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2013.03.006

Zeng, W., Miwa, T., Wakita, Y., Morikawa, T., 2015. Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2015.04.018

Zhang, Y., Jun, Y., Wei, G., Wu, L., 2010. Find multi-objective paths in stochastic networks via chaotic immune PSO. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2009.07.025

Zhang, Y., Max Shen, Z.J., Song, S., 2017. Lagrangian relaxation for the reliable shortest path problem with correlated link travel times. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2017.04.006

Zockaie, A., Mahmassani, H.S., Nie, Y., 2016. Path Finding in Stochastic Time Varying Networks with Spatial and Temporal Correlations for Heterogeneous Travelers. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.3141/2567-12

Zockaie, A., Nie, Y., Wu, X., Mahmassani, H.S., 2013. Impacts of Correlations on Reliable Shortest Path Finding. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.3141/2334-01

Zockaie, A., Nie, Y.M., Mahmassani, H.S., 2015. Simulation-Based Method for Finding Minimum Travel Time Budget Paths in Stochastic Networks with Correlated Link Times. Transp. Res. Rec. J. Transp. Res. Board 2467, 140–148. https://doi.org/10.3141/2467-15

Zockaie, A., Nie, Y.M., Mahmassani, H.S., 2014. Simulation-based method for finding minimum travel time budget paths in stochastic networks with correlated link times. Transp. Res. Rec. 2467, 140–148.