Development and Application of Environmentally Friendly Intelligent Transportation System (ECO-ITS) Freight Strategies

July 2022

A Research Report from the National Center for Sustainable Transportation

Kanok Boriboonsomsin, Alexander Vu, Peng Hao, Zhensong Wei, Dylan Brown, and Matthew Barth

University of California, Riverside

Yihang Zhang, Faisal Alasiri, Filipe Vital, and Petros Ioannou

University of Southern California

National Center for Sustainable Transportation

UCR College of Engineering – Center for Environmental Research and Technology

USC University of Southern California

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. NCST-UCR-RR-22-25	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Development and Application of Environmentally Friendly Intelligent Transportation System (ECO-ITS) Freight Strategies		**5. Report Date** July 2022 **6. Performing Organization Code** N/A
7. Author(s) Kanok Boriboonsomsin, Ph.D., https://orcid.org/0000-0003-2558-5343 Alexander Vu, https://orcid.org/0000-0003-4265-6350 Peng Hao, Ph.D., https://orcid.org/0000-0001-5864-7358 Zhensong Wei, https://orcid.org/0000-0003-3523-5689 Matthew Barth, Ph.D., https://orcid.org/0000-0002-4735-5859 Yihang Zhang, Ph.D. https://orcid.org/0000-0003-1188-8343 Faisal Alasiri, Ph.D. https://orcid.org/0000-0002-4381-3322 Filipe Vital, Ph.D., https://orcid.org/0000-0001-5987-5993 Petros Ioannou, Ph.D., https://orcid.org/0000-0001-6981-0704		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address College of Engineering - Center for Environmental Research and Technology University of California, Riverside 1084 Columbia Ave, Riverside, CA 92507 METRANS Transportation Center University of Southern California 650 Childs Way, Rm 216, Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** SCAQMD Contract # 17276 USDOT Grant 69A3551747114
12. Sponsoring Agency Name and Address South Coast Air Quality Management District 21865 Copley Dr, Diamond Bar, CA 91765 U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (August 2017 – January 2022) **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes
DOI: https://doi.org/10.7922/G22J6964
Dataset DOI: https://doi.org/10.6086/D1BT3K

16. Abstract
In the last few decades, efforts to reduce emissions from heavy-duty diesel trucks (HDDTs) and their health impacts have been focused on imposing increasingly stringent emissions standards. This has led to significant advancements in emission control technologies and alternative fuel vehicle technologies. While these technologies are effective at reducing emissions from HDDTs, the turnover of the existing HDDT population to these advanced technologies would require a large amount of investment and a long time. In the near term, other efforts to reduce emissions of the existing HDDTs and mitigate their impacts on communities are needed. Many studies have shown the promise of intelligent transportation systems (ITS) technologies in reducing the energy consumption and environmental footprint of people and goods movement through various means. This research is aimed at developing and evaluating eco-friendly ITS strategies for freight vehicles and traffic, with a focus on strategies that are applicable to the transportation systems in the South Coast Air Basin. Four specific strategies are examined in this research, including: 1) connected eco-driving, 2) truck eco-routing, 3) integrated traffic control, and 4) intelligent parking assist. This report describes the evaluation of each strategy, discusses results, and provide recommendations for future implementation.

17. Key Words Heavy-duty trucks, freight movement, intelligent transportation sytems, traffic, emissions	**18. Distribution Statement** No restrictions.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 358	**22. Price** N/A

NCST

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; University of California, Riverside; University of Southern California; California State University, Long Beach; Georgia Institute of Technology; and University of Vermont. More information can be found at: ncst.ucdavis.edu.

U.S. Department of Transportation Disclaimer

South Coast Air Quality Management District Disclaimer

Acknowledgments

This report was submitted in fulfillment of SCAQMD contract #17276 "Development and Application of Environmentally Friendly Intelligent Transportation System (ECO-ITS) Freight Strategies" by the University of California at Riverside (UCR) and the University of Southern California (USC) under the sponsorship of the South Coast Air Quality Management District. Work was completed as of December 31, 2021.

This study was funded, partially or entirely, by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation (USDOT) and the South Coast Air Quality Management District (SCAQMD) through the University Transportation Centers program. The authors would like to thank the NCST, the USDOT, and the SCAQMD for their support of university-based research in transportation, and especially for the funding provided in support of this project.

The UCR research team would like to acknowledge the California Energy Commission, the California Air Resources Board, and the Los Angeles County Metropolitan Transportation Authority for co-funding the development of Eco-Drive and associated infrastructure, as well as Los Angeles County Department of Public Works, City of Carson, City of Los Angeles Department

of Transportation, Volvo Technology of America, Econolite, McCain, and Western Systems for their technical support. The UCR research team is also thankful to Mr. Troy Musgrave of Dependable Highway Express for explaining the operations of the fleet and sharing sample fleet operations data.

Development and Application of Environmentally Friendly Intelligent Transportation System (ECO-ITS) Freight Strategies

A National Center for Sustainable Transportation Research Report

July 2022

Kanok Boriboonsomsin, Alexander Vu, Peng Hao, Zhensong Wei, Dylan Brown, Matthew Barth

College of Engineering – Center for Environmental Research and Technology, University of California, Riverside

Yihang Zhang, Faisal Alasiri, Filipe Vital, Petros Ioannou

Department of Electrical and Computer Engineering, University of Southern California

NCST

[page intentionally left blank]

TABLE OF CONTENTS

List of Tables

List of Figures

Development and Application of Environmentally Friendly Intelligent Transportation System (ECO-ITS) Freight Strategies

EXECUTIVE SUMMARY

Over the last several years, many studies have shown the promise of intelligent transportation systems (ITS) technologies in reducing the energy consumption and environmental footprint of people and goods movement through various means. This research is aimed at developing and evaluating eco-friendly ITS strategies for freight vehicles and traffic, with a focus on strategies that are applicable to the transportation systems in the South Coast Air Basin. Four specific strategies are examined in this research, including: 1) connected eco-driving, 2) truck eco-routing, 3) integrated traffic control, and 4) intelligent parking assist. The description of the evaluation of each strategy and key findings are summarized below.

Connected Truck Eco-Driving

Eco-approach and departure (EAD) at signalized intersections has been shown to be one of the most effective connected vehicle applications for energy savings and emission reductions. It uses signal phase and timing (SPaT) information from the upcoming traffic signal along with the information about the state of the host vehicle and preceding traffic to determine the best course of action for the vehicle to pass through the intersection. Research and development of EAD applications had been focused on passenger cars, and it was only recently that EAD was applied to heavy-duty trucks. Nevertheless, most of the research on EAD for heavy-duty trucks to date has primarily been conducted in numerical or traffic simulation environment. In this project, the research team designed and evaluated user interfaces of an EAD system in a heavy-duty truck driving simulator. Then, we implemented the promising user interface design in an actual truck EAD system and conducted an evaluation of its effectiveness in real world.

The EAD user interface designs that were evaluated include the audio-only interface and the audio & visual interface. Results from the driving experiment show that in general, the audio & visual interface results in a similar or higher level of reduction in travel time, fuel consumption, as well as CO2, NOx, and PM2.5 emissions than the audio-only interface. Surveys were also conducted of the truck drivers who participated in the driving experiment. According to the survey results, a majority of the drivers thinks that both the SPaT information and the advisory speed are very useful or extremely useful. It is also clear that most drivers would prefer the audio & visual interface and that they would always use the truck EAD system with this type of user interface if it is available on their trucks. Based on these findings and the superior performance of the audio & visual interface in the driving experiment, the research team implemented this user interface on a truck EAD system, called “Eco-Drive”, for a real-world evaluation on two connected corridors in Carson, California.

The results from the real-world performance evaluation show that driving with Eco-Drive resulted in less fuel consumption than driving without it by 6% to 15%, but the underlying reasons for which the fuel savings were achieved varied by corridor. On the Alameda St corridor, Eco-Drive helped the driver better comply with the speed limit of the road. This helped reduce speed fluctuations around the cruising speed, leading to improvements in fuel efficiency and potentially traffic safety. On the Wilmington Ave corridor, Eco-Drive helped cut down number of stops at connected intersections considerably, resulting in not only fuel savings but also travel time savings.

It should be noted that during the data collection the truck was not pulling any load (i.e., bobtailing). It is expected that Eco-Drive would provide a higher level of fuel savings than observed in this evaluation if the truck pulls a load, especially a heavy one. It is recommended that the effectiveness of Eco-Drive be also evaluated under those conditions in future work. In addition, it is recommended that the potential traffic safety benefits of Eco-Drive be studied in more depth as part of future research. Although the data collected in this project provide anecdotal evidence that the truck driver better complied with the speed limit of the road when driving with Eco-Drive, more data from additional drivers should be collected to verify whether Eco-Drive also improves traffic safety in addition to climate and air quality.

Truck Eco-Routing

Over the last decade, there has been much research and development on new routing techniques for navigation systems. These include eco-routing that is aimed at finding the route that would minimize vehicle energy consumption and/or emissions for the trip. To date, most of the research and development of eco-routing has been focused on passenger cars. Although there have been some research studies that apply eco-routing to heavy-duty trucks, they are based on hypothetical trips that may not be representative of the typical operations of truck fleets. Thus, it is not possible to assess how often and how much truck drivers or fleet operators could benefit from eco-routing in their typical operations. Therefore, in this project the research team applied eco-routing to the real-world truck operation data of a regional distribution fleet in Inland Southern California to determine the eco route for each trip, and then compared trip metrics of the eco route with those of the fastest route to evaluate the fuel savings potential from eco-routing.

Based on the results of 456 trips made by 48 trucks in a typical day, it was found that for 52% of the trips the fastest route is already the most fuel-efficient route. These trips tend to be very short, less than 3 miles on average, and may not have more than one realistic route option to begin with. For another 23% of the trips, the eco route would take up to one minute (1% to 8%) longer travel time than the fastest route on average, but would result in 5% to 7% fuel savings. This would be considered attractive from the fuel savings versus travel time increase tradeoff perspective. For another 11% of the trips, the eco route would take up to 3.5 minutes (12% to 17%) longer travel time on average, but would result in 7% to 8% fuel savings, which could still be considered to be acceptable. In other words, there is potential for truck eco-routing to help the trucks in this analysis save fuel (and reduce CO2 emission) on about a third of their trips.

The choice of travel route is usually made by truck drivers or fleet operators, and the primary metric for making that route choice decision is travel time. In this research, it is demonstrated that the eco route for any given trip can very well be different from the fastest route. By providing additional route options and the corresponding travel costs to the drivers, they can consider the tradeoff among the routes and make a choice with regards to which route to take based on their preferences and circumstances. For example, the eco route may be chosen if it would still allow the truck to arrive at the destination within a specified time window, while the fastest route may be preferred if the truck needs to meet a stringent delivery schedule. Even if the drivers take the eco route in some of their trips, it could still contribute to significant fuel savings and CO2 emission reductions, which are beneficial to the drivers and fleet operators as well as to the society. Thus, it is recommended that eco-routing be promoted or incentivized for use among truck drivers and fleet operators.

Integrated Traffic Control Strategies

This chapter aims to develop traffic flow control strategies that elevate traffic congestion near highway bottleneck locations as well as improve traffic mobility and reduce vehicle emissions. Highway traffic congestion is detrimental to traffic mobility, safety, and the environment. An effective and cost-efficient solution to mitigate traffic congestion, especially at highway bottlenecks, is through Intelligent Transportation Systems (ITS), such as variable speed limit (VSL), lane change (LC) control, ramp metering (RM), and many others. An integrated control design of VSL, LC, and RM for highway traffic is expected to smooth highway traffic flow, resulting in improving traffic mobility, increasing safety, and enhancing environmental sustainability. In this project, we developed a coordinated variable speed limit (VSL), ramp metering (RM), and lane change (LC) control, which stabilizes the bottleneck flow at the maximum possible level and while enhancing safety and reducing vehicle emissions.

First, we demonstrated that one of the primary reasons for the disordered traffic flow behavior near highway bottlenecks is the forced lane changes at the vicinity of the bottleneck. A lane change controller is proposed, which provides lane change recommendations to upstream vehicles to avoid creating a queue. Two types of variable speed limit controllers are designed to improve the flow rate at highway bottleneck together with the lane change controller. The combined lane change and feedback-linearization variable speed limit controller, built based on the first-order macroscopic Cell Transmission Model (CTM), can analytically guarantee the global exponential convergence to the desired equilibrium point at which the maximum possible flow rate is achieved. Then the combined LC and VSL controller is extended to coordinate with ramp metering controllers. The coordinated VSL, RM, and LC control scheme is able to improve system performance, maintain the queue length on-ramps, and keep the fairness between mainline and on-ramp flows. Microscopic simulations show consistent improvement under different traffic demands and scenarios.

Second, the proposed controller is compared to the widely used Model Predictive Control (MPC) strategy. Both macroscopic and microscopic simulations show that the performance and robustness with respect to model parameter errors and measurement noise of our proposed controller (feedback-linearization VSL) are better than that of the MPC controller. Furthermore,

we modify the CTM to include the effect of capacity drop and the decreasing discharging flow of the road section and rigorously investigate its stability properties under all possible traffic flow scenarios. The analysis is used to motivate the design of variable speed limit control to overcome capacity drop without lane change control and achieve the maximum possible flow under all feasible traffic situations.

Finally, we also consider the case where the system disturbance is included and extend the VSL control design by adding the integral action in order to reject the disturbance while avoiding the capacity drop. Microscopic simulations based on the commercial software VISSIM are used to demonstrate the effectiveness of the proposed robust VSL controller and associated benefits for traffic scenarios along a large segment of I-710 in Southern California.

Intelligent Parking Assist System

This chapter developed methods to integrate parking availability information into the planning process for long-haul trucking and studied truck parking shortages' potential impacts on the industry. First, we studied the truck driver scheduling problem (TDSP), which considers a fixed route and aims to determine a minimum duration regulation-compliant schedule. We proposed a mixed integer programming formulation that uses conditioned time-window constraints to model the parking availability at parking facilities and moving window constraints to model the hours-of-service (HOS) constraints for long trips. Simulation results illustrate that schedules calculated without accounting for parking availability are often infeasible. Although parking constraints increased trip duration in some scenarios, these scenarios also showed lower feasibility rates when ignoring parking information.

We followed by extending the TDSP under parking availability constraints to include path planning. We proposed a resource-constrained shortest path formulation that uses a resource vector to keep track of the HOS and time constraints. The problem is solver over an auxiliary network that explicitly models the different activities available to drivers and how they affect each regulation constraint. We proposed a tailored label-correcting algorithm that solves the problem to optimality. Computational experiments showed that parking conditions could significantly affect the route choice, illustrating the importance of accounting for parking availability information early in the planning process. We also simulated the potential costs of disregarding parking information under different parking shortage severity levels and how they compare to the cost increase caused by imposing parking restrictions. The results vary greatly depending on the available routes' quality, the parking shortage severity, and the expected cost of illegal parking. The results underline the importance of including parking information as early as possible to increase the quantity and quality of available routes and schedules. In addition, it also elicits the importance of further research on estimating the potential costs and risks of illegal truck parking.

Finally, we extended the resource-constrained shortest path formulation to the case of battery-electric trucks (BETs). We studied the impact of coordinating rest and recharge needs on BETs' performance and its comparison to diesel trucks. Computational experiments were used to estimate the effects of different levels of charging and parking infrastructure. Although BETs

generate only a small fraction of diesel trucks' CO2 emissions, BETs require longer trip durations in most scenarios. However, this gap in trip duration depends on battery capacity, charging infrastructure (power of regular and fast chargers), and parking/charging facilities' availability (regular chargers' time-windows, and fast chargers number and wait time). A common concern regarding the utilization of BETs for long-haul trucking is the infrastructure required to quickly charge large batteries, reducing the disparity to diesel trucks' refueling time. Nevertheless, our experiments show that, although fast-chargers can significantly reduce trip duration in many scenarios, trip duration is even more sensitive to the power and availability of the regular chargers used for long (overnight) rests. It is important to note that these results do not mean that fast chargers are without benefit. The advantages of particular infrastructure decisions will vary for each case. Our experiments only illustrate that, while trying to have BETs operating in similar itineraries to current diesel trucks (e.g., using fast chargers to reduce recharging time) might be the instinctive way to approach truck electrification, it is not the only one. It is likely not the best approach either.

In this project, we exposed the importance of using truck parking availability information during planning, and proposed methods to do so. Besides helping individual truck drivers with trip planning, the methods developed in this project can simulate different scenarios and aid policymakers in estimating the impacts of infrastructure investment decisions.

Introduction

Heavy-duty diesel trucks (HDDTs), the majority of which are used for goods movement, are significant contributors of nitrogen oxides (NOx) and particulate matter (PM) emissions in many regions of the U.S. such as the South Coast Air Basin in California (South Coast Air Quality Management District, 2017). As a result, communities close to freight hubs such as ports, railyards, and distribution centers often experience elevated levels of these pollutant emissions (Brunekreef et al., 1997; Kozawa et al., 2009). Studies have shown strong evidence of the health damage associated with diesel-related air pollution (Reis et al., 2018; Robinson et al., 2018; Wilson et al., 2018). In recent years, more regulatory attention and investment have been directed toward disadvantaged communities (DACs), such as those heavily impacted by air pollution associated with goods movement, in order to improve the health, quality of life, and economic opportunities of their residents (De León, 2012; Gomez, 2016; Garcia, 2017).

In addition, HDDTs represent the second largest share of the nation's transportation energy use, which contributes significantly to greenhouse gas emissions from transportation (U.S. Energy Information Administration, 2020). From 2007 to 2017, the share of energy used by medium and heavy vehicles grew from 19.5% of the transportation sector total to 24.4% (Davis et al., 2010; Davis and Boundy, 2020). In general, truck traffic has been growing nationwide, especially in urban areas, as a result of increased freight demand spurred by international trade and e-commerce. Vehicle miles traveled of freight trucks are expected to increase from 300 billion miles in 2019 to 415 billion miles in 2050 (U.S. Energy Information Administration, 2020). This will put a lot of pressure on the roadway infrastructure, potentially leading to increased traffic congestion, energy consumption, and emissions. Thus, innovative solutions are needed to address the growing freight demand that outpace the rate of expansion in supporting infrastructure.

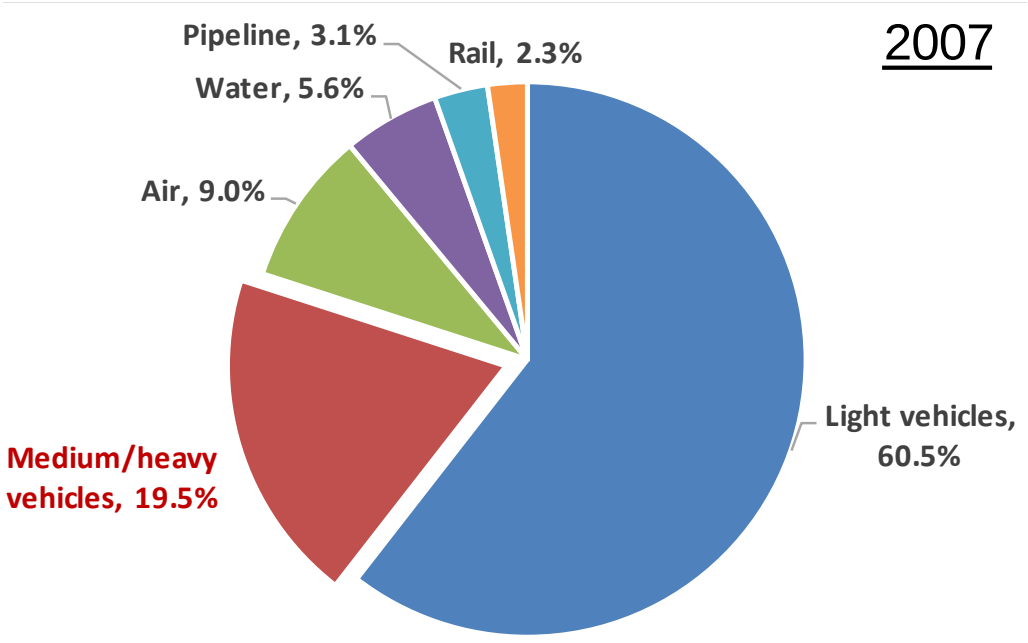


Figure 1. Energy use by mode of transportation in the U.S. in 2007 (Davis et al., 2010)

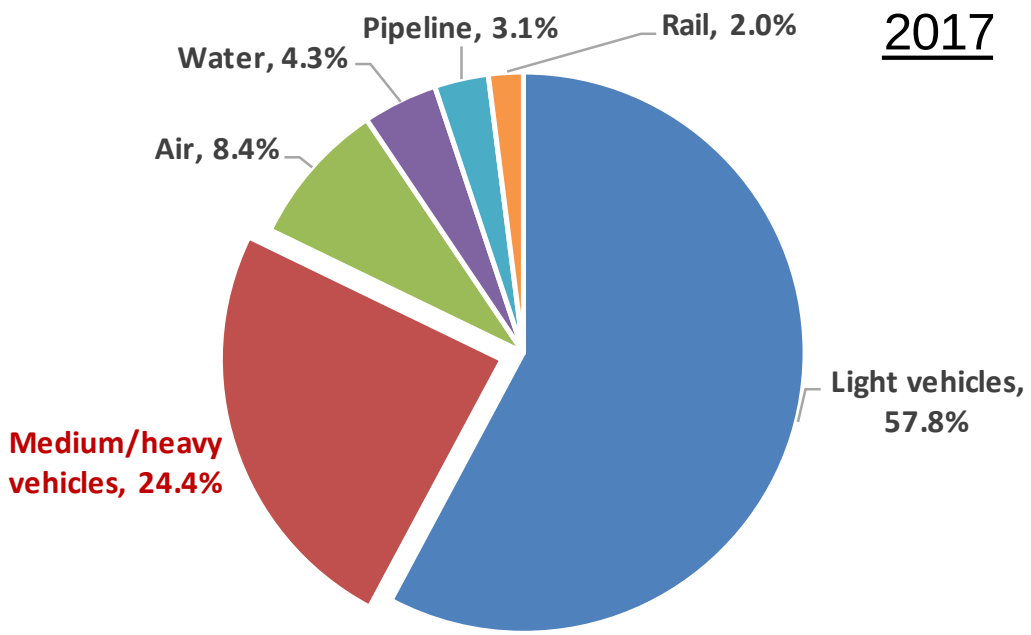


Figure 2. Energy use by mode of transportation in the U.S. in 2017 (Davis and Boundy, 2020)

In the last few decades, efforts to reduce HDDT emissions and their health impacts have been focused on imposing increasingly stringent emissions standards (California Air Resources Board, 2008). This has led to significant advancements in emission control technologies (Jiang et al., 2018) and alternative fuel vehicle technologies (McCaffery et al., 2021). While these technologies are effective at reducing emissions from HDDTs, the turnover of the existing HDDT population to these advanced technologies would require a large amount of investment and a long time. In the near term, other efforts to reduce emissions of the existing HDDTs and mitigate their impacts on communities are needed.

Over the last several years, many studies have shown the promise of intelligent transportation systems (ITS) technologies in reducing the energy consumption and environmental footprint of people and goods movement through various means, such as more energy-efficient driving (Barth and Boriboonsomsin, 2009; Boriboonsomsin, 2015; Huang et al., 2018) and choosing environmentally friendly travel route (Boriboonsomsin et al., 2012; Scora et al., 2015; Wang et al., 2019a). This research is aimed at developing and evaluating eco-friendly ITS strategies for freight vehicles and traffic, with a focus on strategies that are applicable to the transportation systems in the South Coast Air Basin. Four specific strategies are examined in this research, including: 1) connected truck eco-driving, 2) truck eco-routing, 3) integrated traffic control, and 4) intelligent parking assist. The details of the development and evaluation of each strategy are given in the individual chapters that follow.

This research is a joint effort between researchers from the University of California at Riverside (UCR), who are responsible for the connected eco-driving and truck eco-routing strategies, and researchers from the University of Southern California (USC), who are responsible for the integrated traffic control and intelligent parking assist strategies.

Connected Truck Eco-Driving

Introduction

Background

Located in the South Coast Air Basin, the Ports of Los Angeles and Long Beach are the busiest port complex in the U.S. and the ninth busiest port complex in the world by container volume. Together, they handle over 40% of all U.S. international containerized cargo. Although the two ports provide significant contribution toward economic growth and vitality for communities in the South Coast Air Basin, port-related goods movement operations and traffic are also responsible for a large portion of criteria pollutants and greenhouse gases in the region, with disproportionate impacts on communities adjacent to the ports and along major goods movement corridors. While progress has been made in reducing port truck-related emissions through fleet turnover (Bishop et al., 2012), more emission reductions from these trucks are needed in order to attain air quality and climate goals.

It was estimated that the Ports of Los Angeles and Long Beach generated about 63,000 truck trips per day in 2018, the majority of which were made by HDDTs. Unlike long-haul trucks, these port trucks or drayage trucks spend a large amount of time traveling on arterial corridors with traffic signals. Thus, they often experience traffic congestion and delays at signalized intersections. These delays result in inefficiencies in the form of increased travel time, fuel use, and emissions.

The recent advance in Connected Vehicle (CV) technology has brought forward new opportunities to enhance mobility, energy efficiency, and sustainability of people and goods movement via information sharing and better cooperation through vehicle-to-vehicle and vehicle-to-infrastructure communications. Pilot CV applications include cooperative adaptive cruise control (Milanés et al., 2014), cooperative lane change (Luo et al., 2016), cooperative ramp merging (Xie et al., 2017), and many eco-friendly CV applications initialized in the USDOT's Applications for the Environment: Real-Time Information Synthesis (AERIS) research program (U.S. Department of Transportation, 2014), such as eco-traffic signal timing, eco-traffic signal priority, and eco-speed harmonization.

Related Work

Among all those CV applications, connected eco-driving such as eco-approach and departure (EAD) at signalized intersections has been shown to be one of the most effective CV applications for energy savings and emission reductions. The EAD application uses signal phase and timing (SPaT) information from the upcoming traffic signal along with the information about the state of the host vehicle and preceding traffic to determine the best course of action for the host vehicle to pass through the intersection. Possible scenarios as shown in Figure 3 include: 1) cruising through the green light; 2) speeding up (while staying under the speed limit) to pass through the intersection before the signal turn red; 3) slowing down in advance so that the vehicle reaches the intersection just when the signal turns green; and 4) coasting to a stop if the red light is unavoidable.

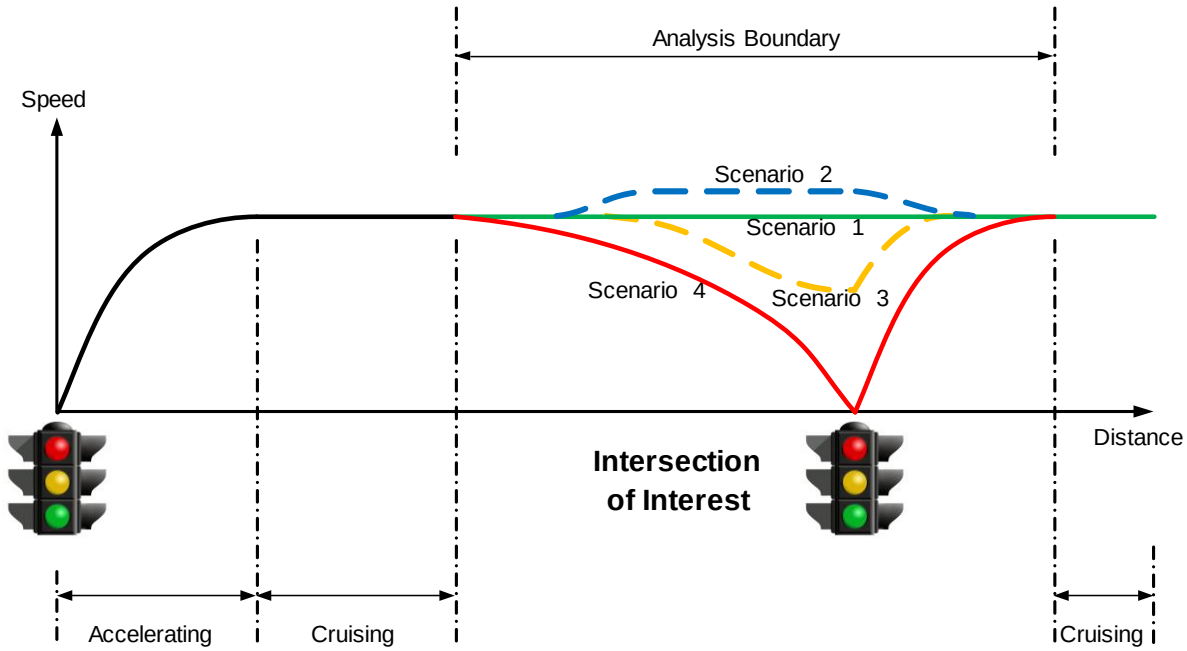


Figure 3. Scenarios when driving through an intersection with traffic signal

Once the application has determined the best course of action, it then designs a driving speed profile that would minimize fuel consumption and delay, and provides the recommended driving speed to the driver. Previous studies have shown that with a well-designed speed profile, the host vehicle would travel through the intersection in a way that reduces the number of stops as well as unnecessary acceleration and deceleration, which results in energy savings and emission reductions, while maintaining a similar travel time (Xia et al., 2013; Altan et al., 2017; Hao et al., 2019).

As most of the eco-friendly CV applications to date have been developed for passenger cars, less attention has been given to similar applications for heavy-duty trucks [Boriboonsomsin, 2015; European Commission, 2021; Lesage, 2013]. Borek et al. (2020) developed a truck optimal control model in highway environment with an offline dynamic programming optimization module along with an online model predictive control module, showing 3.7%–8.3% fuel savings without traffic and 6.5%–10% savings with traffic. Hao et al. (2021) designed an EAD algorithm for HDDTs that also accounts for road grade along the travel path, which was shown through numerical simulations to provide average fuel savings of 11% for level terrain, 6% for uphill, and 20% for downhill. On the other hand, numerical simulations of a dynamic programming-based EAD algorithm for heavy-duty trucks developed by Rodriguez and Fathy (2018) showed 32-72% fuel savings on different arterial corridor configurations. In Wang et al. (2019b), the researchers implemented an EAD system on a HDDT and demonstrated its performance in real-world traffic on urban freight corridors in Carson, California.

Research Objectives

Most of the research on EAD for heavy-duty trucks to date have primarily been conducted in numerical or traffic simulation environment. While there has been a demonstration of EAD system for heavy-duty trucks (Wang et al., 2019b), there has yet to be a systematic evaluation of the efficacy of the EAD application in real-world settings. And unlike numerical or traffic simulation studies, real-world studies of the EAD application require an actual implementation of EAD system on the vehicle. The system will include a user interface for providing information or driving recommendation to the driver. The user interface needs to be carefully designed to be user friendly and effective.

Therefore, there are two objectives for this portion of the research project: 1) to design and evaluate user interfaces for an EAD system that will be suitable for use on heavy-duty trucks, and 2) to implement the EAD system on a heavy-duty truck and conduct a systematic evaluation of its effectiveness in the real world.

User Interface Design and Evaluation

User Interface Design

The UCR research team designed a user interface for an EAD system in a previous research (Hao et al., 2019). The design was implemented on a 7-inch display for a real-world experiment of the system with a passenger car traveling on a connected corridor in the Bay Area. This user interface includes multiple features:

- Speedometer with advisory speed
- Engine speed (RPM) tachometer
- Traffic signal graphic indicating real-time SPaT
- Distance bar with real-time vehicle location in relative to the upcoming intersection
- Signal strength indicators for the dedicated short-range communication (DSRC) and global positioning system (GPS) receivers
- Radar detection indicator (i.e., indicating if a preceding vehicle was within the radar detection range)
- Distance from the preceding vehicle within the radar detection range

As shown in Figure 4, the advisory speed is only displayed when there is no preceding vehicle detected. If a preceding vehicle is detected, the advisory speed will not be given while the radar detection indicator along with the distance from it will appear instead. The logic for such design is that when there is a preceding vehicle within a close distance, the speed of the host vehicle will be primarily governed by the speed of the preceding vehicle. Also, under such situation the driver of the host vehicle should focus more on following the preceding vehicle safely.

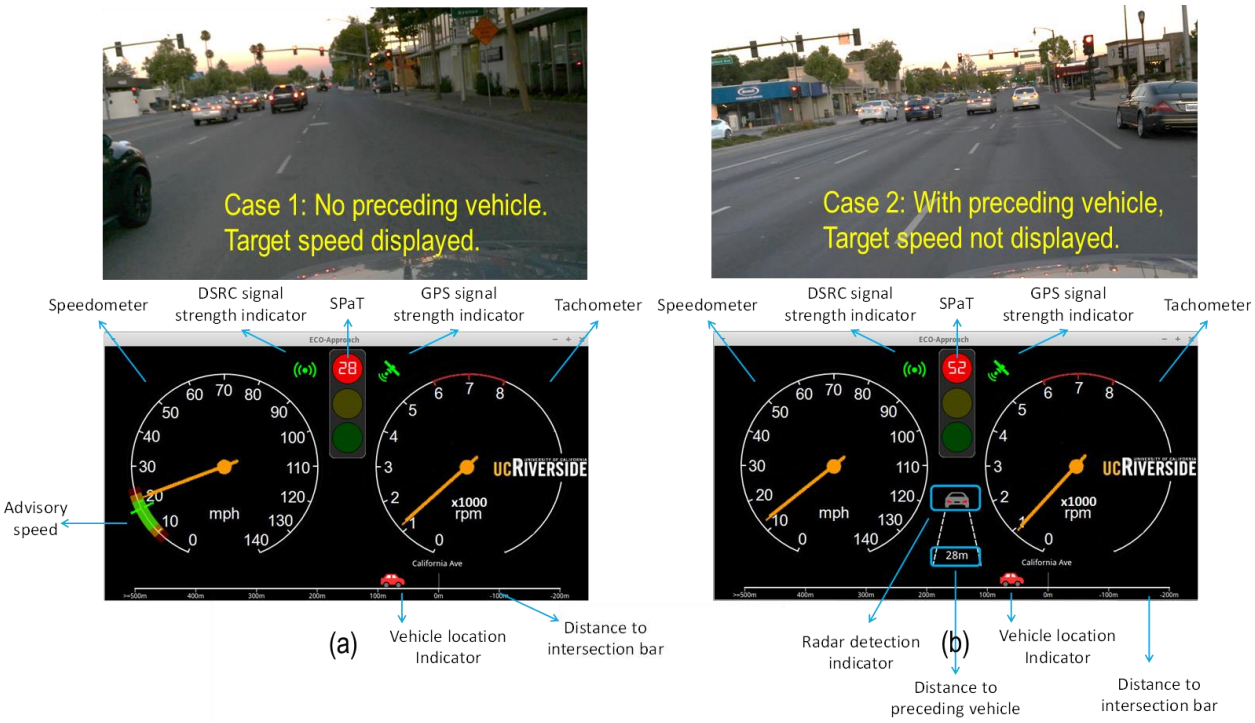


Figure 4. Visual interface of UCR's EAD system used in real-world experiment with passenger car (Hao et al., 2019)

In this project, the UCR research team redesigned the user interface with inputs from truck drivers and truck fleet owners. The goal of the new designs is to provide traffic signal information and driving recommendation to the driver in a simple and clear manner so that the driver can drive through signalized intersections smoothly, reducing vehicle energy consumption and emissions. In the redesign process, we reached out and scheduled meeting with truck drivers and truck fleet owners. At the meetings, we presented them a mock-up video of how the existing design works, and solicited their inputs. Key inputs that we received are listed below:

- [Driver] Consider making it also available on smartphones. This will be useful for trucks without an existing display on them.
- [Driver] Adding the option to orient the user interface vertically similar to commonly used driving navigation apps (e.g., Google Maps). Some drivers are more familiar with vertical orientation.
- [Driver] RPM tachometer is not important as there is one on the trucks already.
- [Driver] Distance from the preceding vehicle is not important as the driver can see the preceding vehicle.
- [Fleet owner] Consider audio interface instead of visual in order to keep the driver's eyes on the road.

Based on the inputs received, we redesigned the visual interface of the EAD system specifically for heavy-duty trucks. Most of the features were carried over from the previous design. The changes in the new design include adding the option to orient the user interface vertically (see Figure 5), removing the RPM tachometer, removing the distance from the preceding vehicle information, changing the host vehicle graphic from car to truck, and adding the cross street name of the upcoming intersection.

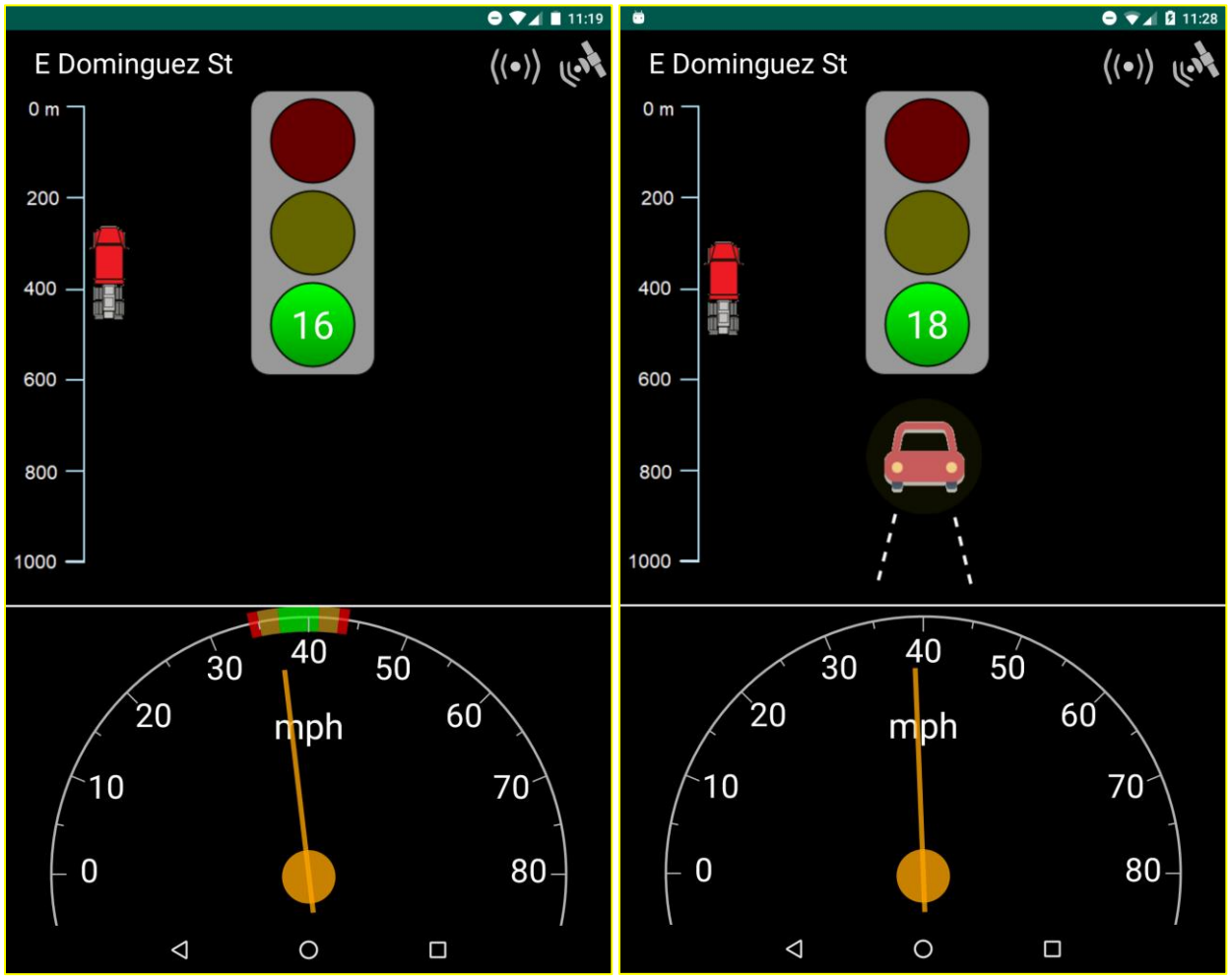


Figure 5. Visual interface of UCR's truck EAD system designed in this project

In addition, we designed a new audio interface for the truck EAD system. The flowchart of the audio interface is presented in Figure 6. This audio interface is only activated when the vehicle is within 200 meters from the intersection, which is far enough for the driver to perceive the SPaT information provided by the truck EAD system, process it, make decision, and take any necessary action. Depending on the traveling speed, the amount of time to traverse the 200 meters will range from 10 seconds at 45 mph to 22 seconds at 20 mph. When the vehicle is more than 200 meters away from the intersection, the SPaT information provided may not be highly relevant and the audio message could be deemed distractive.

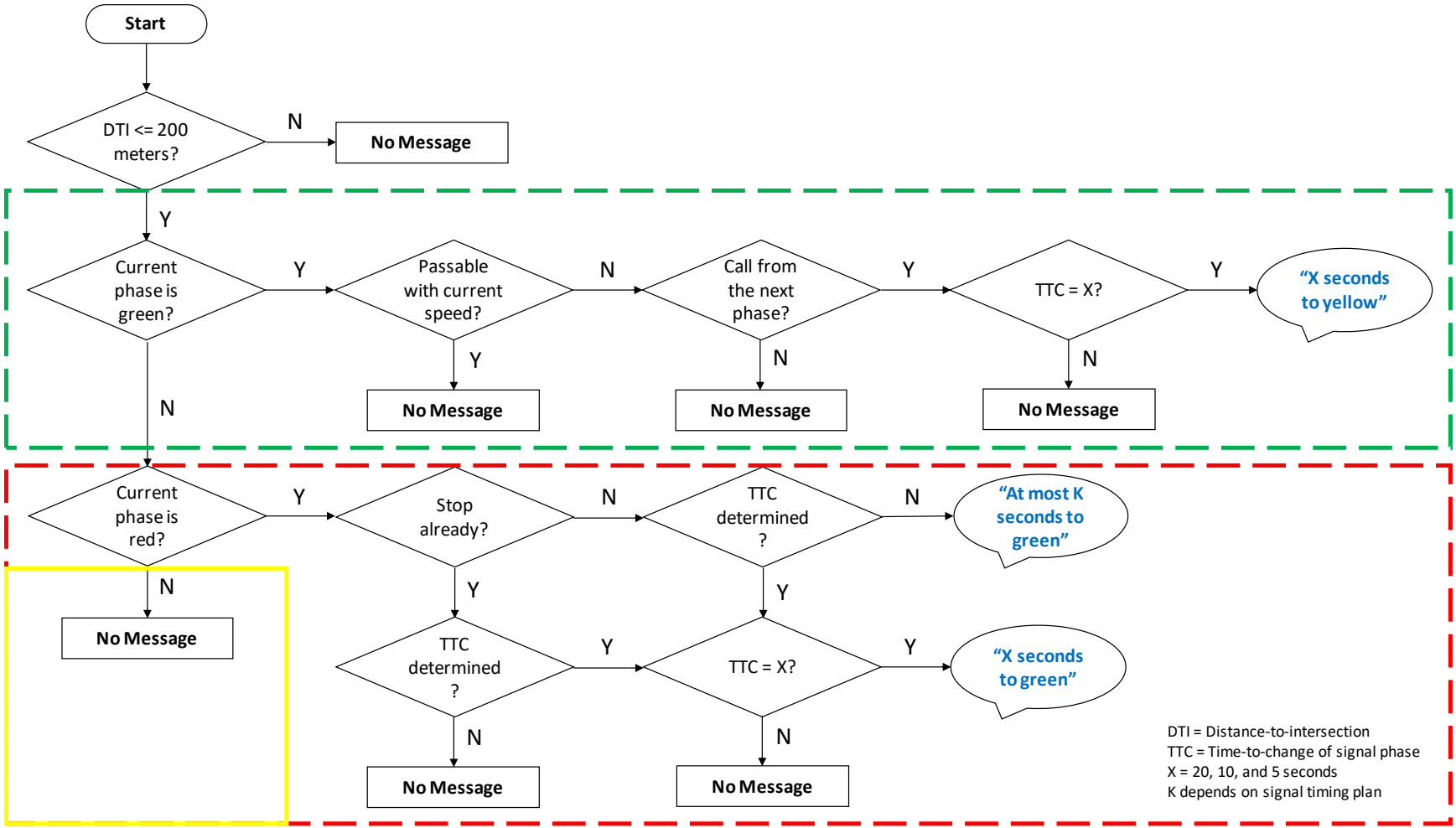


Figure 6. Flowchart of the audio interface of UCR's truck EAD system

The green, red, and yellow bounding boxes in Figure 6 represent the decisions that the truck EAD system will make when the current phase of the upcoming traffic sign is green, red, and yellow, respectively. Note that this flowchart is designed for traffic signals with actuated control, which do not always know the precise time-to-change from the current signal phase to the next phase. That is why one of the audio messages is "At most K seconds to green" where K depends on the signal timing plan. For traffic signals with fixed time control, the answer to the questions "Call from the next phase?" and "TTC determined?" will be yes.

In designing the audio interface, the research team opted for messages that provide SPaT information rather than driving recommendation. This is because a driving recommendation such as an advisory speed is determined based on multiple factors (e.g., distance-to-intersection, driving speed, SPaT, etc.) at that point in time, and it can change if the circumstance changes. Unlike visual interface, it would be difficult for the driver to follow the advisory speed provided through audio messages every few seconds. And providing changing or conflicting recommendations to the driver could undermine the credibility of the system. Thus, it was deemed that providing SPaT information to the driver and let the driver use that information to support his or her decision on how best to pass through the upcoming intersection would be more useful.

In providing the SPaT information, the research team opted to provide as few audio messages as necessary. Too many audio messages could be distractive, especially if they are irrelevant such as when the vehicle is still far away from the intersection. Therefore, instead of a constant count down of the remaining time in the current signal phase, we designed the system to only provide audio message when the vehicle is within 200 meters from the intersections and there are 20, 10, and 5 seconds left in the current phase. Under these circumstances, the information is highly relevant and could impact the driver's decision on what action to take.

With the redesigned visual interface and the newly developed audio interface, the research team set up an experiment in a truck driving simulator to evaluate the effectiveness of different user interface configurations. Given the resources available in this project, only two configurations could be evaluated. The two configurations were:

1. *Audio-only interface* – This configuration was based on the desire to limit driver's distraction. It would also make the installation of the system easier as there is no need to mount a display on the vehicle.
2. *Audio & visual interface* – This configuration was motivated by the desire to make more information available to the driver. As opposed to the visual-only interface that requires the driver to look at the display to receive information, this configuration allows the driver more flexibility on what information to receive and how to receive them. For example, drivers who do not wish to receive an advisory speed can still receive the SPaT information through audio messages without having to glance at the display.

Truck Driving Simulator

The evaluation of the two user interface configurations was conducted in a state-of-the art truck driving simulator, called Minisim™. It is a sophisticated computer-based driving simulator that consists of multiple displays, instrument panel, and high-quality steering wheel and pedals built into a system the size of a quarter cab of a truck. As shown in Figure 7, the simulator's visual graphics are displayed on three 42'' plasma screens, and there is a separate LCD screen immediately below that replicates the instrument panel on a truck. The simulator conveys sound (e.g., engine noise, horn) through a 2.1 audio system. The driver uses an adjustable steering wheel, gas and brake pedals, and gear shift lever (8 gears) when he or she drives in the simulator, the same way as driving in a real truck.

Figure 7. Minisim™ truck driving simulator at UCR

The simulator comes with several software tools including Tile Mosaic Tool (TMT) and Interactive Scenario Authoring Tool (ISAT) that are used to create driving scenarios. TMT allows roadmaps to be built using a database of premade tiles. ISAT uses the roadmap built in TMT and places objects and coordinators on it. The objects include static ones, such as signs and obstacles, as well as Autonomous Dynamic Objects (ADOs) and Dependent Dynamic Objects (DDOs), such as vehicles. ADOs can be simply placed on a roadmap, and will drive autonomously while obeying preset controls, including lane deviation, maximum and minimum speed, acceleration, and gap between vehicles. DDOs require a designated path, and will not move unless given a set of nodes to travel to on the roadmap. Coordinators, including traffic signals, global time triggers, and road pads, provide a way to control large groups of ADOs. On the other hand, DDOs are not affected by coordinators. Once a driving scenario has been created and the external driver has been given a spawn point on the roadmap, the scenario can be run by Minisim™.

In this project, the truck EAD system was integrated with the truck driving simulator as a parallel system that interfaces with Minisim™ in real time. As shown in Figure 8, a data acquisition and processing computer is connected to the truck driving simulator. It collects real-time vehicle, traffic, and SPaT information from the simulator, and uses them to determine

driving recommendation. The SPaT information and driving recommendation are then conveyed to the driver through the artificial dashboard. In this project, the artificial dashboard is a 7-inch tablet that serves as both visual and audio interfaces of the truck EAD system. That is, the audio messages from the truck EAD system are provided through the tablet's speaker as opposed to the driving simulator's audio system. This is a realistic representation of how the truck EAD system will be set up in the real world.

Figure 8. Integration of truck EAD system with truck driving simulator

Driving Scenarios

The driving scenarios in the truck driving simulator were designed to consist of a series of intersections that form a rectangular virtual test track with the speed limit of 40 mph throughout. Each of the long sides of the virtual test track consists of 17 intersections. The distance between two consecutive intersections is 500 meters. The traffic signals are set to be fixed time control with the cycle length, green time, yellow time, and red time being 60, 30, 3, and 27 seconds, respectively.

On one of the long sides of the virtual test track (i.e., 17 intersections), there is no other vehicle, while there is traffic with other vehicles around on the other long side. The former represents the scenario where the truck is the lead vehicle in the traffic stream that approaches the intersections, while the latter represents the scenario where the truck is a following vehicle. To complete one round of driving experiment, the driver will drive through a total of 34 intersections that make up various combinations of whether the truck is the lead or a following vehicle, current signal phase, remaining phase time, etc. Figure 9 shows an example scenario where the truck was the lead vehicle approaching an intersection when the traffic signal was in red phase.

Figure 9. Example driving scenario in the truck driving simulator

Experimental Procedure

Once the truck driving simulator was set up and the driving scenarios programmed, truck drivers in the local area were recruited to take part in the driving experiment. The experiment consisted of several steps as described below.

1. *Pre-Survey* – First, the participants were asked to fill out a pre-survey. This survey contains questions about their typical trucking operations and how useful the different information and driving recommendation that can be provided by the truck EAD system would be to them. A copy of the pre-survey in both English and Spanish is provided in Appendix A.
2. *Introduction to Driving Simulator* – The participants were introduced to the truck driving simulator, and given a chance to test drive it using a sample driving scenario that comes with the simulator (not the ones developed in this project). They would practice driving the simulator until they felt comfortable with it.
3. *Baseline Driving* – The participants were asked to drive the driving scenario developed in this project in a way that they would normally drive in real life.
4. *Introduction to Truck EAD System* – The participants were introduced to the truck EAD system and both of its user interface configurations. The participants were explained about all the features on the user interface. Then, they were given a chance to try the truck EAD system in another test drive using a sample driving scenario that comes with the simulator (not the ones developed in this project).

5. *Driving with Truck EAD System* – The participants were asked to drive the driving scenarios developed in this project with the use of the truck EAD system. They would drive two rounds, one with the audio-only configuration and the other with the audio & visual configuration. The order of which configuration was used first would alternate. That is, one participant would drive with the audio-only configuration first. Then, the next participant would drive with the audio & visual configuration first.
6. *Post-Survey* – Lastly, the participants were asked to fill out a post-survey. This survey contains questions about their opinion of the truck EAD system and some basic demographics. A copy of the post-survey in both English and Spanish is provided in Appendix B.

Data Collection and Processing

Driving data were automatically recorded through the Minisim™ DAQ (Data Acquisition) tool. Several parameters such as truck speed, position, engine speed, etc., were recorded at the frequency of 60 Hz. The data were aggregated into 1 Hz, and the truck speed data were then used to calculate the corresponding acceleration/deceleration rates.

Since the truck driving simulator does not have a real truck engine, truck fuel consumption was estimated using fuel consumption rate data from the U.S. Environmental Protection Agency's Motor Vehicle Emission Simulator (MOVES) model (U.S. Environmental Protection Agency, 2015) for a 2009 model year HDDT, which is the most common model year of heavy-duty trucks serving the Port of Los Angeles in 2018 (Port of Los Angeles, 2018). The estimation was based on the second-by-second speed and acceleration data of the driving as captured by the simulator. MOVES defines 23 different operating mode (OpMode) bins for a vehicle in operation based on its second-by-second speed and vehicle specific power (VSP), as shown in Figure 10. Each OpMode bin is associated with a fuel consumption or emission rate. Figure 11 shows the fuel consumption rates by OpMode bin for the truck assumed in this project. In addition to fuel consumption, several types of emissions were also estimated using emission rate data from the MOVES model. These include carbon dioxide (CO_2), carbon monoxide (CO), hydrocarbons (HC), oxides of nitrogen (NO_X), and fine particulate matter ($PM_{2.5}$). Of particular interest are NO_X and $PM_{2.5}$, which are major pollutants generated by diesel engines.

Driving Results

The research team was able to recruit 15 truck drivers to participate in the experiment. During the experiment for one driver, the equipment broke down, impacting the collected data. Thus, the data for this driver was excluded from subsequent analyses. The driving data for the other 14 drivers were applied to the MOVES model to estimate the associated fuel consumption and emissions. Table 1 shows the differences in travel time, fuel consumption, and emissions when using the audio-only interface as compared to the baseline. On average, the audio-only interface results in reductions in travel time, fuel, NOx, and PM2.5 by 10%, 2%, 4%, and 1% respectively. Table 2 shows the differences when using the audio & visual interface as compared to the baseline. On average, travel time, fuel, NOx, and PM2.5 are reduced by 10%, 4%, 5%, and 2% respectively.

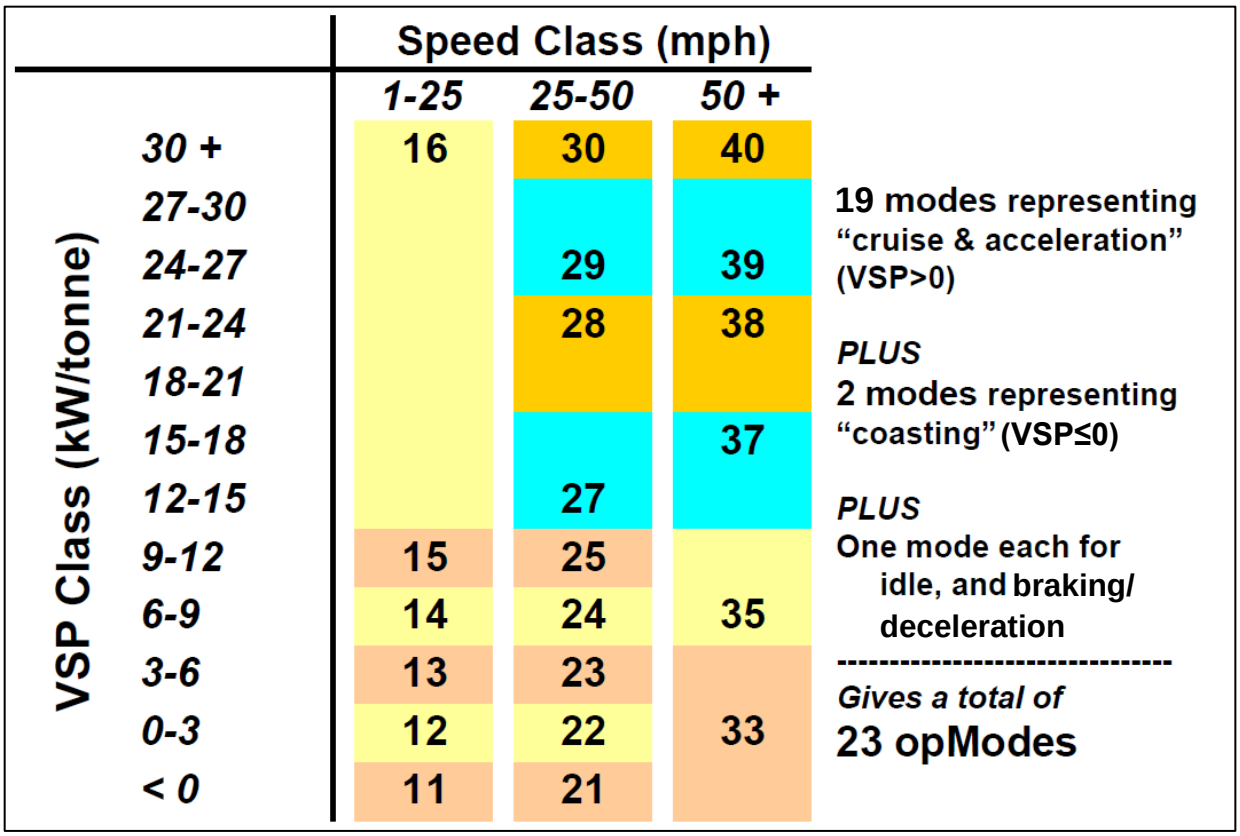


Figure 10. Definition of vehicle operating mode bins in MOVES model

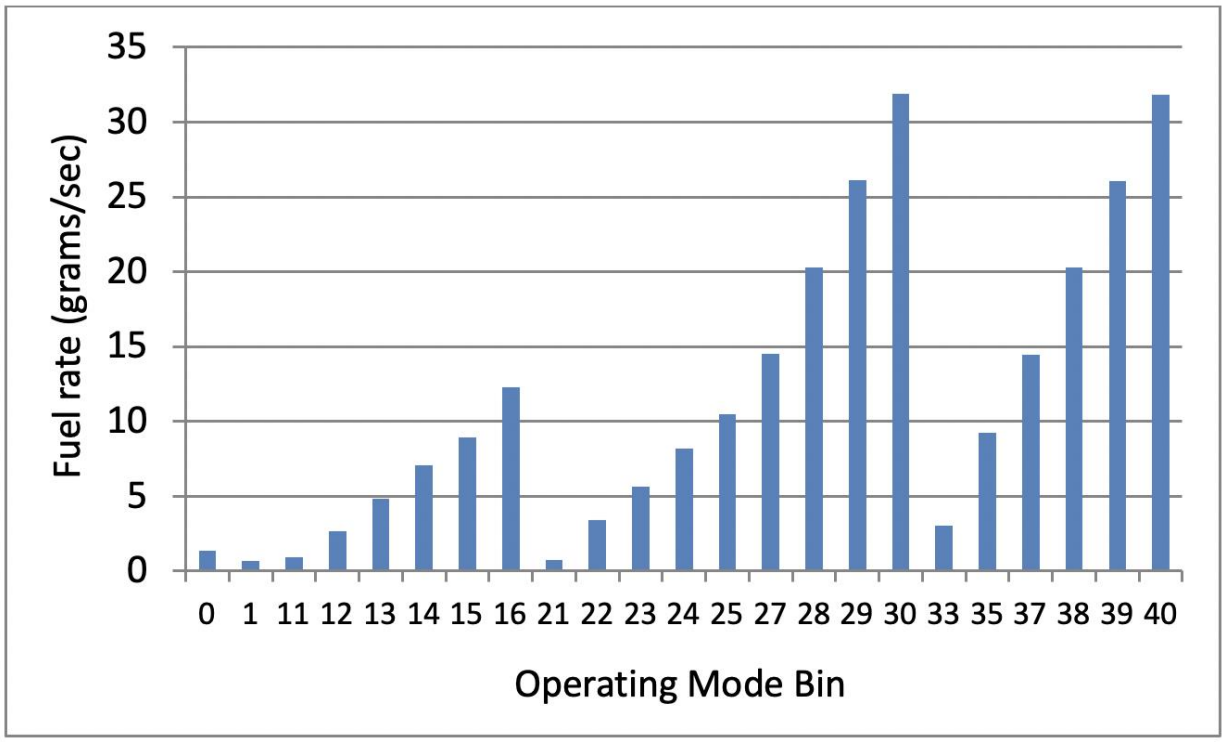


Figure 11. Fuel consumption rate of 2009 model year heavy-duty diesel truck

Table 1. Differences in driving metrics for the audio-only interface compared to baseline

Driver	Time	Fuel	CO2	NOx	PM2.5	CO	HC
001	-12%	5%	5%	2%	7%	-4%	-11%
002	-15%	-2%	-2%	-5%	0%	-4%	-14%
003	-5%	1%	1%	-1%	3%	-4%	-7%
004	-22%	-6%	-7%	-10%	-5%	-10%	-18%
005	16%	8%	7%	9%	2%	11%	17%
006	-10%	-4%	-4%	-5%	-1%	-7%	-13%
007*	-	-	-	-	-	-	-
008	-9%	4%	4%	1%	5%	0%	-8%
009	1%	0%	0%	0%	-1%	0%	4%
010	-21%	-11%	-11%	-13%	-11%	-7%	-15%
011	-24%	3%	4%	-2%	6%	-10%	-22%
012	-8%	-6%	-6%	-7%	-7%	-7%	-5%
013	-33%	-8%	-8%	-14%	-6%	-16%	-27%
014	13%	-11%	-11%	-6%	-10%	2%	11%
016	-7%	6%	5%	2%	9%	-8%	-12%
Avg.	*-10%*	*-2%*	*-2%*	*-4%*	*-1%*	*-5%*	*-9%*

*Data impacted by equipment failure during the experiment

Table 2. Differences in driving metrics for the audio & visual interface compared to baseline

Driver	Time	Fuel	CO2	NOx	PM2.5	CO	HC
001	-6%	-7%	-7%	-6%	-6%	-4%	-2%
002	-9%	-6%	-6%	-7%	-4%	-1%	-7%
003	-2%	-1%	-1%	-1%	1%	-5%	-4%
004	-19%	-3%	-3%	-6%	-1%	-6%	-19%
005	-1%	-2%	-2%	-1%	-3%	4%	0%
006	-10%	-13%	-13%	-12%	-9%	-8%	-10%
007*	-	-	-	-	-	-	-
008	-15%	4%	4%	0%	6%	-8%	-14%
009	-6%	-4%	-4%	-4%	-2%	-6%	-3%
010	-10%	-11%	-11%	-11%	-10%	-7%	-6%
011	-21%	-1%	-1%	-5%	0%	-8%	-15%
012	-22%	-12%	-12%	-15%	-12%	-14%	-18%
013	-21%	-4%	-4%	-8%	-2%	-8%	-16%
014	6%	1%	1%	2%	2%	2%	3%
016	8%	7%	7%	6%	10%	1%	-3%
Avg.	*-10%*	*-4%*	*-4%*	*-5%*	*-2%*	*-5%*	*-8%*

*Data impacted by equipment failure during the experiment

The driving speed profiles, in terms of speed vs. distance, of one of the participants are presented in Figure 12. The first half of the driving is where there was no other vehicle around. The second half is where other vehicles might present, which made the speed profiles more fluctuated. In general, the driver rarely needed to come to a full stop at any of the intersections. However, the truck EAD system was able to help the driver avoid slowing down significantly at some intersections, especially during the first half of the driving.

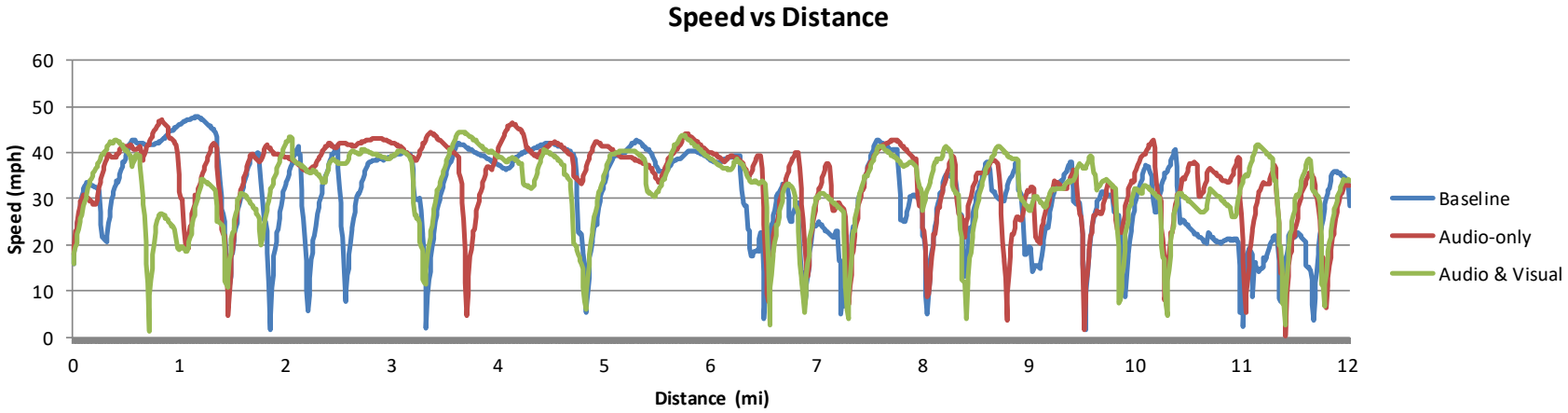


Figure 12. Speed profiles of Driver #006

We also further analyzed the data by differenciating between the first half and the second half of the driving. The results in terms of percent differences in travel time, fuel consumption, and emissions as compared to the baseline are shown in Figure 13 and Figure 14 for the audio-only configuration and the audio & visual configuration, respectively. These results represent the average values of the 14 drivers whose driving data are complete.

In general, the audio & visual configuration results in a similar or higher level of reduction in almost all the metrics. When the truck was the lead vehicle with no other preceding vehicle, the truck EAD system results in a higher level of reduction in fuel consumption as well as CO2, NOx, and PM emissions. On the other hand, the truck EAD system provides a higher level of reduction in travel time as well as CO and HC emissions when there were other vehicles present. These trends are true for both the audio-only and the audio & visual configurations.

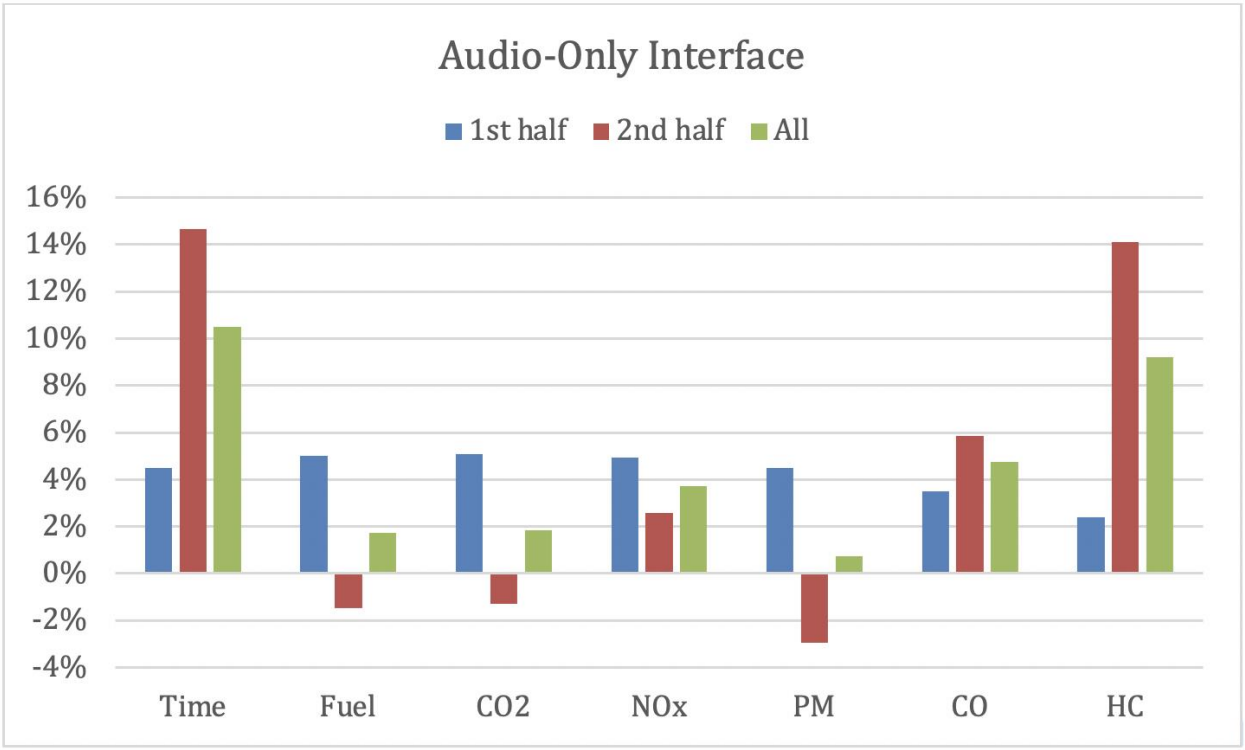


Figure 13. Differences in time, fuel, and emissions for the audio-only interface

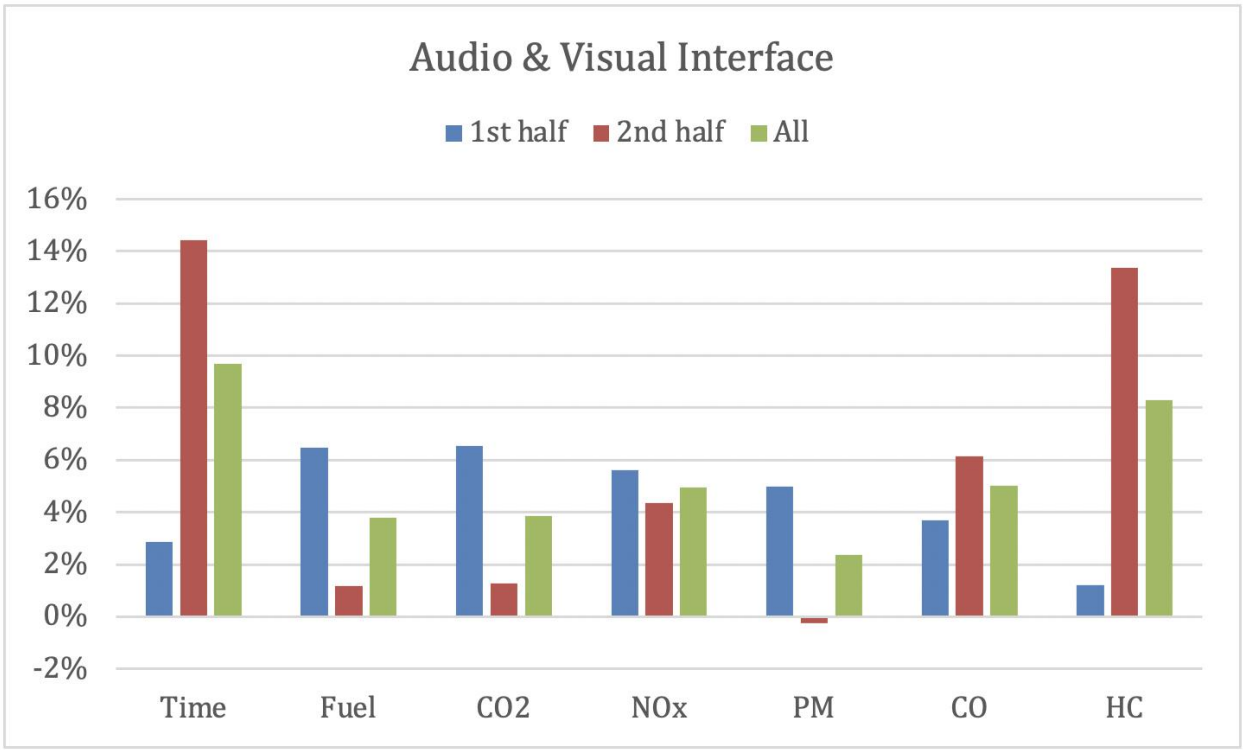


Figure 14. Differences in time, fuel, and emissions for the audio & visual interface

Survey Results

As part of the experiment, we asked the participants to fill out surveys before and after the driving. There are a total of seven questions in the "before" survey and eight questions in the "after" survey. We analyzed the survey responses, which are presented in Figure 15 through Figure 24. Highlights from these figures are summarized below:

- More than half of the participants spend over 40% of their driving jobs on surface streets with traffic signals.
- Over 90% of the participants think that it is very useful or extremely useful to know how many seconds are left before the traffic signal will change from green to yellow.
- About 75% of the participants think that it is very useful or extremely useful to know how many seconds are left before the traffic signal will change from red to green when approaching a red light. However, if already stopped at a red light, only about 45% think that such information is very or extremely useful.
- 60% of the participants thinks that a device that can recommend what speed they should be driving at in order to pass through the intersection on green is very useful or extremely useful.
- 87% of the participants thinks that a device that can recommend them to slow down ahead of time because it knows that they will not be able to pass through the intersection on green is very useful or extremely useful.
- 60% of the participants would prefer to use the audio & visual version of the truck EAD system while 33% would prefer to use the audio-only version. The other 7% would use either version.
- About 75% of the participants would always use the audio & visual version of Eco-Drive when driving on surface streets with traffic signals if it is already equipped on their trucks. This number drops to about 45% for the audio-only version.

Note that the truck driver participants were recruited from the local area. They primarily perform drayage operation or regional distribution, and thus, it is not surprising that over 40% of their driving jobs are on surface streets. Out of the 15 drivers, one is female and the rest are male. The age of these drivers ranges from 29 years old to 64 years old, with an average of 44 years old. Their professional driving experience ranges from 1.5 years to 43 years, with an average of 15 years.

According to the survey results, it is encouraging to find that a majority of the drivers thinks that both the SPaT information and the advisory speed are very useful or extremely useful. It is also clear that most drivers would prefer the audio & visual configuration and that they would always use the truck EAD system with this configuration of the user interface if it is available on their trucks. Based on these findings and the superior performance of the audio & visual configuration, we implemented this configuration of the truck EAD system for a real-world evaluation that is described in the next section.

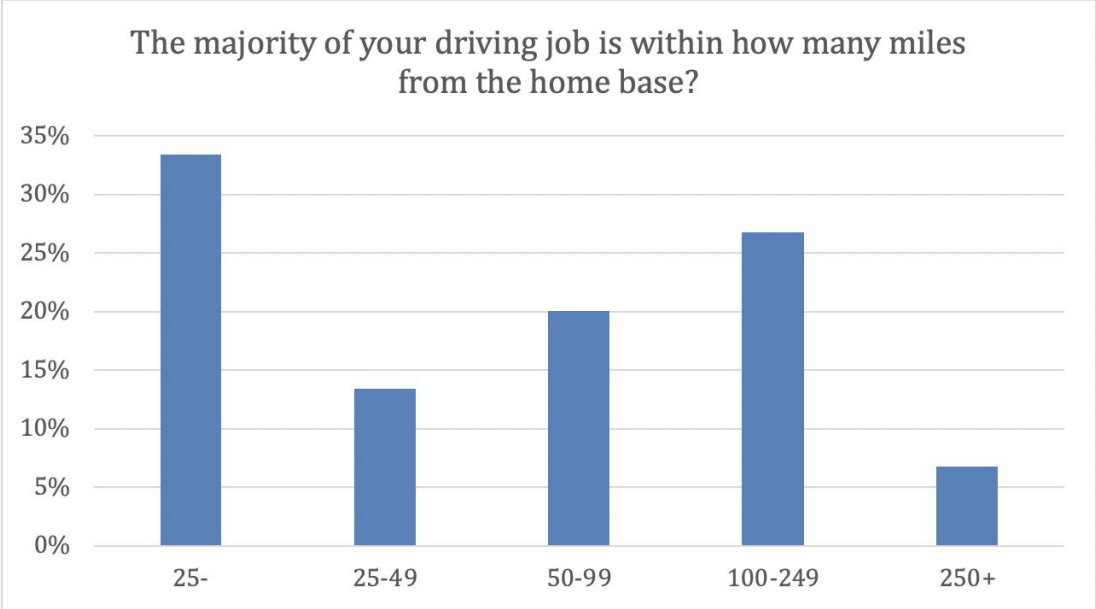


Figure 15. Distribution of responses to Question # 1 in the pre-survey

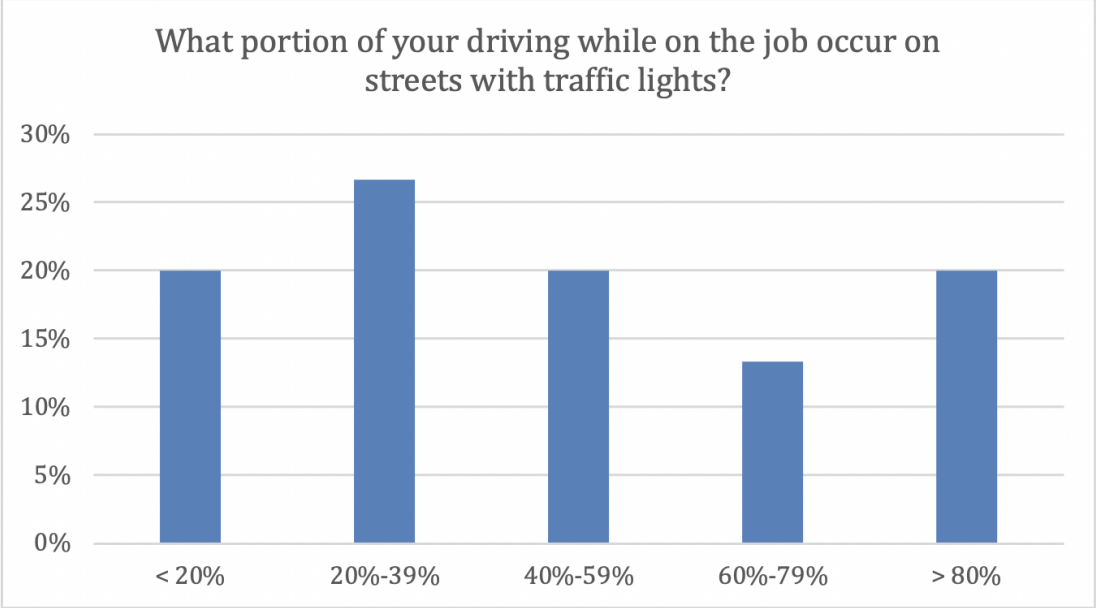


Figure 16. Distribution of responses to Question # 2 in the pre-survey

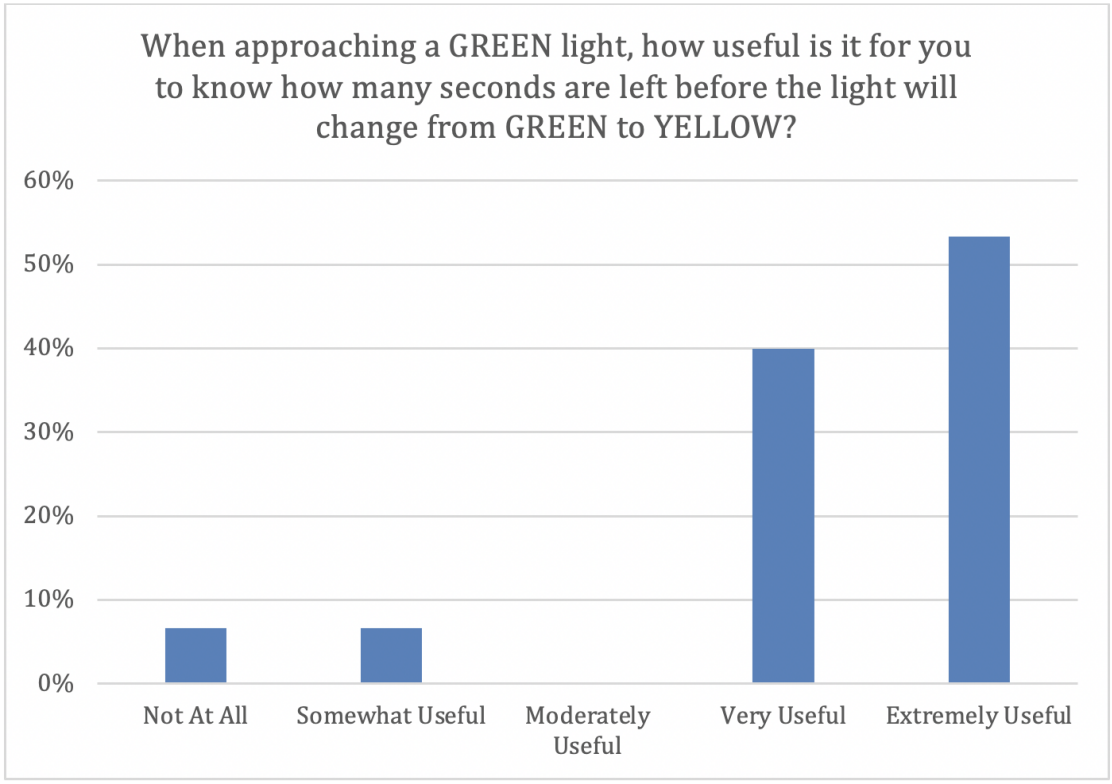


Figure 17. Distribution of responses to Question # 3 in the pre-survey

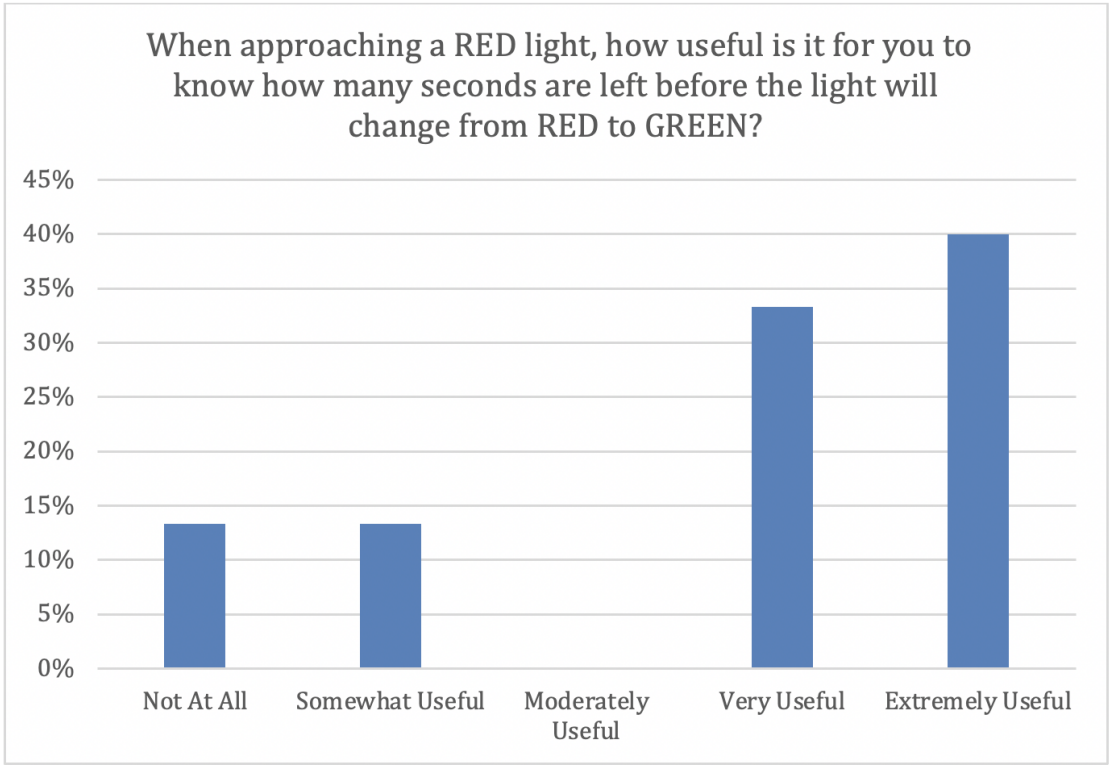


Figure 18. Distribution of responses to Question # 4 in the pre-survey

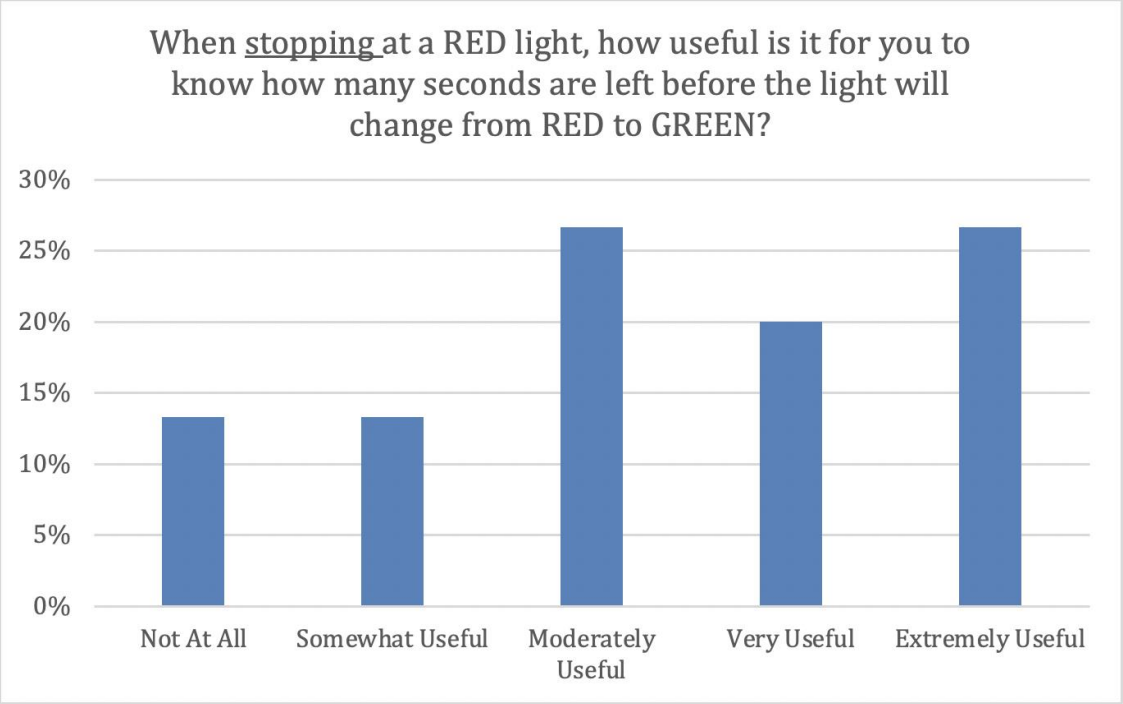


Figure 19. Distribution of responses to Question # 5 in the pre-survey

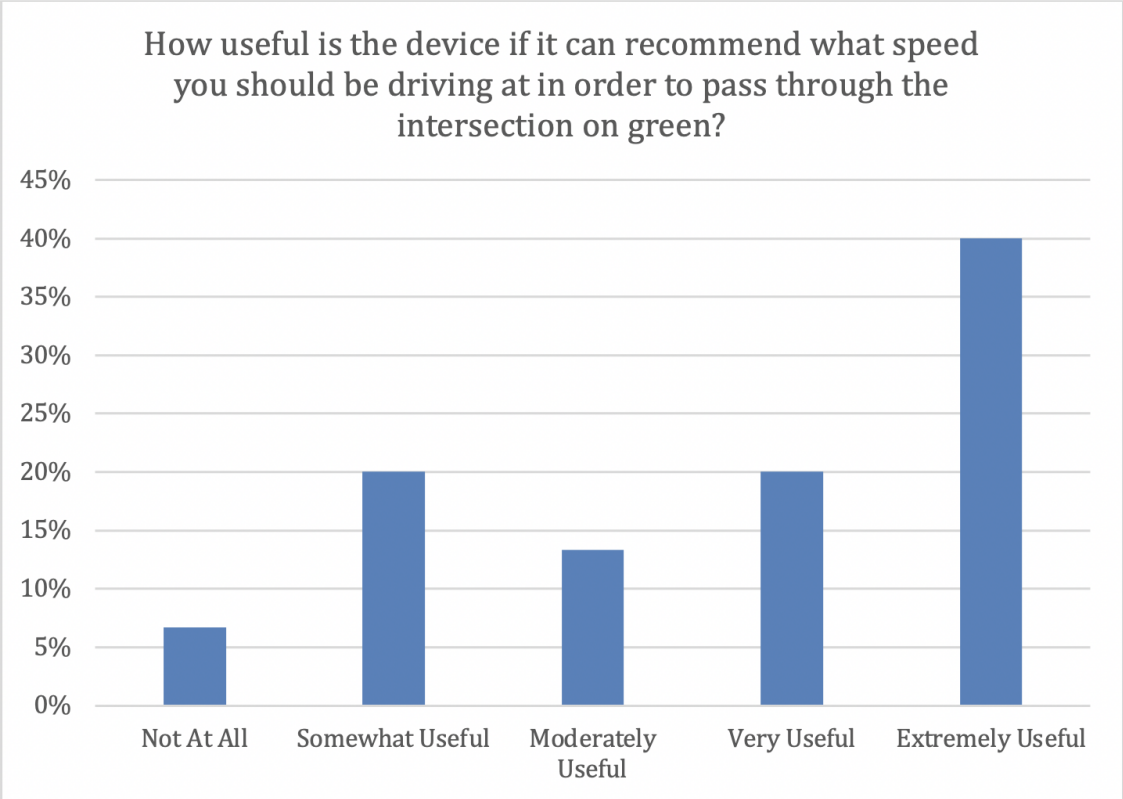


Figure 20. Distribution of responses to Question # 6 in the pre-survey

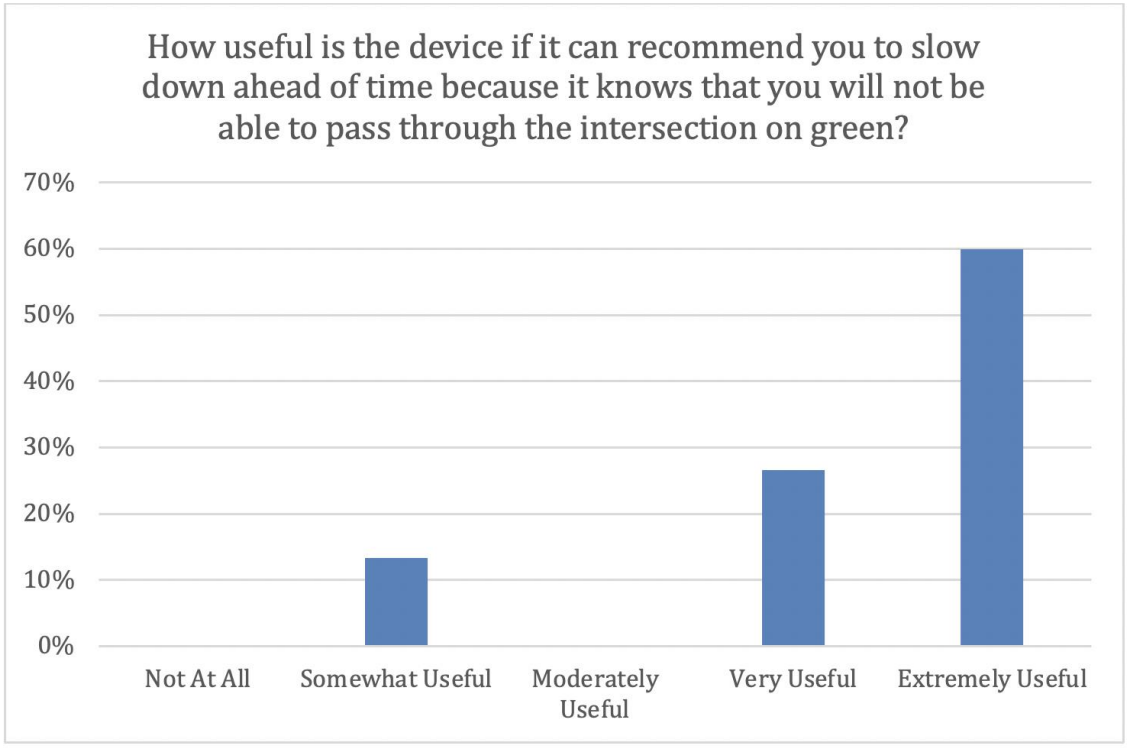


Figure 21. Distribution of responses to Question # 7 in the pre-survey

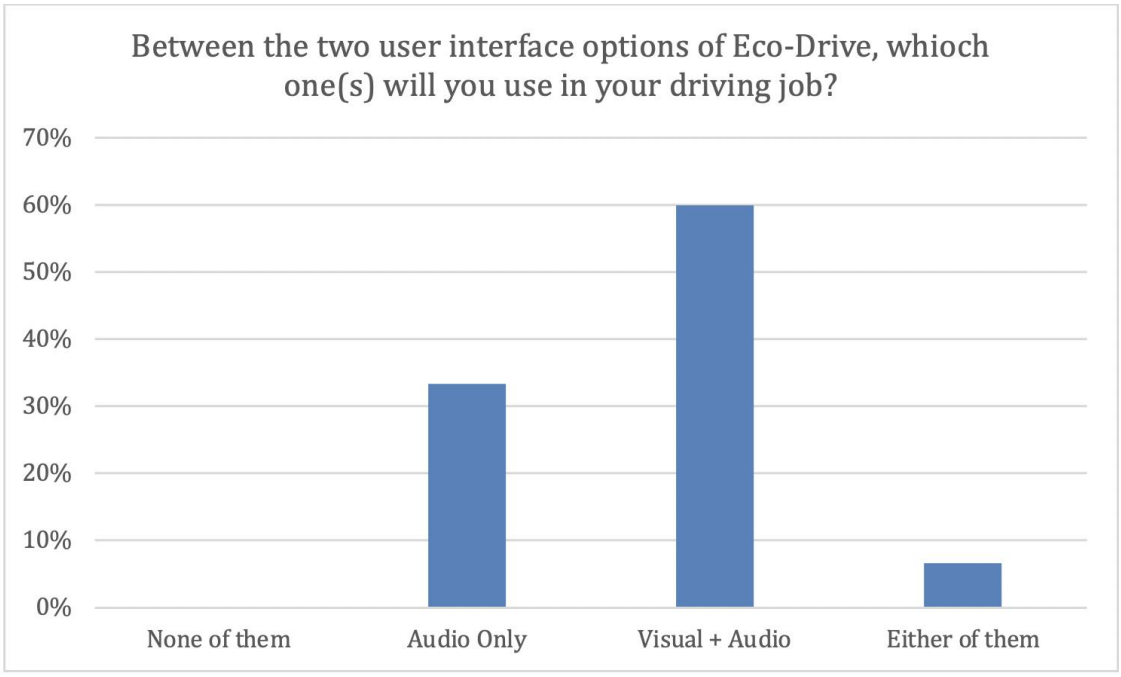


Figure 22. Distribution of responses to Question # 1 in the post-survey

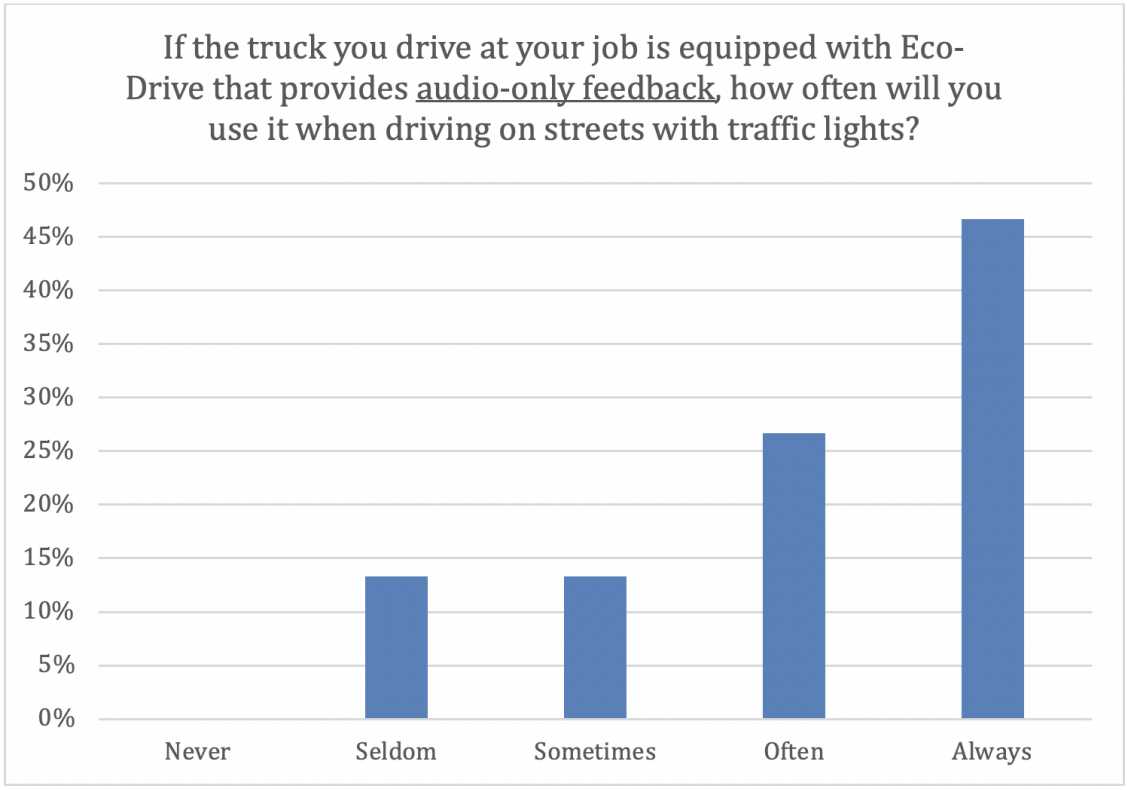


Figure 23. Distribution of responses to Question # 2 in the post-survey

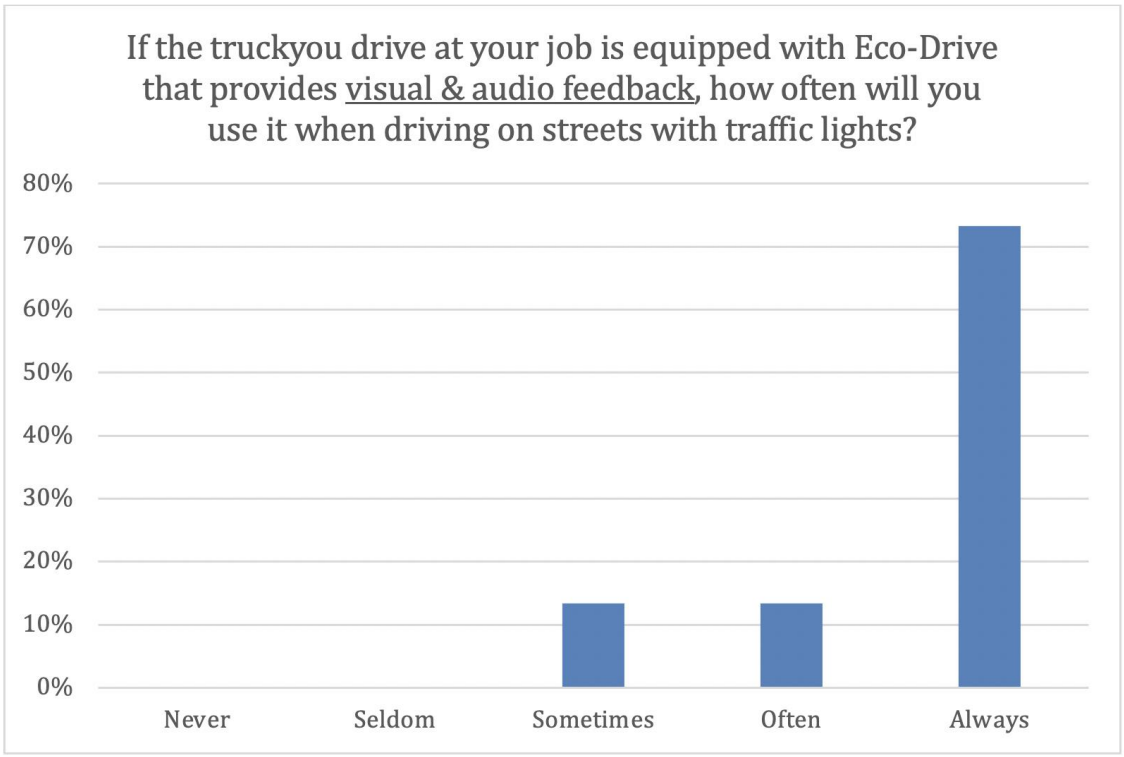


Figure 24. Distribution of responses to Question # 3 in the post-survey

Real-World Implementation of Truck EAD System

To enable the truck EAD system in the real world, the research team developed and implemented a series of hardware and software for both the vehicle and the roadway infrastructure (i.e., traffic signals). These development and implementation efforts were made with support from California Energy Commission (CEC), California Air Resource Board (CARB), South Coast Air Quality Management District (SCAQMD), Port of Los Angeles (POLA), Los Angeles County Metropolitan Transportation Authority (LA Metro), Los Angeles County's Department of Public Work (LADPW), City of Carson, City of Los Angeles' Department of Transportation (LADOT), Econolite, McCain, Western System, and Volvo Technology of America.

On the roadway infrastructure side, the research team worked with POLA, LA Metro, LADPW, City of Carson, and LADOT to deploy 15 connected signalized intersections nearby the San Pedro port complex to support a variety of connected vehicle applications. The 15 connected signalized intersections are located on three urban freight corridors, which carry a high volume of truck traffic: 1) Alameda St, 2) S. Wilmington Ave, and 3) W. Harry Bridges Blvd, as shown in Figure 25. On each of the corridors, five signalized intersections were chosen and enabled to send real-time SPaT data to the Traffic Signal Information System (TSIS) server at UCR.

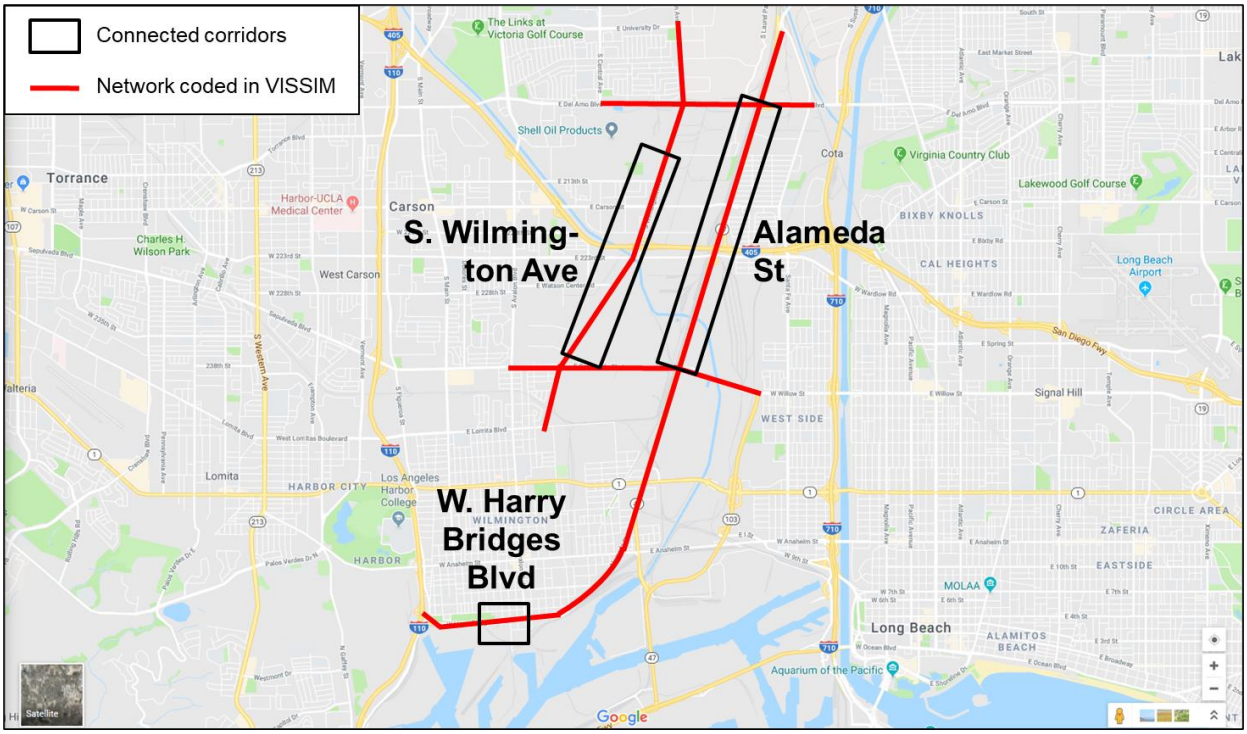


Figure 25. Connected signalized corridors near San Pedro port complex

For the five connected intersections on W. Harry Bridges Blvd, real-time SPaT data are obtained from the Traffic Management Center (TMC) of LADOT. On the other hand, the connectivity of the 10 connected intersections on Alameda St and S. Wilmington Ave is enabled by 4G/LTE modem where real-time SPaT data is sent to the TSIS server at UCR via cellular communication.

Figure 26 shows the instrumented hardware inside the traffic signal controller cabinet at one of the intersections. The router mounted in the cabinet is a rugged, industrial-graded router that can withstand temperature of up to 160 degree Fahrenheit. After connecting to the traffic signal controller, the cellular modem forwards SPaT messages from the traffic signal controller to the TSIS server over the 4G/LTE cellular network. The level of latency varies, but is usually around 1-2 seconds, which is acceptable for the truck EAD system.

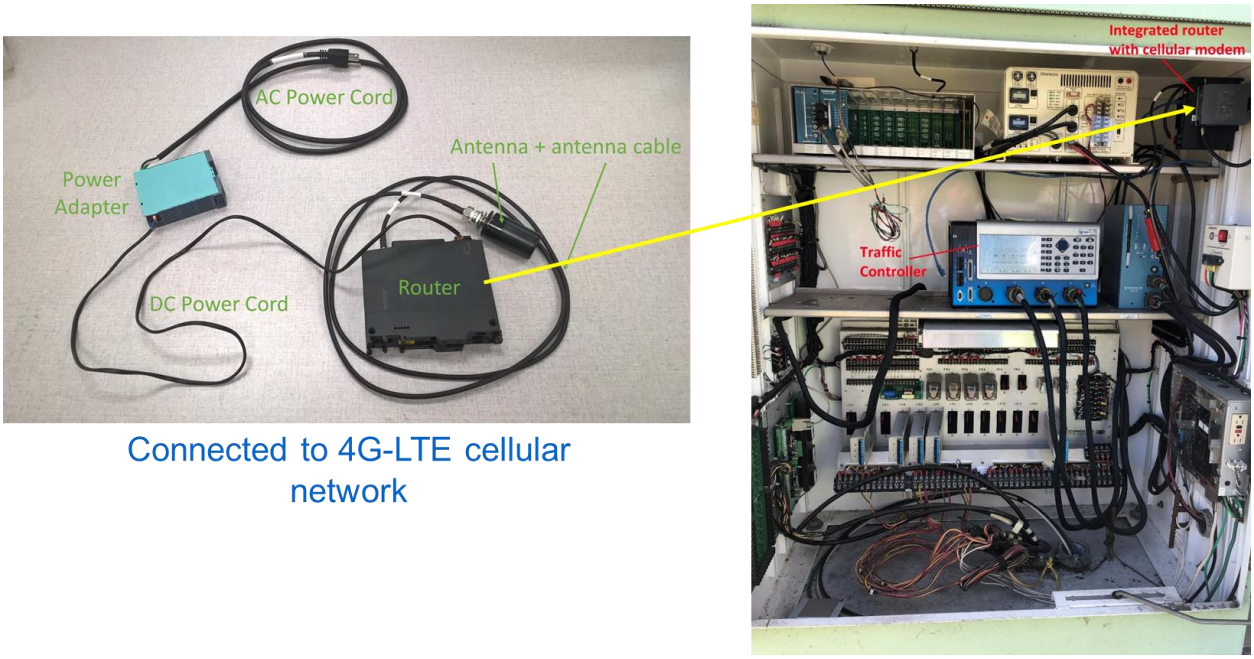


Figure 26. Instrumented communication devices in traffic signal controller cabinet

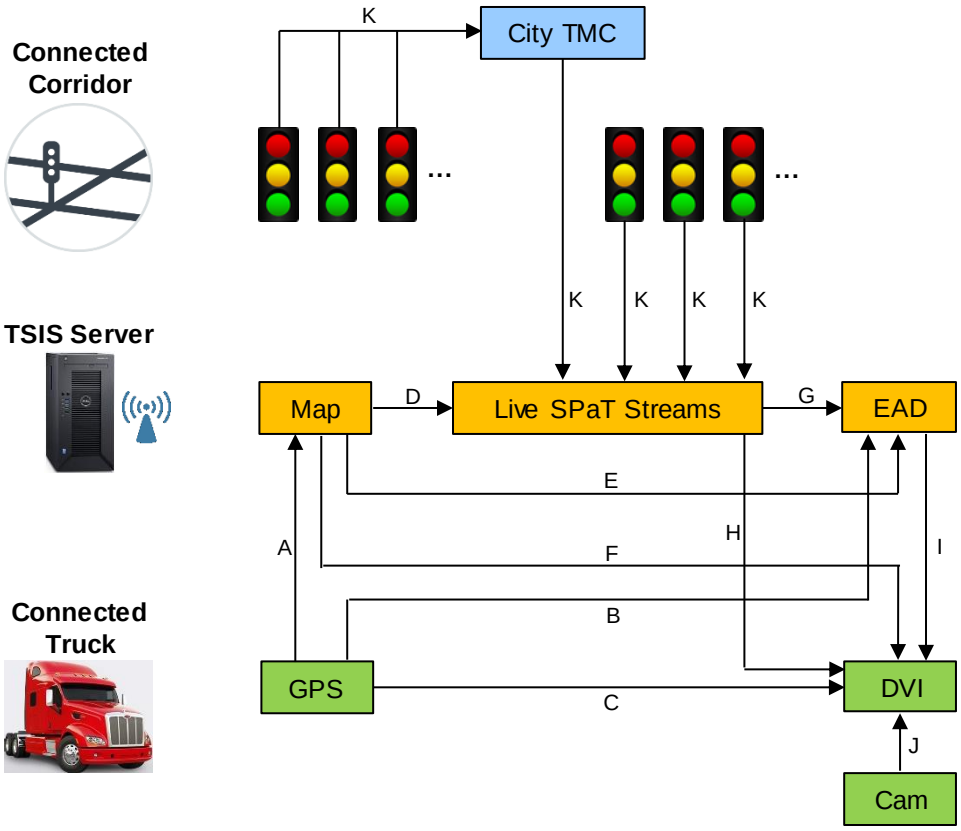
On the vehicle side, the hardware consists of an Android tablet and a Mobileye camera installed as shown in Figure 27. The Android tablet houses the EAD software application called Eco-Drive. The tablet also comes equipped with a GPS receiver and a cellular (4G/LTE) modem, allowing it to communicate with the TSIS server. The Mobileye camera is used to detect the presence of preceding vehicles and send the detection information directly to the Eco-Drive application.

Figure 28 presents the components and data flow of the cellular-based Eco-Drive application. The flow of SPaT data from the connected intersections to the connected truck is represented by arrows K and H. The TSIS server hosts a digital map of the connected signalized corridors and the EAD algorithm for calculating the advisory speed. The TSIS server uses the digital map, along with the global positioning system (GPS) data of the truck location from the tablet onboard the truck (arrow A), to identify the upcoming intersection, estimate the distance to the intersection, and determine the speed limit of the road before returning the results to the tablet to be displayed (arrow F). At the same time, the EAD algorithm takes the vehicle speed data from the GPS (arrow B), the distance-to-intersection result from the map (arrow E), and the SPaT information of the upcoming intersection (arrow G) to calculate the recommended driving speed for the connected truck, which is then sent to the tablet (arrow I). The Eco-Drive

application running on the tablet then displays the information received from all the sources on the driver-vehicle interface or user interface.

The real-world implementation of the truck EAD system was conducted in conjunction with the "The Port of Los Angeles Eco-FRATIS Drayage Truck Efficiency Project" led by the City of Los Angeles Harbor Department. More details about the instrumentation of the connected corridors, the development of Eco-Drive and TSIS server, and the implementation of EAD algorithm can be found in (City of Los Angeles Harbor Department, 2021).

Figure 27. On-board hardware of UCR's truck EAD system



Component	Description
City TMC	Traffic management center
Map	Digital map of Southern California
Live SPaT Streams	Real-time SPaT from connected intersections
EAD	Eco-Approach and Departure algorithm
GPS	GPS for identifying truck position
Cam	Camera for detecting preceding vehicle
DVI	Driver-vehicle interface on tablet

ID	Data Elements	Update Frequency (Hz)
A	Latitude, longitude, & heading	1
B	Veicle speed (GPS)	1
C	Veicle speed (GPS)	10
D	Upcoming intersection & direction	1
E	Distance to intersection	1
	Roadway speed limit	1
F	Distance to intersection	1
	Roadway speed limit	1
G	Upcoming SPaT	1
H	Upcoming SPaT	1
I	Recommended driving speed	10
J	Preceding vehicle detection signal	10
K	Real-time SPaT	1

Figure 28. Components and data flow of cellular-based truck EAD system

In addition to the truck EAD system, we also installed a combined GPS and engine control unit (ECU) data logger on the truck (see Figure 29) to collect its real-world vehicle and engine operation data at the frequency of 1 Hz. The data collected by this data logger include GPS parameters such as timestamp, speed, latitude, and longitude as well as ECU parameters such as engine speed and fuel consumption. Both sets of parameters are time aligned, which allows us to analyze truck fuel consumption at different locations and times.

Figure 29. J1939 Mini Logger™ used for data collection

Field Evaluation

The research team used the truck shown in Figure 30 for the field evaluation of the truck EAD sytem. It is a class 8 truck with 13-liter diesel engine and rated power of 455 horsepower. We hired a professional truck driver to drive this truck on a designated driving route. The driving route was a loop that included four connected intersections on Alameda St and five connected intersections on Wilmington Ave, as marked by the white circles in Figure 31. This driving route was designed to maximize the driving time on the portion of the two corridors with connected intersections. The driver was instructed to drive the route in both clockwise and counterclockwise directions so that data were collected on both directions of each corridor. The data without the use of Eco-Drive (i.e., baseline data) were collected in July 2020, and the data with the use of Eco-Drive were collected in August and September, 2020. Durign the data collection in September, the condition of the road on East Lomita Blvd that connects the two connected corridors on the south side of the loop became deteriorated. Thus, we modified the driving route by using East Sepulveda Blvd instead (i.e., brown dashed line in Figure 31) to connect between the two corridors on the south side. As a result, the data analysis was conducted on the 2.5-mile segments on both corridors, as shown in Figure 31.

Figure 30. Truck used for data collection

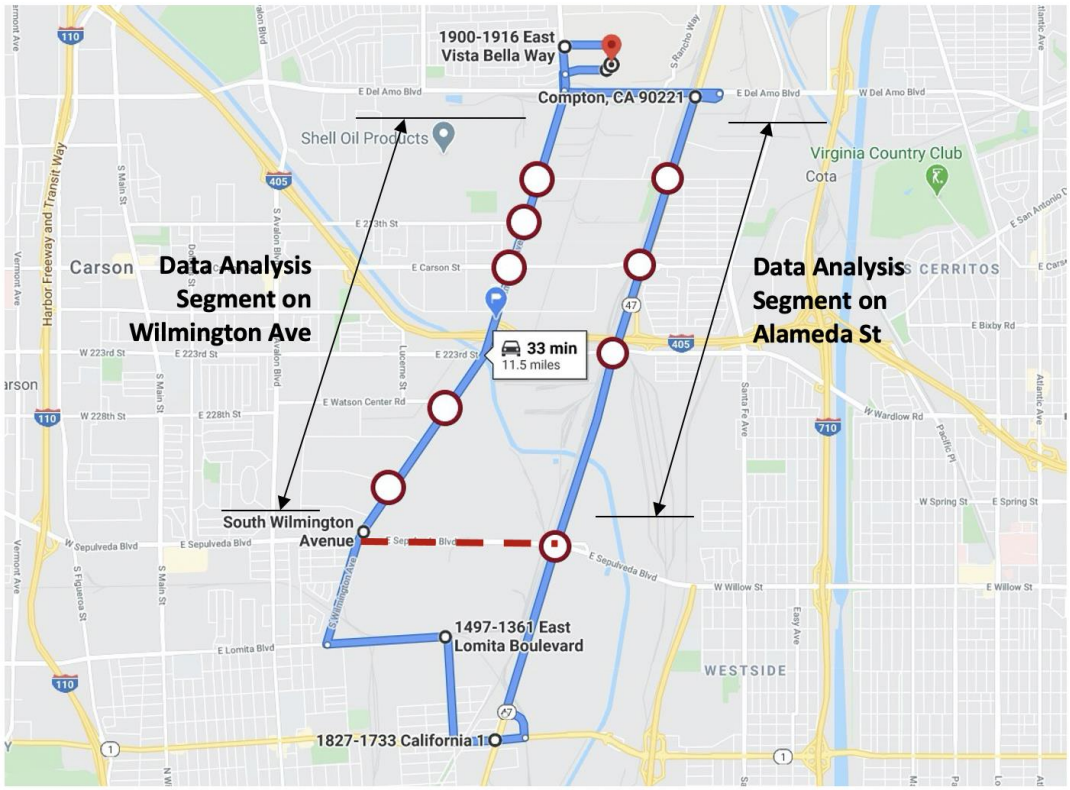


Figure 31. Driving route during data collection

Data Processing

The analysis of both baseline and Eco-Drive datasets was conducted following the steps described below.

Data Reduction

During the data collection, the data logger collected data continuously throughout the driving period. However, the analysis was focused only on the 2.5-mile segments shown in Figure 31. Note that these analysis segments exclude the portion of driving that involves making a turn at each of the four corners of the driving loop. This is to remove the influence of those turning movements on the driving performance metrics. Thus, the first step in the data processing was to reduce the data to only the portion when the truck was on the analysis segments. This was accomplished by a spatial analysis technique called *geofencing* where virtual boundaries of the analysis segments (i.e., geofences) were created on the map, and the data points inside these boundaries were identified based on their latitude and longitude information.

Road Identification

After that, the selected data points were processed through another spatial analysis technique called *map matching*. Using latitude, longitude, and heading information, each data point was assigned to a road link inside the geofences based on its proximity (a data point usually belongs to the closest road link), orientation (a data point heads into the same direction as the road link), and history (a data point is more likely to be on the same road link as the few previous data points than not). The map matching results allowed us to identify data associated with each of the four analysis segments below:

1. Alameda St Northbound
2. Alameda St Southbound
3. Wilmington Ave Northbound
4. Wilmington Ave Southbound

Performance Metrics Calculation

Using the data that were processed through the steps described above, we separated the data on each analysis segment into multiple observations where each observation represented a driving from the start of the analysis segment to the end of the analysis segment. For each data observation, we then calculated the performance metrics for the truck as listed below:

1. Travel distance (miles)
2. Travel time (seconds)
3. Average travel speed (miles per hour)
4. Travel delay (seconds)
5. Number of stops
6. Mean of acceleration (mph/s)
7. Variance of acceleration (mph^2/s^2)

8. Mean of deceleration (mph/s)
9. Variance of deceleration (mph^2/s^2)
10. Fuel consumption (liters)

Because the use of GPS data to derive these metrics could result in minor discrepancies in travel distance of each data observation, we normalized the travel time, travel delay, number of stops, and fuel consumption by the travel distance of each data observation. This resulted in the performance metrics below, which were used in the analysis instead of their original metrics:

1. Travel time per mile (seconds/mi)
2. Travel delay per mile (seconds/mi)
3. Number of stops per mile
4. Fuel consumption per mile (liters/mi)

Data Analysis

The processed data were in the form of a data table with a number of records where each record contained the calculated performance metrics for each data observation. This master data table was divided into four data tables, one for each of the four analysis segments. Then, each data table was further divided into two tables, one for the baseline case and the other for the Eco-Drive case. Therefore, there were a total of eight data tables. For each of these data tables, the data analysis was performed as follows:

Data Filtering

Data filtering was performed to remove data records with incomplete or erroneous driving data. This could occur due to unexpected interruptions during the driving, poor GPS signal resulting in erroneous data, etc. This was determined based on the travel distance where data records with the travel distance outside the range of 2.5 ± 0.3 miles were removed.

Outlier Detection

Because the driving data were collected in real-world traffic conditions which could vary greatly, there could be circumstances that caused some data observations to be drastically different from the rest. For example, an incident or a construction on the road could increase the travel time, travel delay, number of stops, and fuel consumption of a data observation considerably, making that data observation an outlier. Therefore, outliers were removed to minimize their impact on the performance metrics. Out of the 10 performance metrics that were calculated, travel time per mile, travel delay per mile, number of stops per mile, and fuel consumption per mile were the primary focus as the promise of Eco-Drive was that it could help reduce fuel consumption along signalized corridors without significantly impacting travel time by reducing number of stops and associated travel delays at connected intersections. Therefore, the data records whose values of these four metrics were considered outliers were removed. A value was considered to be an outlier if it was outside the range of mean ± 3 times the standard deviation.

Statistical Analysis

After the data filtering and outlier detection steps, the remaining data records were used to calculate various description statistics (e.g., mean and standard deviation) of each performance metric in each of the eight data tables. Then, the difference in the mean values between the baseline and the Eco-Drive cases were calculated, and the t-test was conducted to determine if the difference was statistically significant at 5% significance level.

Results

Table 3 provides descriptive statistics of the 10 performance metrics for both the baseline and the Eco-Drive cases.

Table 3. Descriptive statistics of performance metrics

	Baseline					Eco-Drive				
	Count	**Max**	**Min**	**Mean**	**S.D.**	**Count**	**Max**	**Min**	**Mean**	**S.D.**
Alameda St Northbound										
Travel distance (mi)	41	2.4	2.3	2.4	0.0	76	2.4	2.3	2.4	0.0
Travel time (s/mi)	41	117.8	79.6	98.4	10.2	76	133.3	81.4	104.9	11.3
Fuel consumed (liters/mi)	41	0.47	0.25	0.33	0.05	76	0.40	0.24	0.31	0.04
Travel delay (s/mi)	41	19.2	0.0	5.0	6.0	76	19.7	0.0	5.7	5.3
No. of stops per mile	41	0.85	0.00	0.36	0.36	76	1.26	0.00	0.39	0.30
Average speed (mph)	41	45.2	30.6	37.0	3.8	76	44.2	27.0	34.7	3.8
Mean of accel (mph/s)	41	1.6	0.2	0.8	0.3	76	1.4	0.2	0.7	0.3
Variance of accel (mph^2/s^2)	41	2.4	0.1	0.8	0.5	76	1.8	0.0	0.8	0.5
Mean of decel (mph/s)	41	-0.3	-1.8	-1.0	0.3	76	-0.2	-1.5	-0.8	0.3
Variance of decel (mph^2/s^2)	41	2.6	0.0	1.4	0.7	76	3.2	0.0	1.2	0.7

	Baseline					Eco-Drive				
	Count	**Max**	**Min**	**Mean**	**S.D.**	**Count**	**Max**	**Min**	**Mean**	**S.D.**
Alameda St Southbound										
Travel distance (mi)	56	2.4	2.3	2.4	0.0	73	2.4	2.3	2.4	0.0
Travel time (s/mi)	56	118.6	72.6	93.3	10.6	73	125.2	77.4	97.2	10.8
Fuel consumed (liters/mi)	56	0.42	0.23	0.32	0.04	73	0.40	0.22	0.30	0.04
Travel delay (s/mi)	56	14.8	0.0	3.6	3.8	73	15.2	0.0	4.2	5.0
No. of stops per mile	56	0.85	0.00	0.34	0.33	73	1.27	0.00	0.30	0.33
Average speed (mph)	56	49.6	30.4	39.1	4.5	73	46.5	28.8	37.5	4.0
Mean of accel (mph/s)	56	1.2	0.2	0.7	0.2	73	1.2	0.2	0.7	0.2
Variance of accel (mph^2/s^2)	56	1.7	0.0	0.8	0.5	73	1.9	0.0	0.7	0.5
Mean of decel (mph/s)	56	-0.2	-1.6	-0.9	0.4	73	-0.2	-1.4	-0.8	0.3
Variance of decel (mph^2/s^2)	56	2.5	0.0	1.2	0.6	73	3.9	0.0	1.0	0.7
Wilmington Ave Northbound										
Travel distance (mi)	45	2.6	2.6	2.6	0.0	63	2.6	2.6	2.6	0.0
Travel time (s/mi)	45	204.7	99.0	148.3	25.0	63	187.7	103.2	137.9	18.9
Fuel consumed (liters/mi)	45	0.48	0.26	0.38	0.05	63	0.44	0.25	0.33	0.04
Travel delay (s/mi)	45	84.2	0.0	35.3	20.4	63	59.6	1.5	25.0	16.1
No. of stops per mile	45	2.31	0.00	1.25	0.57	63	1.91	0.38	0.89	0.36
Average speed (mph)	45	36.4	17.6	25.0	4.5	63	34.9	19.2	26.6	3.6
Mean of accel (mph/s)	45	1.5	0.5	1.1	0.2	63	1.4	0.5	0.9	0.2
Variance of accel (mph^2/s^2)	45	2.4	0.1	1.3	0.4	63	2.1	0.4	1.2	0.4
Mean of decel (mph/s)	45	-0.5	-1.9	-1.3	0.3	63	-0.5	-1.7	-1.0	0.2
Variance of decel (mph^2/s^2)	45	3.1	0.5	1.7	0.5	63	2.1	0.4	1.3	0.4

	Baseline					Eco-Drive				
	Count	Max	Min	Mean	S.D.	Count	Max	Min	Mean	S.D.
Wilmington Ave Southbound										
Travel distance (mi)	34	2.6	2.5	2.6	0.0	60	2.6	2.5	2.6	0.0
Travel time (s/mi)	34	196.1	112.2	146.6	23.7	60	165.3	99.1	130.2	15.3
Fuel consumed (liters/mi)	34	0.49	0.28	0.40	0.05	60	0.40	0.26	0.34	0.03
Travel delay (s/mi)	34	58.3	1.2	29.5	16.8	60	38.4	0.0	15.1	9.8
No. of stops per mile	34	2.69	0.38	1.28	0.59	60	1.92	0.00	0.87	0.44
Average speed (mph)	34	32.1	18.4	25.1	4.1	60	36.3	21.8	28.0	3.3
Mean of accel (mph/s)	34	1.6	0.6	1.1	0.2	60	1.3	0.5	0.9	0.2
Variance of accel (mph^2/s^2)	34	2.3	0.5	1.3	0.4	60	1.6	0.4	1.0	0.3
Mean of decel (mph/s)	34	-0.6	-2.1	-1.3	0.3	60	-0.4	-1.4	-1.0	0.2
Variance of decel (mph^2/s^2)	34	4.1	0.7	2.0	0.7	60	2.5	0.2	1.3	0.4

As mentioned earlier, travel time per mile, travel delay per mile, number of stops per mile, and fuel consumption per mile are the primary focus of Eco-Drive performance evaluation. This is because the promise of Eco-Drive is that it can help reduce fuel consumption along signalized corridors without significantly impacting travel time by reducing number of stops and associated travel delays at connected intersections. In addition, Eco-Drive can help smooth the driving speed profile through reductions in the frequency and magnitude of acceleration and deceleration events, which will translate to lower mean acceleration and mean deceleration values. Table 4 shows the differences in mean values of these key performance metrics between the baseline case and the Eco-Drive case, along with the indicator of their statistical significance. The results for each analysis segment are discussed below.

Table 4. Differences in mean values of performance metrics between baseline and Eco-Drive

	Alameda St NB	Alameda St SB	Wilmington Ave NB	Wilmington Ave SB
Travel time (s/mi)	6.5%*	4.1%*	-7.1%*	-11.2%*
Travel delay (s/mi)	13.3%	19.4%	-29.2%*	-48.7%*
No. of stops per mile	7.6%	-11.3%	-28.8%*	-32.0%*
Fuel consumed (liters/mi)	-6.1%*	-6.2%*	-12.0%*	-15.0%*
Mean of accel (mph/s)	-8.2%	-11.8%*	-15.5%*	-18.8%*
Mean of decel (mph/s)	-18.1%*	-16.3%*	-23.4%*	-23.6%*

**Statistically significant at 5% significance level*

Alameda St Northbound

The results show that, on average, fuel consumption in the Eco-Drive case was 6.1% less than in the baseline case, and the difference was statistically significant at 5% significance level. However, travel time, travel delay, and number of stops in the Eco-Drive case were 6.5%, 13.3%, and 7.6% more than in the baseline case, respectively, although it should be noted that only the travel time difference was statistically significant. The lower fuel consumption in the Eco-Drive case is as expected, but the higher travel time is not. To understand the reasons behind these results, we plot the speed profiles of the truck along Alameda St Northbound for the baseline case and the Eco-Drive case as shown in Figure 32 and Figure 33, respectively. Note that there are 41 speed profiles in the baseline case and 76 speed profiles in the Eco-Drive case. It can be seen that the free-flow speeds of the truck in the baseline case in Figure 32 were generally higher than those in the Eco-Drive case in Figure 33. This explains why the travel time in the baseline case was less than in the Eco-Drive case. Note that the speed limit on this analysis segment is 45 mph. It is obvious that the truck was exceeding the speed limit more often in the baseline case. This implies that Eco-Drive might help the driver better comply with the speed limit as Eco-Drive never suggested a driving speed higher than the speed limit. Also, the speed limit information was shown on the Eco-Drive screen all the time, reminding the driver of the speed limit of the road that the driver was driving on. In addition, the better compliance with speed limit also helped reduce speed fluctuations around the cruising speed. As reported in Table 4, driving with Eco-Drive resulted in 8.2% lower mean acceleration and 18.1% lower mean deceleration than driving without it, where the lower mean deceleration was statistically significant. These effects likely contributed to less fuel consumption in the Eco-Drive case as smooth driving with few acceleration and deceleration events is known to result in higher fuel efficiency.

Alameda St Southbound

The results in Table 4 shows 11.3% fewer number of stops and 6.2% less fuel consumption in the Eco-Drive case as compared to the baseline case, where the fuel consumption difference was statistically significant. On the other hand, travel time and travel delay in the Eco-Drive case were 4.1% and 19.4% more than in the baseline case, respectively, where the travel time difference was statistically significant. These results are similar to the results for Alameda St Northbound. Figure 34 and Figure 35 show the speed profiles of the truck along Alameda St Southbound for the baseline case and the Eco-Drive case, respectively. By comparing these two figures, it can be seen that the free-flow speeds of the truck in the baseline case were generally higher than those in the Eco-Drive case, which explains why the travel time in the baseline case was less than in the Eco-Drive case. As in the case of Alameda St Northbound, Eco-Drive seemed to help the driver better comply with the speed limit, which is 45 mph on this analysis segment. In addition, the better compliance with speed limit also helped reduce speed fluctuations around the cruising speed. As reported in Table 4, driving with Eco-Drive resulted in 11.8% lower mean acceleration and 16.3% lower mean deceleration than driving without it, both of which were statistically significant. These effects likely contributed to less fuel consumption in the Eco-Drive case.

Wilmington Ave Northbound

The results in Table 4 show that fuel consumption in the Eco-Drive case was 12.0% less than in the baseline case, which was statistically significant. In addition, travel time and travel delay in the Eco-Drive case were 7.1% and 29.2% less than in the baseline case, respectively, which were statistically significant as well. These fuel and travel time savings can be attributable to the 28.8% reduction in number of stops, 15.5% lower mean acceleration, and 23.4% lower mean deceleration in the Eco-Drive case, all of which were statistically significant. These effects can be seen by comparing the truck speed profiles in the baseline case in Figure 36 with those in the Eco-Drive case in Figure 37, especially over the three consecutive connected intersections with Carson St, 213th St, and Dominguez St. Taking the intersection with 213th St as an example, the truck came to a full stop once out of 63 times that it passed that intersection with Eco-Drive, which is a smaller fraction than one out of 45 times that it passed that intersection without Eco-Drive. In addition, at the intersection with Dominguez St, it is evident that the fraction of passes without stopping at the intersection in the Eco-Drive case was higher than in the baseline case.

Wilmington Ave Southbound

Similar to Wilmington Ave Northbound, the results for Wilmington Ave Southbound show that fuel consumption in the Eco-Drive case was 15.0% less than in the baseline case, and the difference was statistically significant. In addition, travel time and travel delay in the Eco-Drive case were 11.2% and 48.7% less than in the baseline case, respectively, and both differences were statistically significant. These fuel and travel time savings can be attributable to the 32.0% reduction in number of stops per mile, the 18.8% lower mean acceleration, and the 23.6% lower mean deceleration in the Eco-Drive case. These effects can be seen by comparing the truck speed profiles in the baseline case in Figure 38 with those in the Eco-Drive case in Figure 39, especially over the connected intersections with 213th St and Dominguez St. It is evident that the fraction of passes without stopping at those intersections in the Eco-Drive case was higher than in the baseline case.

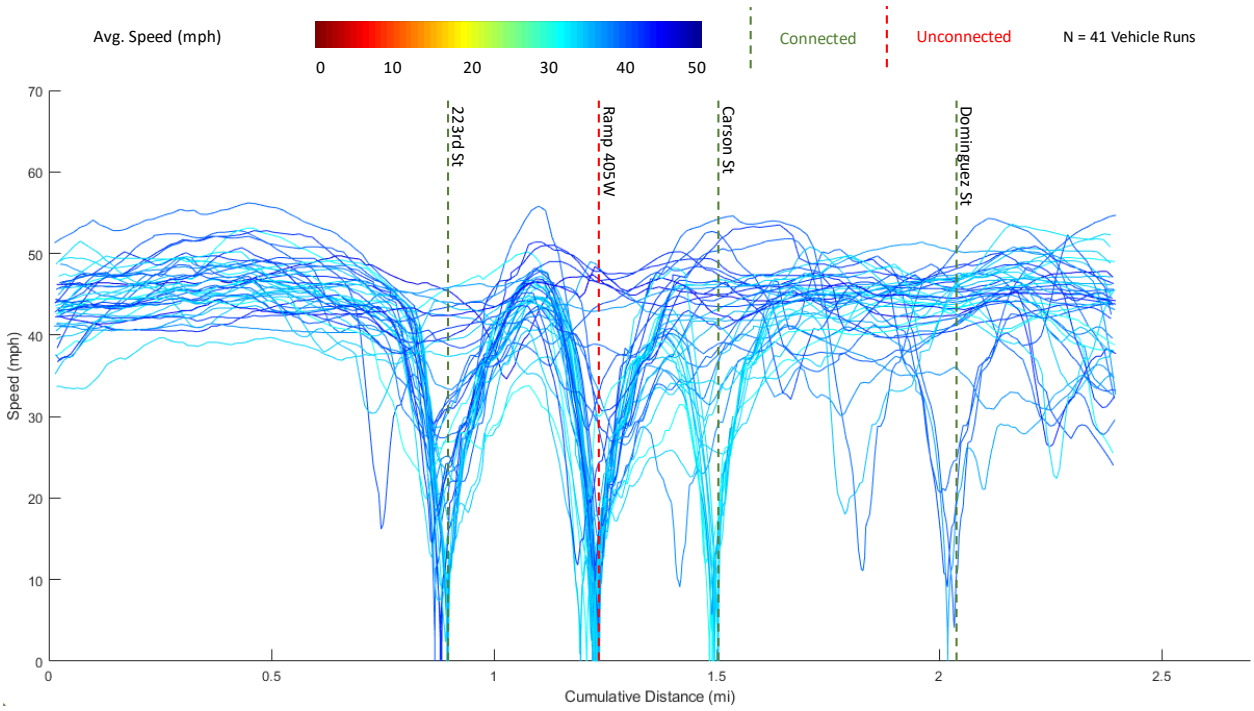


Figure 32. Speed profiles of the truck along Alameda St Northbound without Eco-Drive

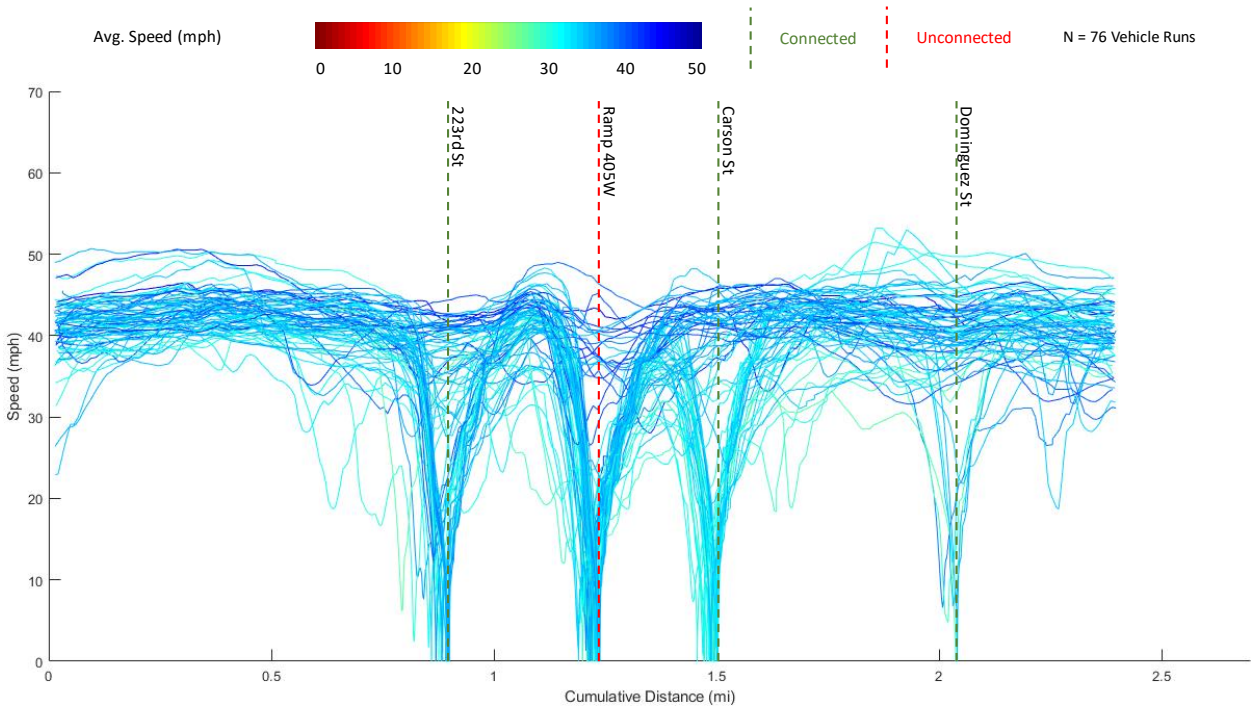


Figure 33. Speed profiles of the truck along Alameda St Northbound with Eco-Drive

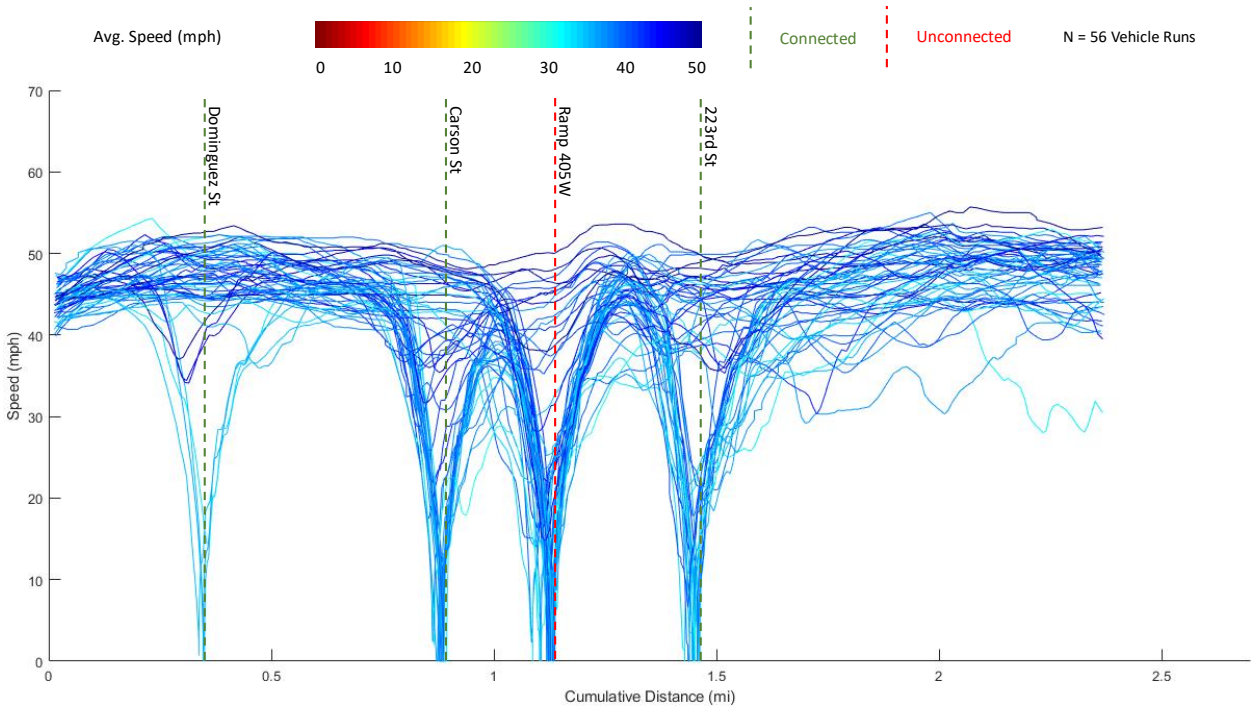


Figure 34. Speed profiles of the truck along Alameda St Southbound without Eco-Drive

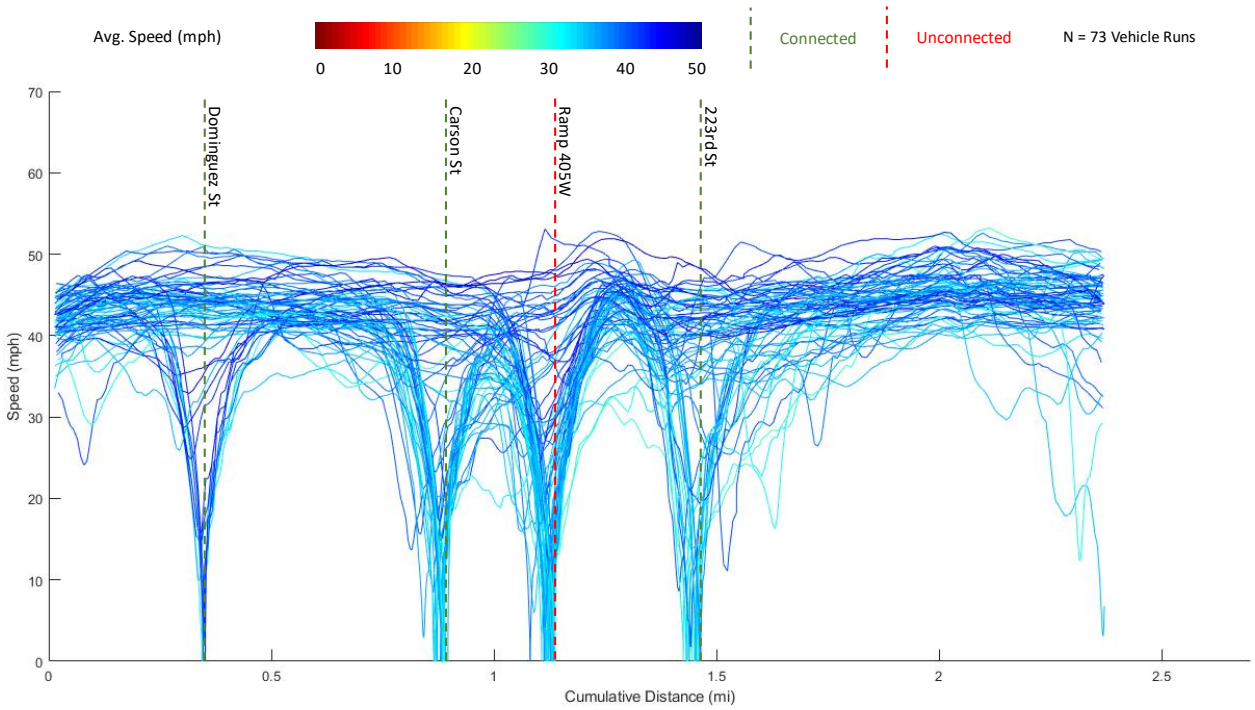


Figure 35. Speed profiles of the truck along Alameda St Southbound with Eco-Drive

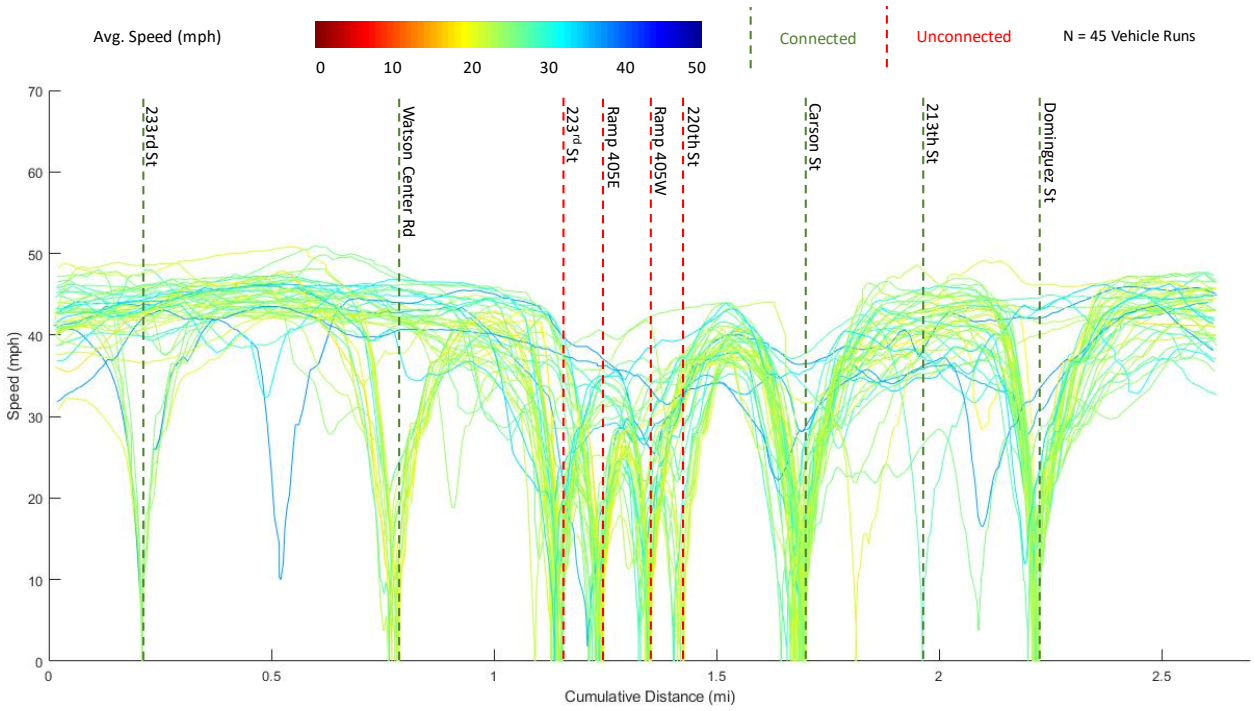


Figure 36. Speed profiles of the truck along Wilmington Ave Northbound without Eco-Drive

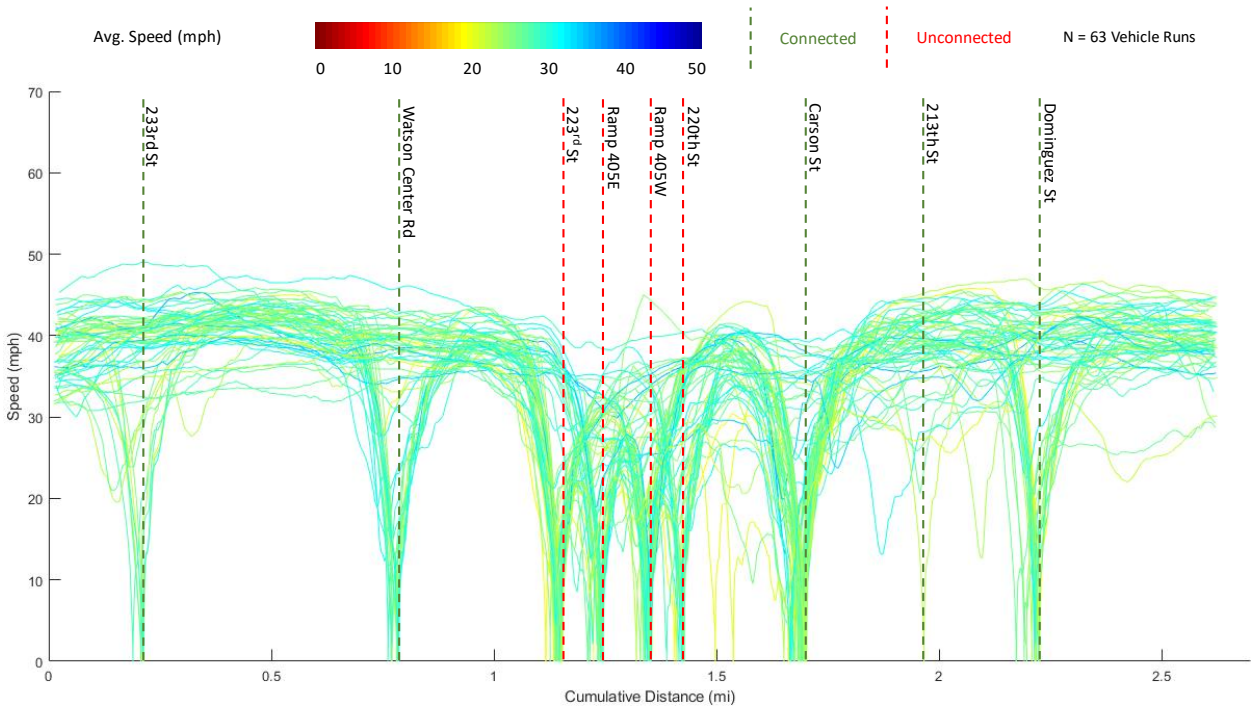


Figure 37. Speed profiles of the truck along Wilmington Ave Northbound with Eco-Drive

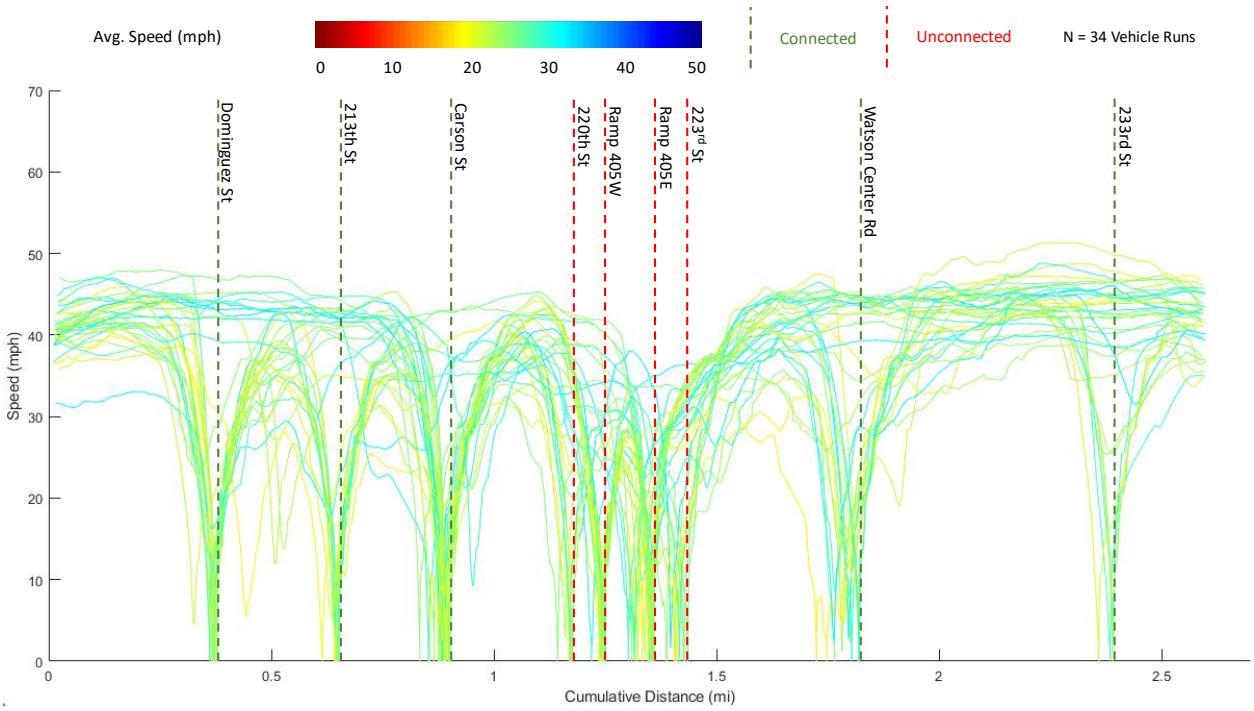


Figure 38. Speed profiles of the truck along Wilmington Ave Southbound without Eco-Drive

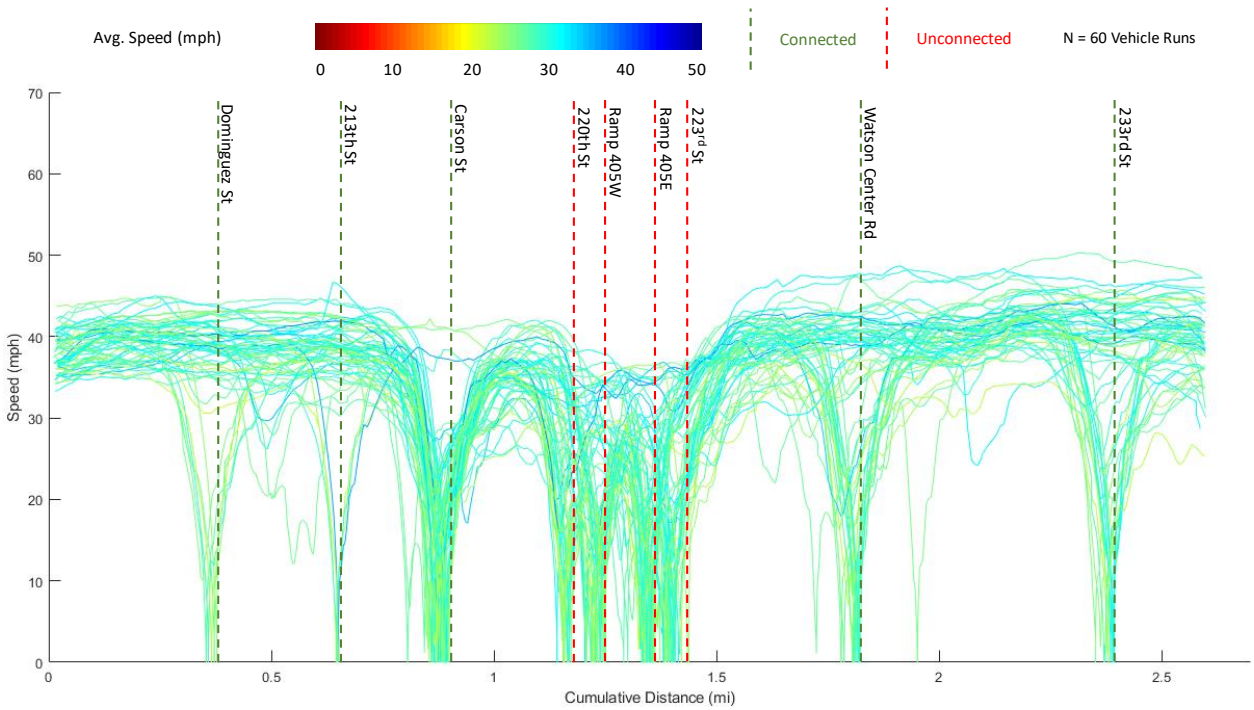


Figure 39. Speed profiles of the truck along Wilmington Ave Southbound with Eco-Drive

Conclusions

Eco-approach and departure or EAD at signalized intersections has been shown to be one of the most effective CV applications for energy savings and emission reductions. Research and development of EAD applications had been focused on passenger cars, and it was only recently that EAD was applied to heavy-duty trucks. Nevertheless, most of the research on EAD for heavy-duty trucks to date has primarily been conducted in numerical or traffic simulation environment. In this project, the research team designed and evaluated user interfaces of an EAD system in a heavy-duty truck driving simulator. Then, we implemented the promising user interface design in an actual truck EAD system and conducted an evaluation of its effectiveness in real world.

The EAD user interface designs that were evaluated include the audio-only interface and the audio & visual interface. Results from the driving experiment show that in general, the audio & visual interface results in a similar or higher level of reduction in almost all the metrics. On average, the audio-only interface results in reductions in travel time, fuel, CO2, NOx, and PM2.5 by 10%, 2%, 2%, 4%, and 1% respectively, while the audio & visual interface results in reductions in travel time, fuel, CO2, NOx, and PM2.5 by 10%, 4%, 4%, 5%, and 2% respectively. When the truck was the lead vehicle with no other preceding vehicle, the truck EAD system results in a higher level of reduction in fuel consumption as well as CO2, NOx, and PM emissions. On the other hand, the truck EAD system provides a higher level of reduction in travel time when there were other vehicles present. These trends are true for both the audio-only and the audio & visual configurations.

Surveys were also conducted of the truck drivers who participated in the EAD user interface evaluation experiment. According to the survey results, a majority of the drivers thinks that both the SPaT information and the advisory speed are very useful or extremely useful. It is also clear that most drivers would prefer the audio & visual interface and that they would always use the truck EAD system with this type of user interface if it is available on their trucks. Based on these findings and the superior performance of the audio & visual interface in the driving experiment, we implemented this user interface on a truck EAD system, called Eco-Drive, for a real-world evaluation.

The results from the Eco-Drive performance evaluation show that driving with Eco-Drive resulted in less fuel consumption than driving without it by 6% to 15%, but the underlying reasons for which the fuel savings were achieved varied by analysis segment. On both Northbound and Southbound of Alameda St, Eco-Drive helped the driver better comply with the speed limit, which is 45 mph on both analysis segments. It can be observed from the truck speed profiles that the truck was exceeding the speed limit less often in the Eco-Drive case than in the baseline case. In addition, the better compliance with speed limit also helped reduce speed fluctuations around the cruising speed, resulting in lower mean acceleration and mean deceleration values. Smooth driving with few acceleration and deceleration events is known to result in higher fuel efficiency. On the other hand, the better compliance with speed limit while driving with Eco-Drive caused the travel time to be longer than driving without it. This should not be viewed negatively as it was due to the baseline driving exceeding the speed limit more

frequently. In fact, the better compliance with speed limit can be viewed as another benefit of Eco-Drive as it could help improve safety for the driver and the surrounding traffic.

On both Northbound and Southbound of Wilmington Ave, Eco-Drive helped cut down number of stops at connected intersections considerably, which resulted in lower mean acceleration and mean deceleration values. These effects contributed to not only fuel savings but also travel time savings when driving with Eco-Drive on these analysis segments. Note that the fuel savings observed on Wilmington Ave Northbound and Wilmington Ave Southbound (12% and 15%, respectively) are much higher than those on Alameda St Northbound and Alameda St Southbound (6% for both). This may be because the driver was able to use Eco-Drive at five intersections on Wilmington Ave Northbound and Wilmington Ave Southbound while he could do so at only three intersections on Alameda St Northbound and Alameda St Southbound.

It should be noted that during the data collection the truck was not pulling any load (i.e., bobtailing). It is expected that Eco-Drive would provide a higher level of fuel savings than observed in this evaluation if the truck pulls a load, especially a heavy one. This is because the effects of acceleration and deceleration events on fuel consumption would be more pronounced when pulling a heavy load. Thus, avoiding those events would result in more fuel savings. Since drayage trucks pull some load for at least half of the time, whether it be empty container or loaded container, it is recommended that the effectiveness of Eco-Drive be also evaluated under those circumstances in future work.

In addition, it is recommended that the potential traffic safety benefits of Eco-Drive be studied in future research. Although the data collected in this project show anecdotal evidence that the truck driver better complied with the speed limit of the road when driving with Eco-Drive, more data from additional drivers should be collected to verify this finding. If Eco-Drive is also found to improve traffic safety in addition to climate and air quality, it will provide more reasons to support the investment in connected infrastructure that enables Eco-Drive and other CV applications.

Truck Eco-Routing

Introduction

Background

Routing is a major strategy for managing truck traffic in communities. Many cities have designated truck routes for carrying commercial vehicles between the highways and commercial zones in the city. The designation of truck routes typically takes into account road type, available right-of-way, traffic volume, clearance, safety, among others. Cities also often avoid routing trucks through residential zones due to concerns regarding traffic safety as well as air and noise pollution. Nevertheless, as land use, population, and truck traffic pattern in the area evolve, these concerns may re-emerge and truck routes may need to be updated to mitigate the impacts of truck traffic.

From the truck driver perspective, routing involves determining a specific travel route to take from an origin (e.g., the current location) to a destination (e.g., delivery location). Over the past several years, there has been proliferation of navigation systems in multiple platforms to assist truck drivers with that task. Some navigation systems can take truck-specific restrictions such as truck routes and clearance into consideration. These navigation systems primarily find the shortest distance or shortest time route between an origin and a destination. It is commonly assumed that taking either of these routes will also result in minimum fuel consumption and emissions from the vehicle. However, there are several cases where this may not be true. A shortest distance route may include roadway sections with steep road grades, requiring more energy for the vehicle to climb the hills while producing more emissions in the process. The route may also have the vehicle travel through heavily congested roadways, resulting in longer travel time and more fuel consumption and emissions. A shortest time route may have the vehicle travel longer distance, albeit on less congested roadways. Traveling at high speeds for longer distance will result in higher fuel consumption (and emissions) compared to a more direct route at lower speeds. This is especially true for heavy-duty trucks whose power-to-weight ratio is low.

Related Work

Over the last decade, there has been much research and development on new routing techniques for navigation systems. Instead of finding the shortest distance or shortest time route for the trip, these new routing techniques are aimed at finding the route that would minimize vehicle energy consumption and/or emissions. These so-called “eco-routing” techniques were focused initially on energy consumption and mostly on passenger cars (Boriboonsomsin et al., 2012; Boriboonsomsin et al., 2014). Figure 40 shows an example eco-routing application that displays multiple route options—shortest distance (blue), shortest time (purple), and least fuel consumption (green)—for a trip from Los Angeles Airport to Downton Los Angeles.

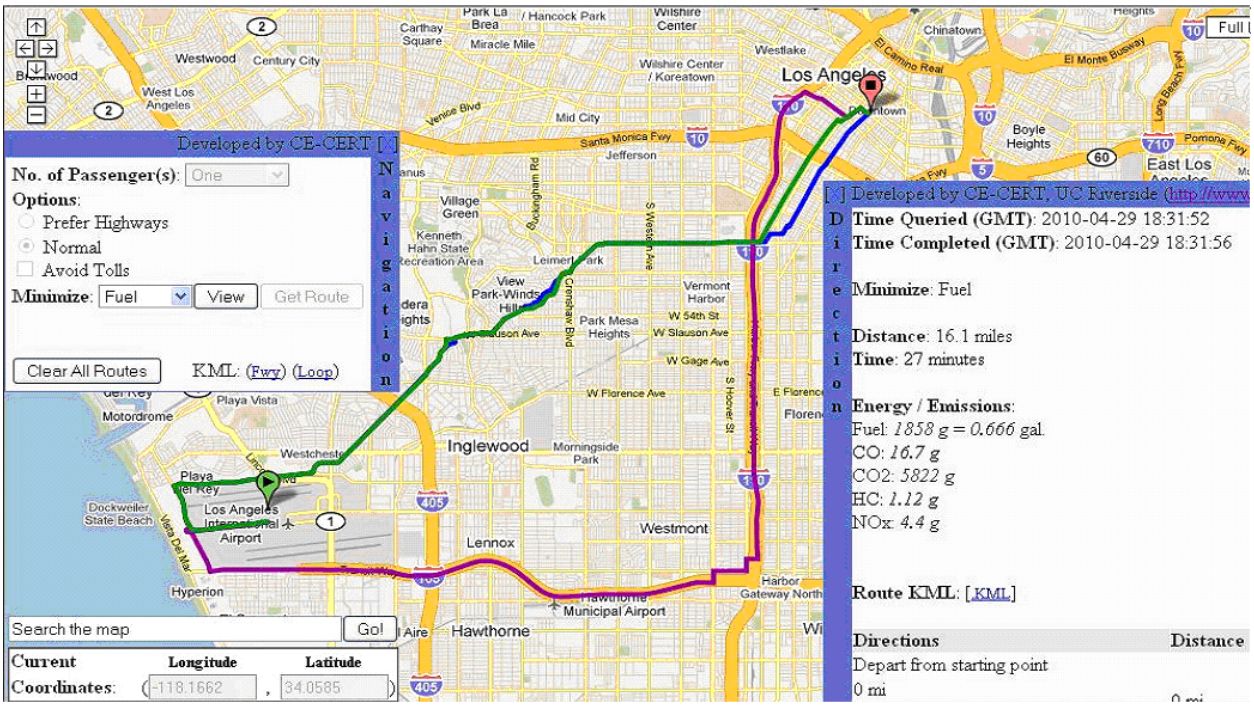


Figure 40. Eco-routing application for passenger cars (Boriboonsomsin et al., 2012)

Eco-routing techniques have also been applied to other types of vehicle including heavy-duty trucks. For instance, Scora et al. (2015) developed an eco-routing application for HDDTs and compared the least fuel consumption route with the shortest time route for more than 500,000 simulated truck trips in the Greater Los Angeles Metropolitan Area. It was found that, as compared to the shortest time route, the least fuel consumption route would require 4% to 33% less fuel, but it would increase travel time by 6% to 53%. By converting fuel and travel time into monetary values, it was found that the least fuel consumption route would result in net dollar savings for about 50% of the simulated trips. Based on this finding, it was suggested that eco-routing could be beneficial to truck drivers and fleet operators. They can choose to use the fuel-optimized route for those trips where the fuel savings justify the extra travel time.

Eco-routing techniques are also aimed at finding the route that would minimize vehicle emissions. Note that the least fuel consumption route is also the least CO2 emission route. However, this may not be true for other pollutant emissions such as NOx and PM2.5. The reason is that different emissions have different relationships with travel speed. Compared to energy-based or fuel-based routing, less attention has been given to emission-based routing, especially for pollutant emissions. This may be because while fuel cost accounts for about 20-25% of total operating cost of commercial trucking (Hooper and Murray, 2018), there is currently no incentive for truck drivers and fleet operators to reduce pollutant emissions from their trucks through routing.

A recent study by Scora et al. (2019) investigated the effect of route choice on fuel consumption and emission of HDDTs through real-world experiments. In each experiment, two identical model year 2014 HDDTs left the same origin at the same time, but took different

routes to the same destination. A total of four experiments were conducted and their results are shown in Table 5.

Table 5. Comparison of two alternative routes taken by identical heavy-duty diesel trucks for the same trips (Scora et al., 2019)

Trip	Route	Location	Duration (min)	Distance (miles)	Average Trip Speed (mph)	Fuel (Liters)	Measured NOx (g)
1	1	Riverside to	24.32	21.53	53.12	8.2	15.53
	2	Ontario	19.03	14.12	44.52	5.57	18.83
	Route Difference (%)		***21.75***	***34.42***	***16.19***	***32.07***	***-21.41***
2	1	Ontario to	48.63	43.4	53.55	15.22	27.19
	2	Vernon	50.22	42.63	50.93	13.01	31.91
	Route Difference (%)		***-3.27***	***1.77***	***4.88***	***14.52***	***-17.37***
3	1	Vernon to	45.75	24.98	32.76	10.29	37.92
	2	San Pedro	47.77	30.43	38.22	11.34	26.5
	Route Difference (%)		***-4.42***	***-21.82***	***-16.67***	***-10.20***	***30.11***
4	1	San Pedro to	48.23	28.44	35.38	12.01	20.18
	2	Corona	60.68	25.66	25.37	11.13	38.53
	Route Difference (%)		***-25.81***	***9.77***	***28.29***	***7.33***	***-90.98***

The experiment results in Table 5 confirm that the choice of travel route can have significant impacts on trip metrics including distance, duration, average speed, fuel consumption, as well as pollutant emission (in this case, NOx). Among the four trips experimented, the route with less NOx emission was not necessarily the shorter or faster route. In fact, the route with less NOx emission took longer distance, had higher average speed, and consumed more fuel (and thus, produced more CO2 emission) in all four experiments. This may be explained by the fact that the effectiveness of selective catalytic reduction (SCR), commonly used in HDDTs of model year 2010 and newer, in controlling NOx emission depends on having high enough exhaust gas temperatures, which usually occur when the truck travels at high speeds, incurring high engine load and consuming more fuel (Misra et al., 2013).

It should be noted that the impacts of travel route on NOx emission may be different for HDDTs of model years 2009 and older that are not equipped with SCR, or for future HDDTs that comply with the upcoming CARB's low-NOx engine standards for new on-road heavy-duty engines (California Air Resources Board, 2021). In the latter case where NOx emissions from HDDTs would be significantly lower than the current level, the ability of eco-routing to find a travel route that would require less fuel consumption and produce less CO2 emission from HDDTs would be beneficial.

Research Objectives

The choice of travel route for HDDTs is usually made by truck drivers or fleet operators, and the primary metric for making that route choice decision is travel time. The experiment results in Table 5 show that sometime the fastest route is already the more fuel-efficient or eco route (i.e., Trips #1 and #3). Other times when that is not the case, the eco route may not be practical. For instance, Route #2 for Trip #4 took over 12 minutes longer (26%) than Route #1, but saved less than one liter of fuel (7%). On the other hand, Trip #2 represents a scenario where truck eco-routing could be attractive for truck drivers or fleet operators. For this trip, Route #2 took just a little over a minute longer (3%) but consumed 1.2 liters or about a third of a gallon (15%) less than Route #1.

Currently, the type of information presented in Table 5 is not available to truck drivers, and it is not possible to assess how often and how much truck drivers or fleet operators could benefit from eco-routing in their typical operations. Therefore, the objectives of this research are to: 1) apply truck eco-routing to real-world truck operation data to determine the eco route for each trip, and 2) compare trip metrics of the eco route with those of the fastest route to evaluate the tradeoff between fuel consumption and travel time.

Methodology

Truck Eco-Routing Application

The truck eco-routing application used in this project is based on the UCR's eco-routing application developed in previous research (Boriboonsomsin et al., 2012; Scora et al., 2015). The various components of the application are illustrated in the block diagram in Figure 41. They include:

1. *Dynamic Roadway Network (DynaNet)* – This is a digital map of roadway network that integrates historical and real-time traffic information from multiple sources. DynaNet also contains road grade information of roadway links where available.
2. *Roadway Fuel Consumption and Emission Calculator* – At the heart of this component are mesoscale vehicle energy and emissions models, which estimate fuel consumption and emission factors per unit distance for a vehicle of a certain vehicle (plus cargo) weight traversing a roadway link with a certain road grade at a certain average speed. The fuel consumption and emission factors are later multiplied by the link length to result in the vehicle fuel consumption and emission estimates associated with that roadway link.
3. *Routing Engine* – This component consists of optimization algorithms used for calculating optimal routes.
4. *User Interfaces* – The user interfaces receive trip origin/destination inputs from the user and display route maps to the user based on various routing criteria.

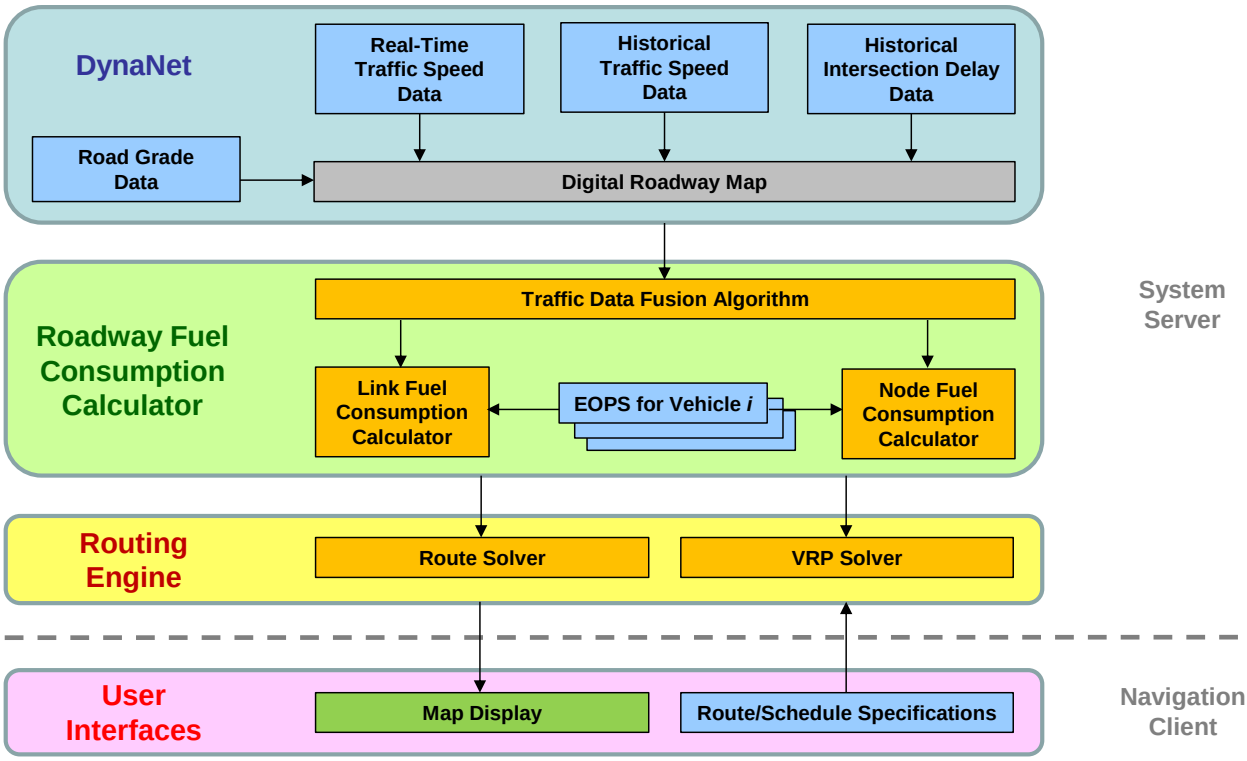


Figure 41. Components of UCR's eco-routing application

Mesoscale Fuel Consumption Model

The energy consumption and emissions production of a vehicle are related to the power requirement at the vehicle's wheels. There are several factors which determine this power requirement and they fall into two general categories—the physical characteristics of the vehicle and the trajectory or operation of the vehicle. A vehicle's physical characteristics include aerodynamic shape, frontal area, tire-road friction and vehicle mass. These characteristics determine some of the major forces that a vehicle encounters during real-world operation. Vehicle weight is an important physical characteristic to consider since it can vary substantially. This is especially true for HDDTs, which can range in weight from roughly 15,000 lbs to 80,000 lbs depending on payload. Vehicle trajectory characteristics such as vehicle speed and road grade are important factors that also significantly influence vehicle power requirements.

In order to represent truck energy and emissions in the truck eco-routing application on a roadway link-by-link basis (as opposed to on a microscale, second-by-second basis), mesoscale models are needed. Scora et al. (2015) applied multiple linear regression to develop a mesoscale truck fuel consumption model that calculates, for each link on the roadway network, fuel consumption factors that are used by the routing engine to determine the least fuel consumption route. The model relates truck fuel consumption with the key variables of average travel speed, vehicle (plus cargo) weight, and road grade (see Figure 42). The model also accounts for the effects of these variables on one another through the use of linear and non-

linear interaction terms as shown in Equation 1. It has an R^2 of 0.88, and its intercept and regression coefficients are provided in Table 6.

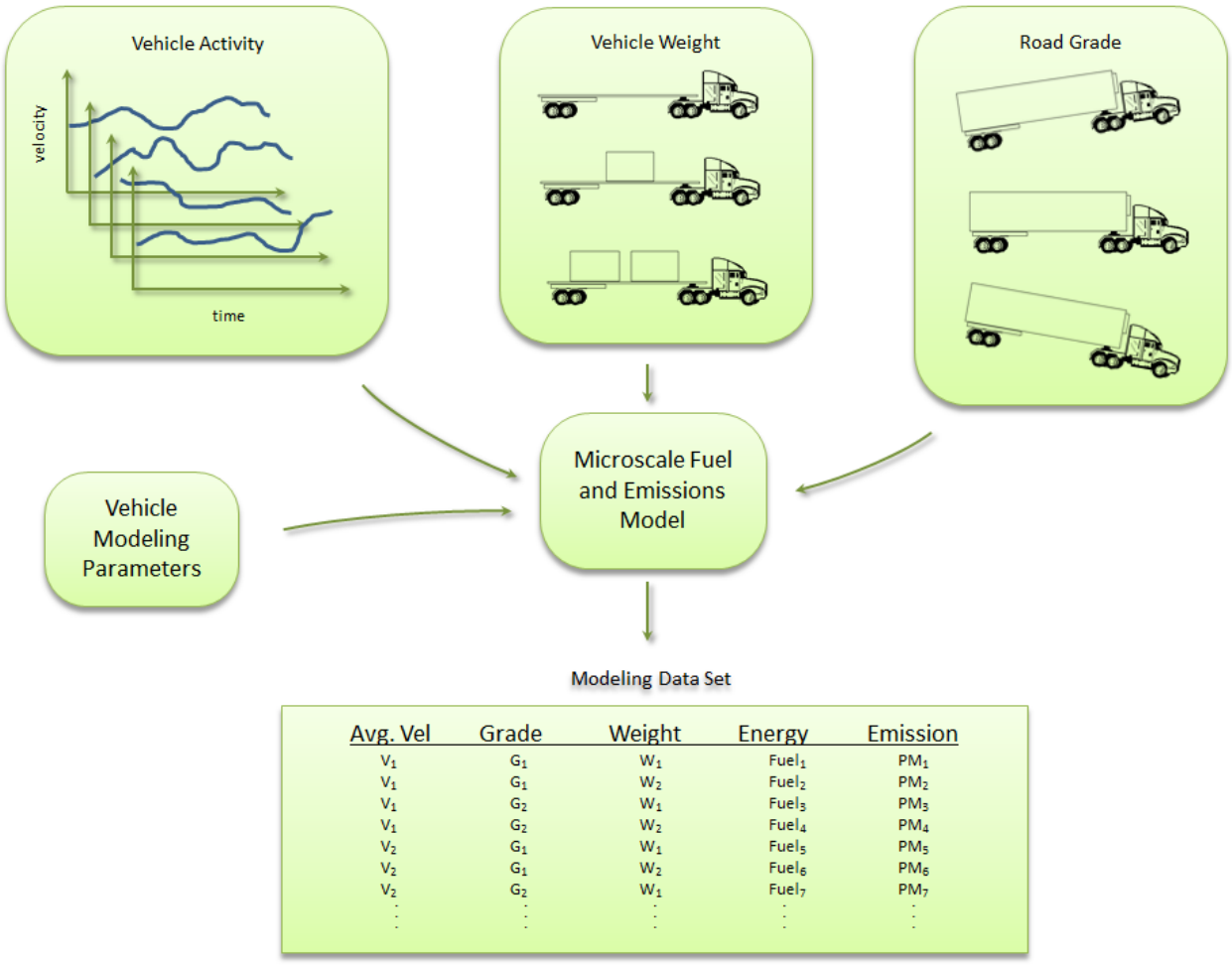


Figure 42. Schematic of modeling data set development based on vehicle activity, vehicle weight, and road grade (Scora et al., 2015)

$$f = a_0 + a_1 \cdot m + a_2 \cdot g + a_3 \cdot g \cdot m + a_4 \cdot g^2 \cdot m + a_5 \cdot v + a_6 \cdot v^2 \quad \text{(Eq. 1)}$$

where:

f	= fuel consumption
m	= vehicle mass
g	= road grade
v	= vehicle velocity
$a_0, \ldots, a_6$	= regression coefficients

Table 6. Regression coefficients in mesoscale truck fuel consumption model (Scora et al., 2015)

a_0	a_1	a_2	a_3	a_4	a_5	a_6
7.14E+02	9.82E-03	4.34E+00	2.86E-03	2.04E-04	-2.84E+01	2.82E-01

Evaluation

Real-World Truck Trip Data

In order to evaluate the potential of truck eco-routing in providing fuel savings to truck drivers and fleet operators, a real-world dataset of 48 HDDTs from a regional distribution fleet in Inland Southern California was used. The dataset contains multiple data files associated with the fleet's operation on June 16, 2020. These include:

- *Order data file* – This file contains the details of pickup and delivery orders such as order type (pickup or delivery), customer location, arrival and departure times, total cargo weight, and total number of pieces.
- *Vehicle data file* – This file contains truck ID, sum of orders, sum of total pieces, sum of total pallets, sum of total cargo weight.
- *Telemetry data file* – This file contains time series data collected from an on-board telematics device. The data includes timestamp, latitude, longitude, heading, geographic location, ignition status, speed, and odometer.

By combining the data in these three files, the research team reconstructed the full itinerary of each truck on that day. Figure 43 shows the itinerary of an example truck in the dataset. It started from the home based in Ontario and made pickups and deliveries at the locations indicated by blue dots southwest of Ontario. The label of each blue dot represents the sequence of stop and the type of stop. For example, "1-P" was the first stop of the day and it was a pickup stop. "3-D" was the third stop and it was a delivery stop. It can be observed that several stop locations are close to each other. This is by design as it is more efficient to have one truck service multiple orders in the same area. As a result, many of the trips that the truck made were short trips on surface streets.

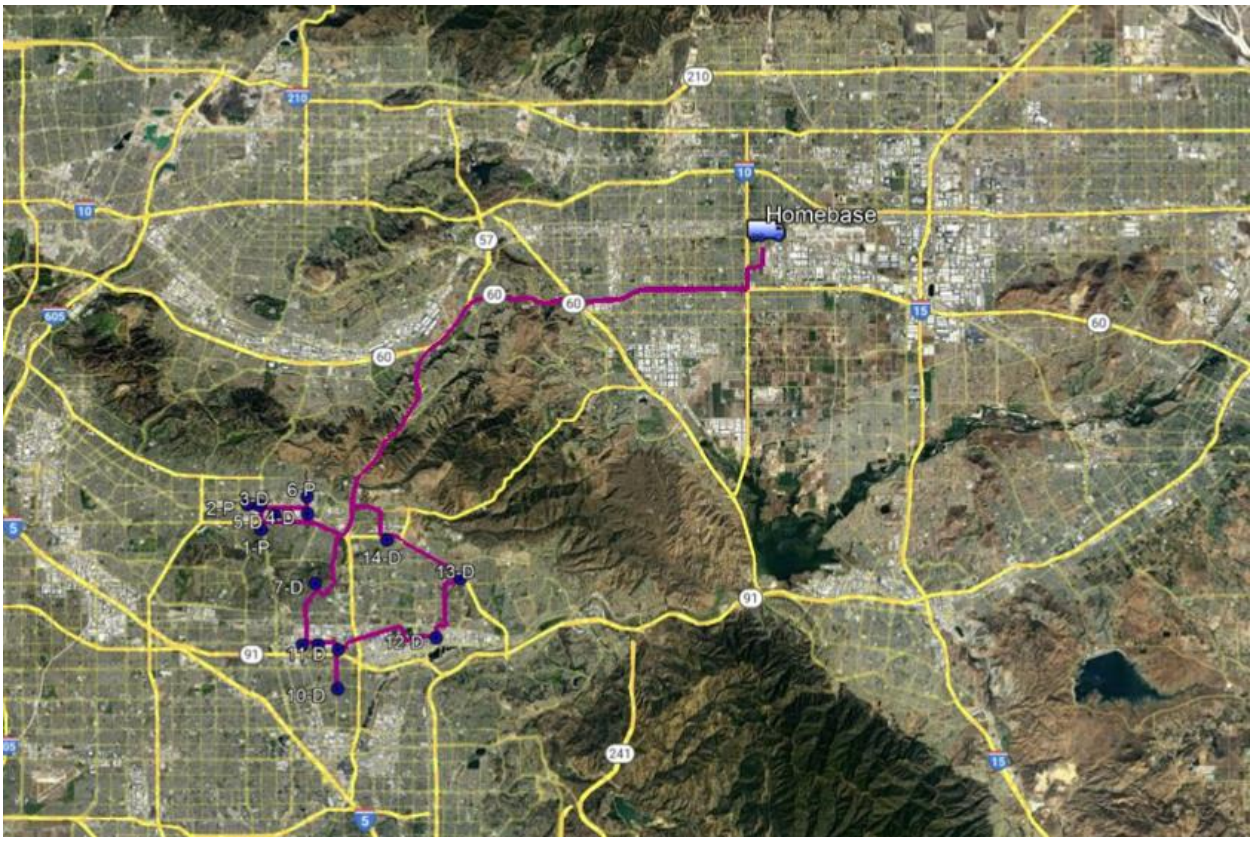


Figure 43. Itinerary of an example truck in the 48-truck dataset

During the itinerary reconstruction process, the research team also cleaned and organized the data. For example, the total number of orders that the trucks serviced on that day was 850. However, there were some cases that a truck serviced multiple orders at the same customer location at the same time. Hence, we combined those orders into one and summed the cargo weight together. In all, the 48 trucks made a total of 456 trips including the end-of-day trip back to the home base. Although the telemetry data is in time series, the time intervals between consecutive data points are not fixed, ranging from 1 minute to 15 minutes. Also, the data points do not always match up with the arrival and departure times for pickup or delivery. This is because the latter data were entered manually by the drivers and are prone to errors. Therefore, we regenerated the itinerary of each truck based on the time instances that it arrived or departed customer locations. Specifically, the itinerary contains:

- Order ID
- Event type (departure from home base, return to home base, pickup, or delivery)
- Truck ID
- Arrival time
- Departure time
- Duration (from arrival time to departure time)
- Address

- Latitude
- Longitude
- Cargo weight (for the order)
- Travel distance (from the previous order location to the current order location)

As the initial weight of each truck at the beginning of the day is unknown, it was assumed that the trucks would deliver all the cargos loaded onto them. As a result, the initial weight at the beginning of the day could be calculated by summing the cargo weight of all the delivery orders in that day. Once the initial weight at the beginning of the day had been established, the cargo weight on the truck during any trips on that day could be calculated. Thus, we had all the necessary input data for truck eco-routing, i.e., trip origin and destination, departure time, and vehicle plus cargo weight. Because the route calculation was made for past trips, historical traffic information based on the trip departure time was used in the calculation. For each of the 456 trips, the fastest route and the eco route, in terms of least fuel consumption, were calculated.

Overall Results

Table 7 summarizes the route attributes of the fastest route and the eco route for all the 456 trips, including trip distance, travel time, average speed, fuel consumption, and NOx emission. NOx emission was estimated for HDDTs of model year 2009 and older as well as HDDTs of model year 2010 and newer, using the mesoscale NOx emission models provided in Scora et al., 2019. In order to present the results from 456 trips succinctly, we grouped the trips based on how much travel time increase it would be to take the eco route instead of the fastest route. For instance, the "Maximum Travel Time Increase (%)" of "0" means that for the trips in this group, the eco route and the fastest route would take the same amount of time. The "Maximum Travel Time Increase (%)" of "2" means for the trips in this group, the eco route would take 0.1% to 2% longer than the fastest route; "5" means for the trips in this group, the eco route would take 2.1% to 5% longer than the fastest route; and so forth. The values of route attributes shown in the table are the average value for the trips in each group.

Based on the results in Table 7, the following observations were made with regards to the tradeoff between fuel savings and travel time increase from taking the eco route:

- There are 236 trips (52% of the total) in group "0" where the eco route would take the same amount of time as the fastest route. Because the average route attributes of both routes are identical, it can be concluded that for all of these trips, the fastest route is already the least fuel consumption route. These trips are mostly very short, less than 3 miles on average, and may not have more than one realistic route option to begin with.

Table 7. Comparison of route attributes between fastest route and eco route for all trips

Maximum Travel Time Increase (%)		0	2	5	10	15	20	25	30	40	50	>50	All
Number of trips		236	21	38	46	26	23	12	19	28	6	1	456
Fraction of all trips (%)		51.8	4.6	8.3	10.1	5.7	5.0	2.6	4.2	6.1	1.3	0.2	100.0
Fastest Route (FR)	Distance (miles)	2.8	6.6	8.3	12.9	15.0	19.1	18.6	23.1	20.5	23.8	24.8	8.6
	Travel time (minutes)	4.2	9.3	10.3	14.7	16.5	20.3	19.4	24.0	21.4	24.3	24.0	10.1
	Average speed (mph)	39.9	42.8	48.2	52.8	54.5	56.6	57.6	57.8	57.5	58.9	62.0	51.3
	Fuel consumption (gal)	0.5	1.1	1.6	2.6	2.8	3.4	3.5	4.6	3.9	5.4	5.3	1.6
	NOx emission, MY2009- (g)	3.1	6.9	9.0	14.2	15.8	20.1	19.9	25.7	21.9	27.1	28.5	9.4
	NOx emission, MY 2010+ (g)	15.5	16.9	21.5	35.9	39.1	57.7	49.0	64.7	72.0	75.4	50.2	28.8
Eco Route (ER)	Distance (miles)	2.8	6.1	7.9	12.3	14.1	18.3	18.2	22.4	20.1	23.6	24.7	8.4
	Travel time (minutes)	4.2	9.3	10.6	15.9	18.4	23.8	23.9	30.6	28.8	34.6	36.2	11.5
	Average speed (mph)	39.9	39.4	44.6	46.7	46.0	46.3	45.8	43.9	41.8	40.9	40.9	43.5
	Fuel consumption (gal)	0.5	1.0	1.5	2.4	2.6	3.2	3.3	4.3	3.7	5.1	5.1	1.5
	NOx emission, MY2009- (g)	3.1	6.5	8.7	13.8	15.2	19.5	19.8	25.8	22.4	28.4	29.8	9.3
	NOx emission, MY 2010+ (g)	15.5	17.7	24.3	41.5	51.1	67.7	59.0	80.5	103.6	140.9	54.0	34.6
Percent Difference (ER vs. FR)	Distance	0%	-7%	-4%	-5%	-6%	-4%	-2%	-3%	-2%	-1%	0%	-3%
	Travel time	0%	1%	3%	8%	12%	17%	23%	27%	35%	42%	51%	14%
	Average speed	0%	-8%	-8%	-11%	-16%	-18%	-21%	-24%	-27%	-30%	-34%	-15%
	Fuel consumption	0%	-7%	-5%	-7%	-8%	-7%	-6%	-6%	-7%	-5%	-5%	-5%
	NOx emission, MY2009-	0%	-5%	-3%	-3%	-4%	-3%	-1%	1%	2%	5%	5%	-1%
	NOx emission, MY 2010+	0%	5%	13%	16%	31%	17%	21%	24%	44%	87%	8%	20%

- There are 105 trips (23% of the total) in groups “2”, “5”, and “10” where the eco route would take 1% to 8% longer time than the fastest route on average, but could result in 5% to 7% fuel savings. The travel time increase from taking the eco route for these trips would be one minute or less on average, which in real world may not be noticeable. Therefore, the eco route would be considered attractive from the fuel savings versus travel time increase tradeoff perspective, especially for the 21 trips (5% of the total) that have the average travel time increase of merely 1%.
- There are 49 trips (11% of the total) in groups “15” and “20” where the eco route would take 12% to 17% longer time than the fastest route on average, but could result in 7% to 8% fuel savings. The travel time increase from taking the eco route for these trips would be 2 to 3.5 minutes on average. For these trips, the eco route could still be considered to be acceptable from the fuel savings versus travel time increase perspective.
- For the other 66 trips (15% of the total) in the remaining groups, taking the eco route would result in more than 20% longer travel time while the fuel savings would be 5% to 7% on overage. For these trips, the eco route would not make economic sense from the fuel savings versus travel time increase perspective.
- According to the average trip distance of the different groups, the eco route typically makes no difference for trips shorter than 3 miles. It is mostly attractive for trips with distance between 3 and 13 miles, generally acceptable for trips with distance between 13 and 20 miles, and usually impractical for trips with distance greater than 20 miles.

In terms of NOx emission associated with taking the eco route, the trends are opposite for HDDTs of model years 2009 and older (MY 2009-) and HDDTs of model years 2010 and newer (MY 2010+). In case of the former, taking the eco route would result in not only fuel savings but also a reduction in NOx emission at a similar level. For example, for those attractive trips where taking the eco route would result in 5% to 7% fuel savings with minimal travel time increase, doing so would also result in 3% to 5% reduction in NOx emission if the truck was of model year 2009 or older. On the other hand, for those same trips, taking the eco route would result in 5% to 16% increase in NOx emissions if the truck was of model year 2010 or newer, which are likely equipped with SCR. However, it should be noted that the CARB low-NOx engine standards will help address the issue of high NOx emission under low engine load, and consequently low SCR temperature, conditions. Therefore, the impact of taking the eco route on NOx emission may not be much of a concern for HDDTs that comply with the low-NOx engine standards.

Example Trips

For brevity, we present here the routing results of three example trips, one from each of the “Maximum Travel Time Increase (%)” groups of “2”, “5”, and “10” in Table 7, where the eco route would be attractive. The route attributes of the fastest route and the eco route for these example trips are given in Table 8, and the maps of the routes are displayed in Figure 44 through Figure 46.

For example trip #1 shown in Figure 44, the fastest route is to take the freeway I-605 to the west of the trip origin and destination, while the eco route is to take the arterial that directly

connects the trip origin and destination. Therefore, the eco route has a 28% shorter distance (8.0 miles vs. 11.1 miles), while the travel times for both routes are about the same (14 minutes). Taking the eco route would consume 27% less fuel (1.5 gallons vs. 2 gallons) and produce less NOx emission (19% for HDDTs MY2009- and 15% for HDDTs MY2010+).

For example trip #2 shown in Figure 45, the fastest route is to take the freeway SR-91 to the south of the trip origin and destination, while the eco route is to take the arterial that runs in parallel to the freeway. For this trip, the eco route has a 18% shorter distance (4 miles vs. 5 miles), but take 4% longer travel time (7.0 vs. 6.7 minutes). Taking the eco route would consume 19% less fuel (0.6 gallons vs. 0.8 gallons) and produce less NOx emission (15% for HDDTs MY2009- and 3% for HDDTs MY2010+).

For example trip #3 shown in Figure 46, the fastest route is to take the freeway I-10 to the north of the trip origin and destination, while the eco route is to take the arterial that arterial that runs in parallel to the freeway. For this trip, the eco route has a 8% shorter distance (7.6 miles vs. 8.3 miles), but take 6% longer travel time (10.2 vs. 9.6 minutes). Taking the eco route would consume 13% less fuel (1.8 gallons vs. 2.0 gallons) and produce 7% less NOx emission if truck is of model year 2009 or older. If the truck is of model year 2010 or newer, taking the eco route for this trip would produce 1% more NOx emission.

Table 8. Comparison of route attributes between fastest route and eco route for example trips

Example Trip No.		**1**	**2**	**3**
Fastest Route (FR)	**Distance (miles)**	11.1	5.1	8.3
	Travel time (minutes)	13.8	6.7	9.6
	Average speed (mph)	47.9	45.7	51.4
	Fuel consumption (gal)	2.0	0.8	2.0
	NOx emission, MY2009- (g)	11.2	5.2	9.8
	NOx emission, MY 2010+ (g)	18.8	8.7	16.4
Eco Route (ER)	**Distance (miles)**	8.0	4.2	7.6
	Travel time (minutes)	13.9	7.0	10.2
	Average speed (mph)	34.6	36.0	44.6
	Fuel consumption (gal)	1.5	0.6	1.8
	NOx emission, MY2009- (g)	9.1	4.4	9.2
	NOx emission, MY 2010+ (g)	16.0	8.4	16.5
Percent Difference (ER vs. FR)	**Distance**	-28%	-18%	-8%
	Travel time	0%	4%	6%
	Average speed	-28%	-21%	-13%
	Fuel consumption	-27%	-19%	-13%
	NOx emission, MY2009-	-19%	-15%	-7%
	NOx emission, MY 2010+	-15%	-3%	1%

According to the routing results of the three example trips, there are some common characteristics of eco route. For all three trips, the eco route is the route on a major arterial that directly connects the trip origin and destination. As a result, it has a shorter distance than

the fastest route, which requires the truck to drive away from the destination in order to get on a freeway. While the truck can travel at a higher speed on the freeway, the extra travel distance between the trip origin and the freeway entrance as well as between the freeway exit and the trip destination adds an extra travel time to the fastest route. At the end, the two routes have comparable travel times, but the eco route has a shorter distance and consumes less fuel.

Figure 44. Fastest route (purple) and eco route (green) for example trip #1

Figure 45. Fastest route (purple) and eco route (green) for example trip #2

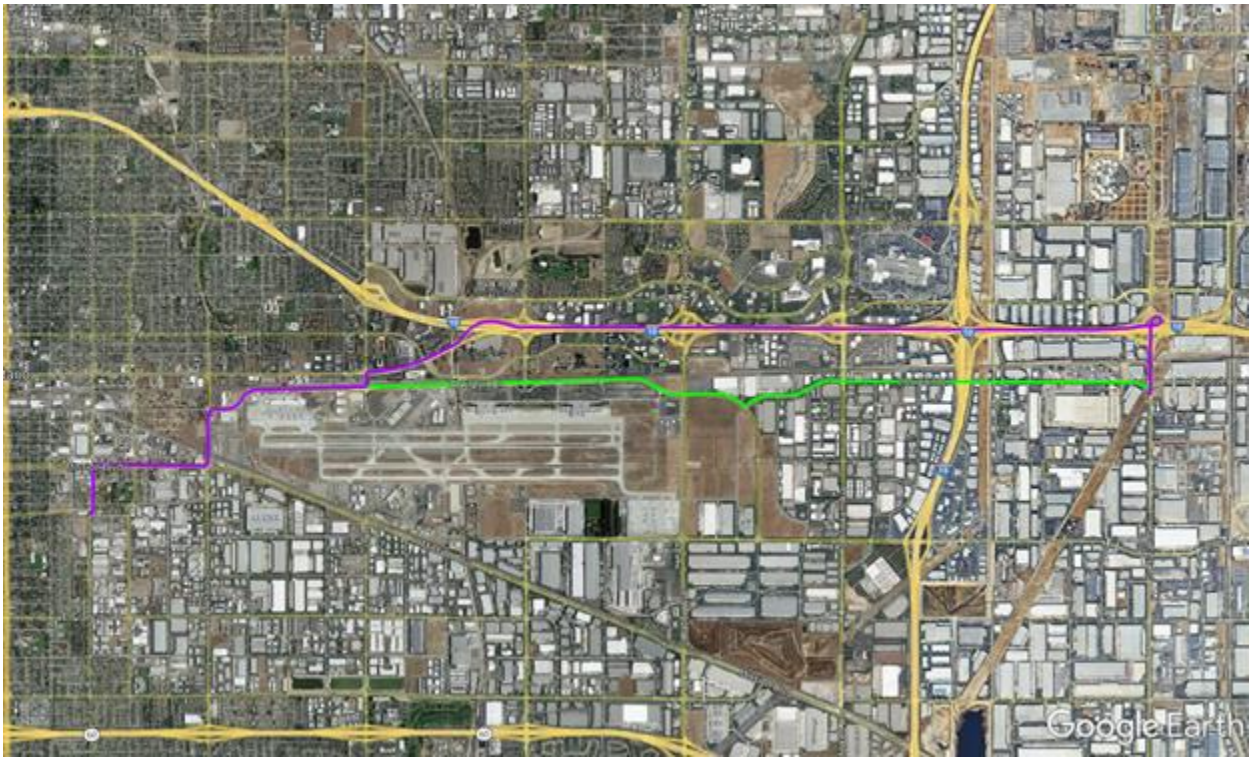


Figure 46. Fastest route (purple) and eco route (green) for example trip #3

Conclusions

Over the last decade, there has been much research and development on new routing techniques for navigation systems. These include eco-routing that is aimed at finding the route that would minimize vehicle energy consumption and/or emissions for the trip. To date, most of the research and development of eco-routing has been focused on passenger cars. Although there have been some research studies that apply eco-routing to heavy-duty trucks, they are based on hypothetical trips that may not be representative of the typical operations of truck fleets. Thus, it is not possible to assess how often and how much truck drivers or fleet operators could benefit from eco-routing in their typical operations. Therefore, in this project the research team applied eco-routing to the real-world truck operation data of a regional distribution fleet in Inland Southern California to determine the eco route for each trip, and then compared trip metrics of the eco route with those of the fastest route to evaluate the fuel savings potential from eco-routing.

Based on the results of 456 trips made by 48 trucks in a typical day, it was found that for 52% of the trips the fastest route is already the most fuel-efficient route. These trips tend to be very short, less than 3 miles on average, and may not have more than one realistic route option to begin with. For another 23% of the trips, the eco route would take up to one minute (1% to 8%) longer travel time than the fastest route on average, but would result in 5% to 7% fuel savings. This would be considered attractive from the fuel savings versus travel time increase tradeoff perspective. For another 11% of the trips, the eco route would take up to 3.5 minutes (12% to 17%) longer travel time on average, but would result in 7% to 8% fuel savings, which could still be considered to be acceptable. In other words, there is potential for truck eco-routing to help the trucks in this analysis save fuel (and reduce CO2 emission) on about a third of their trips.

The choice of travel route is usually made by truck drivers or fleet operators, and the primary metric for making that route choice decision is travel time. In this research, it is demonstrated that the eco route for any given trip can very well be different from the fastest route. By providing additional route options and the corresponding travel costs to the drivers, they can consider the tradeoff among the routes and make a choice with regards to which route to take based on their preferences and circumstances. For example, the eco route may be chosen if it would still allow the truck to arrive at the destination within a specified time window, while the fastest route may be preferred if the truck needs to meet a stringent delivery schedule. Even if the drivers take the eco route in some of their trips, it could still contribute to significant fuel savings and CO2 emission reductions, which are beneficial to the drivers and fleet operators as well as to the society. Thus, it is recommended that eco-routing be promoted or incentivized for use among truck drivers and fleet operators.

Integrated Traffic Control Strategies

Yihang Zhang, Faisal Alasiri, and Petros Ioannou
Department of Electrical and Computer Engineering, University of Southern California

Introduction

Due to the rapidly increasing demand for transportation and mobility, traffic congestion has become a significant problem worldwide. Traffic congestion has a negative impact on traffic mobility, safety, and the environment. In the United States during 2014, for example, the yearly delay time per auto commuter due to congestion was 42 hours, which increased by 13.5% compared to 37 hours in 2000. The fuel wasted in congestion was 19 gallons per commuter per year in 2014, which grew by 26.7% when being compared to 15 gallons per commuter per year in 2000 (Schrank et al. 2015). Moreover, unstable traffic flow conditions on highway segments are known to increase the possibility of collisions (Marchesini and Weijermars 2010). Therefore, efficient traffic flow control strategies are needed to avoid or reduce the severity of congestion, hence reduce fuel consumption and tailpipe emissions along the lanes of highway networks. Furthermore, in order to evaluate the environmental impact of potential traffic flow control strategies precisely and efficiently, well-defined emission models are needed to estimate or predict fuel consumption and tailpipe emissions of vehicles under different traffic scenarios in both simulations and field tests.

In highway traffic, bottlenecks often arise due to incidents, construction, merge or diverge points, and other road conditions. When traffic demand is higher than the capacity of the bottleneck, congestion occurs. One possible way to solve the highway congestion problem is to expand existing highway networks. However, this solution is usually constrained by the long building period and limited capital investment. Hence, increasing the utility of existing road infrastructure with advanced traffic control strategies is a more attractive solution to highway congestion. To prevent or relieve highway congestion, different Intelligent Transportation Systems (ITS) techniques, e.g., dynamic routing, driver information systems, variable speed limit (VSL), ramp metering (RM), and many others, are widely studied and applied to improve the efficiency of existing road networks (Abadi, Ioannou, and Dessouky 2016; Butakov and Ioannou 2015; Y. Zhang and Ioannou 2017a; Markos Papageorgiou, Hadj-Salem, and Blosseville 1991; Lu et al. 2011).

Existing works on the development and evaluation of VSL, RM and LC control have reported consistent improvements in traffic safety in theories, macroscopic simulations, microscopic simulations and field tests (Van den Hoogen and Smulders 1994; Y. Wang and Ioannou 2011a; Lu and Shladover 2014), while the impact on traffic mobility and environment is rather controversial. Although most of previous studies are able to show improvements of traffic mobility in macroscopic simulations with different traffic flow control strategies, when it comes to microscopic simulations and field tests, these improvements are not consistent under different traffic conditions or incident scenarios. In some cases, the travel time is improved and in others deteriorated due to the deployment of traffic flow controllers which raises questions as to the ability of VSL, RM and LC to improve traffic mobility (Hadiuzzaman, Qiu, and Lu 2012;

Baldi et al. 2014; Torne Santos, Rosas, and Soriguera 2011; K. Gao 2012; Kejun et al. 2008; P. Ioannou et al. 2012). Most researchers attribute the inconsistencies in travel time improvement to the highly disordered and stochastic traffic conditions at congested bottlenecks, which are difficult to predicts and regulate (P. Ioannou et al. 2012; Y. Wang and Ioannou 2011b; Torne Santos, Rosas, and Soriguera 2011; Van den Hoogen and Smulders 1994). While these arguments have an element of truth, some questions are need to be asked as follows:

1. What is the behavior of the traffic flow in a road network? Under what condition the road network will get congested and what are the reasons of the disordered behavior of the traffic flow at the bottleneck?
2. Is it possible to reduce the disorder at the bottleneck, therefore the consistency between macroscopic and microscopic simulations can be achieved?
3. Is it possible to find efficient VSL, RM and LC control strategies which are able to improve the traffic mobility at highway bottlenecks and robust to different incident scenarios?
4. Given the complexity of underlying reasons of highway congestion, is it possible to apply multiple traffic flow control strategies simultaneously in a integrated and systematic manner, such that different control strategies can work along with each other coordinately without deteriorate the benefit introduced by other control strategies.
5. Is it possible to find a traffic control strategy that can improve the traffic mobility under all possible traffic scenarios and capacity constraints as well as initial conditions?

Motivated by the above questions, this project dedicates to the design, analysis and evaluation of integrated traffic flow control strategies which is able to provide consistent improvement in traffic mobility, safety and environmental impact. The goal of the integrated controller is to stabilize and homogenize the traffic flow upstream a highway bottleneck, therefore improve the traffic mobility, safety and the environmental impact. We also evaluate the robustness of the integrated controller with respect to different levels of traffic demand, model parameters and measurement noise in both macroscopic and microscopic simulations. Furthermore, the open-loop stability properties of the modified cell transmission traffic flow model (CTM) which takes the capacity drop phenomenon into consideration under all possible traffic flow scenarios are investigated, which motivates the design of a VSL controller which is able to avoid the capacity drop, stabilize the system and maximize the flow rate at the bottleneck. The VSL controller is then extended with integral action in order to reject system disturbance.

Existing Work

In the past several decades, numerous studies have been conducted to explore the effect of VSL, RM and LC control on traffic mobility, safety and the environmental impact.

VSL control has been one of the widely studied highway traffic control technologies since the 1990s (Smulders 1990). Papageorgiou et al. studied the effect of VSL on the fundamental diagram in (Markos Papageorgiou, Kosmatopoulos, and Papamichail 2008). It is shown that VSL control decreases the slope of the fundamental diagram when the vehicle density is lower than the critical value and increases the critical density. The flow at the same density would be higher with VSL in over critical conditions.

Muralidharan et al. proposed a MPC VSL controller based on the LN-CTM model that is able to recover the bottleneck from capacity drop and obtain an optimal trajectory in the absence of capacity drop (Muralidharan and Horowitz 2015). In 2014, Frejo et al. proposed a hybrid MPC controller which combines VSL with ramp metering. The proposed method reduced the computation load of the receding horizon optimization by using genetic and exhaustive algorithms while achieving a good performance in simulation (Frejo et al. 2014).

In (Khondaker and Kattan 2015), Khondaker and Kattan designed a MPC VSL controller based on a microscopic car following model with the assumption of a connected vehicle environment. The proposed method predicts traffic conditions on the microscopic level and optimizes a cost function which is the weighted sum of TTT and time to collision (TTC), therefore improves both traffic mobility and safety. The method was evaluated using a microscopic simulation model based on the commercial software, VISSIM. Significant improvement on travel time is demonstrated. However, the authors assumed that all vehicle information is available in real time and the vehicle states can be accurately predicted, which is very difficult, if at all possible.

In 2013, Carlson et al. (Rodrigo C. Carlson, Papamichail, and Papageorgiou 2013) proposed two local feedback VSL controllers. The local feedback controllers were compared to a nonlinear optimal controller via macroscopic simulations. Results showed that the simple feedback controllers can provide similar improvement with respect to the total time spent (TTS) as the optimal controller by using much lower computational effort. The method is extended to multiple bottlenecks in (Iordanidou et al. 2015) and evaluated to be also effective in microscopic simulations in (Müller et al. 2015). In (H.-Y. Jin and Jin 2015), Jin and Jin proposed a proportional-integral (PI) VSL controller to maximize the bottleneck throughput with only one VSL sign by locally stabilizing the vehicle density at a critical value. Since the analysis is local there is no guarantee that a traffic disturbance would not lead to a capacity drop and unstable situation. In addition it is not clear how the design for one section can be extended to multiple sections upstream the bottleneck.

In (A. Hegyi et al. 2008), Hegyi et al. proposed the SPECIALIST VSL controller strategy based on shockwave theory. The SPECIALIST method detects the shockwave upstream the bottleneck and uses VSL to make the shockwave accumulate slower and dissipate faster thus dampen the shockwave and improve traffic mobility. In (J. Zhang, Chang, and Ioannou 2006), a local feedback VSL control strategy integrated with ramp metering is proposed based on the fundamental diagram. An extended version of this control strategy is evaluated in (Chang et al. 2007) with microscopic simulations. The method is shown to be able to improve freeway efficiency as well as be robust with respect to modeling error and measurement noise.

The effects of VSL on traffic safety and the environment is also assessed. In (Abdel-Aty, Dilmore, and Dhindsa 2006), Abdel-Aty et al. showed that well-configured VSL strategies can decease the crash likelihood but large gaps of speed limit in time and space may increase it. No improvement in travel time is observed in this study. In (Li et al. 2014), a genetic algorithm was used to choose the control parameters in order to minimize the rear-end collision risks near freeway recurrent bottlenecks. With the proposed control strategy, the VSL control reduced the

rear-end crash potential by 69.84% for the high demand scenario and by 81.81% for the moderate demand scenario. (Y. Zhang and Ioannou 2016) evaluated the environmental impact of the VSL and LC control method proposed in (Y. Zhang and Ioannou 2015) with a microscopic emission model CMEM and a macroscopic one MOVES. It is shown that the environmental benefits are evaluated to be qualitatively similar with both models while the microscopic CMEM is more sensitive to transient process.

In (Zegeye et al. 2009), a MPC VSL strategy was proposed using a car-following model to reduce both total time spent (TTS) and total emissions. It is shown that a reduction of TTS alone may not reduce the total emissions. (Khondaker and Kattan 2015) showed that in case of 100% penetration rates of connected vehicles, optimizing for safety alone is enough to achieve simultaneous and optimum improvements in all measures. However, in case of lower penetration rate, a higher collision risk was observed when optimizing for only mobility or fuel consumption.

The aim of RM is to adjust the on-ramp flow into the mainline in order to improve the overall traffic condition. RM has been widely used in United States and the Europe (Horton et al. 2016; Caltrans 2016). ALINEA, one of the most popular RM strategies, is a heuristic local feedback control method with integral action (Markos Papageorgiou, Hadj-Salem, and Blosseville 1991).

In (Oh and Sisiopiku 2001), ALINEA is expanded to MALINEA, which includes the mainline occupancy upstream the on-ramp in the feedback loop. MALINEA addresses two main disadvantages to ALINEA. The first is that although ALINEA optimizes the occupancy downstream of the entrance ramp, congestion can still occur upstream of the ramp. The second is that the optimal detector location can be difficult to determine.

(Smaragdis and Papageorgiou 2003) proposed FL-ALINEA which includes feedback downstream flow rate instead of occupancy and ALINEA/Q algorithm which combines queue control with ALINEA. MALINEA addresses two main disadvantages to ALINEA. The first is that although ALINEA optimizes the occupancy downstream of the entrance ramp, congestion can still occur upstream of the ramp. The second is that the optimal detector location can be difficult to determine. Its formula is identical to the formula used for traditional occupancy-based ALINEA, except that it measures flow, and tries to reach a set point flow rather than a set point occupancy. However, when the occupancy is over the critical occupancy, the metering rate is set to the minimum rate, since the freeway is already over capacity. ALINEA/Q algorithm calculates two metering rates. The first rate is calculated exactly the same as in the traditional ALINEA algorithm. The second rate that is calculated is the minimum rate needed to keep the ramp queue at or below the maximum allowable queue length. The final calculated rate is the greater of either the ALINEA rate or the queue control rate.

Some model-based RM algorithms are also developed. Coordinated ramp metering is based on a second order traffic flow model and an optimal control approach that decides the metering rates of multiple ramps in a coordinated manner (M. Papageorgiou and Kotsialos 2002). Coordinated ramp metering is basically a vectorization of the ALINEA equation, which uses vectors of occupancy, and 2 control gain matrices to return a vector of metering rates.

SWARM is a data-based ramp metering strategy which uses linear regression of measured data to predict the density (Caltrans District 7 2006). Despite the intensive application of RM, it is recognized that ramp metering can only control the vehicle density immediately downstream the on-ramp therefore barely improves the overall traffic condition in practice, especially when the mainstream demand is high (Lu, Qiu, et al. 2010; Scariza 2003). The above limitations of RM motivates the investigation of combining ramp metering with mainline traffic control strategies such as VSL.

Previous efforts to study the effect of lane changes at bottlenecks and develop traffic flow control strategies with consideration of lane management include the following:

In 1986, Rathi et al. (Rathi and Nemeth 1986) developed a microscopic simulation model to evaluate the effect of LC control in a freeway work zone at different driver compliance rate. In 1988, Mahmassani et al. (Mahmassani and Jayakrishnan 1988) applied a macroscopic simulation model to evaluate lane closure strategy for planned work zone. The work in (Rathi and Nemeth 1986; Mahmassani and Jayakrishnan 1988) is focused on long-term lane closure strategies rather than temporary lane closures.

In 1998, Schaefer et al. (Schaefer, Upchurch, and Ashur 1998) assessed the effectiveness of overhead lane control signals. The signals are placed at 1/2 mile intervals ahead of the highway incident area and indicate lane closure with red “x” symbols. A microscopic simulation using SLAM was used to evaluate the performance of the lane change signal on time delay. In 1999, Jha et al. (Jha, Cuneo, and Ben-Akiva 1999) evaluated three different lane control signal settings for the tunnel of I-93 South. Yellow and red overhead signals were applied ahead of incident location and evaluated with microscopic simulator MITSIM. The study showed that under incident condition, TTT is sensitive to upstream road geometry and driver compliance rate. Carelessly configured LC signal settings may result in increase of TTT.

In (W.-L. Jin 2013), Jin stated that systematic lane changes can seriously deteriorate traffic safety and efficiency during lane drop, merge, and other types of bottleneck. The author introduced an entropy condition for the multi-commodity LWR model and solve the Riemann problem inside a homogeneous lane-changing area. In (Laval and Daganzo 2006), Laval and Daganzo also confirmed that lane changes at the bottleneck reduce the flow rate and result in capacity drop at the bottleneck.

In recently years, researchers start to examine the combination and integration of different traffic flow control schemes. In (Baskar, De Schutter, and Hellendoorn 2008), Baskar et al. proposed a MPC approach to find optimal speed limits and lane allocations for platoons. The method is simulated on a 2-lane highway segment and reported to improve travel time by 5% - 10%. It is assumed that all vehicles are controlled by road-side controllers. In 2014, Roncoli et al. (C. Roncoli, Papamichail, and Papageorgiou 2014) proposed a MPC-based traffic control strategy for multi-lane motorways, which integrates VSL, ramp metering and lane allocation. The authors adopted the first order flow model and treated each lane as different cells. MPC is designed based on a cost function which penalizes TTS, queue length on the ramps and

amplitude of oscillations. Simulation results show that VSL performs much better when combined with ramp metering and lane allocation.

The coordination of RM and VSL involves consideration of network mobility, on-ramp queues and fairness between the mainline and the ramps. The objective is to keep a balanced delay time between vehicles on the mainline and the ramps and avoid long queues on the ramps from spilling back to the urban road network. Past efforts to integrate ramp metering with variable speed limit control include the following: (Alessandri et al. 1998; Caligaris, Sacone, and Siri 2007) chose the optimal VSL and RM commands based on a second order model in an open-loop manner. (Andreas Hegyi, De Schutter, and Hellendoorn 2005) developed a combined VSL and RM controller by using model predictive control (MPC) based on the METANET model. (Papamichail et al. 2008) combined VSL and coordinated RM using an optimal control approach. (Lu, Qiu, et al. 2010) used a MPC approach to generate the VSL commands which coordinate with pre-existing RM controllers. (Lu et al. 2011) designed a MPC-based RM controller with a linearized first-order model which is equipped with a heuristic VSL controller.

The design of the coordinated VSL, RM and LC controller is based on the first-order cell transmission traffic flow model, which during the recent years was used to develop variable speed limit (VSL) control strategies. In (Hadiuzzaman and Qiu 2013), Hadiuzzaman et al. proposed a model predictive control (MPC)-based VSL control strategy to relieve congestion caused by active bottleneck which introduces capacity drop. No significant improvement was shown in bottleneck throughput. The reason given by the authors of (Hadiuzzaman and Qiu 2013) for the lack of improvements by the proposed VSL was that the model and data used were not accurate enough. In (Muralidharan and Horowitz 2015), an MPC-based coordinated VSL and ramp metering (RM) controller is proposed based on the link-node CTM. The VSL and RM control commands are computed by relaxing the receding-horizon optimization problem into linear programming. In (Csikós and Kulcsár 2017), the CTM model is expressed in a piecewise affine switching-mode form, based on which an MPC-based VSL controller is developed to attenuate shockwave.

In (Gomes et al. 2008), Gomes et al. performed a thorough analysis of the equilibrium points and their stability properties of the CTM model. However, the authors did not take the capacity drop phenomenon into consideration. In addition, the convergence rate at which the system states converge to the equilibrium points is not specified. Reference (Lovisari et al. 2014) developed sufficient conditions for the stability of the equilibrium points of CTM in terms of connectivity of a graph associated with the traffic network. The results of (Gomes et al. 2008) and (Lovisari et al. 2014) are established based on the monotonicity of CTM. However, if the CTM is modified to account for capacity drop and the fact that the discharging flow rate of a congested road section decreases with density (Y. Zhang and Ioannou 2017a; Srivastava, Jin, and Lebacque 2015; H.-Y. Jin and Jin 2015; Kontorinaki et al. 2016), then the CTM is no longer monotone. A finite horizon optimal routing and flow control strategy is proposed in (Lovisari et al. 2014). The stability and convergence of the closed-loop system to a desired equilibrium however has not been established. In (Coogan and Arcak 2016), the authors analyzed the equilibrium points and their stability properties under feasible and infeasible demand, however

the capacity drop phenomenon and traffic flow control is not considered. In (Karafyllis and Papageorgiou 2015), sufficient conditions for global asymptotic stability and global exponential stability of the equilibrium points of discrete-time CTM model are developed using vector Lyapunov functions. In (Kontorinaki, Karafyllis, and Papageorgiou 2017), the authors proposed a feedback control law that guarantees the global exponential stability of the desired equilibrium point of the CTM model. The control input in this case is the flow rate itself. It is not clear, however, how to implement the flow controller with VSL control.

The rest of this chapter is organized as follows: section Combined Variable Speed Limit and Lane Change Control presents the design and analysis of a combined variable speed limit (VSL) and lane change control. The controller is extended to be integrated with a ramp metering in the section Coordinated Variable Speed Limit, Ramp Metering, and Lane Change Controller. Section Comparison of Feedback Linearization and Model Predictive Strategies for VSL Control devotes to a comparison of feedback linearization and model predictive VSL controller. Section Stability Analysis and Variable Speed Limit Control of a Traffic Flow Model shows the stability analysis of the open-loop CTM and then presents the design of the VSL controller, which avoids the capacity drop and guarantees the global stability of the desired equilibrium point. Section Robust VSL Control of Cell Transmission Model with Disturbance deals with system disturbance, where a robust VSL controller is presented, and the last section Conclusion highlights the main results of this project and concludes it.

Combined Variable Speed Limit and Lane Change Control

Section based on the publication:

Y. Zhang and P.A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," IEEE Transactions on Intelligent Transportation Systems 18.7 (2016): 1812-1823.

Introduction

As introduced earlier, inconsistent performance of variable speed limit and ramp metering controllers have been reported in existing studies. Some researchers attribute the inconsistencies to the highly disordered and stochastic behavior at highway bottlenecks. One of the main factors of the disordered behavior at highway bottlenecks is the capacity drop phenomenon, where the maximum achievable traffic flow rate decreases when queues form (Banks 1991; Hall and Agyemang-Duah 1991). Under certain speed limit, when the density at the vicinity of the bottleneck increases to be higher than some critical value, a queue forms upstream of the bottleneck which decreases the capacity of the bottleneck. Capacity drop makes the dynamics of the traffic flow at bottleneck highly unstable, which is difficult for VSL control to maintain a high flow rate. (Rodrigo C. Carlson, Papamichail, and Papageorgiou 2013) claims that one of the main factors that introduce capacity drop is the inefficient acceleration of vehicles at the bottleneck, thus by providing an acceleration section with reasonable length and regulating the density with VSL, capacity drop can be avoided. Such an approach however has the following drawbacks. First it is difficult to establish in cases of incidents and second enforcing an acceleration section may require reducing the flow upstream considerably. The method in (Muralidharan and Horowitz 2015) is developed under the assumption that the bottleneck never returns to capacity drop mode from free flow mode, i.e., once the VSL controller recovers the bottleneck from capacity drop, the capacity drop never occurs again. While there is no reason to doubt the reported results, our studies and observations of traffic show clearly that forced lane changes in close proximity to the incident or bottleneck is the major cause of capacity drop and once it takes place VSL control will have limited or no effect in improving travel time. Most likely in the reported results which show significant benefits the scenarios did not involve significant forced lane changes or as in the case of (Rodrigo C. Carlson, Papamichail, and Papageorgiou 2013) it was prevented by creating an acceleration area before the bottleneck. It should be intuitively clear that once the forced lane changes bring down the speed of vehicles in neighboring lanes there is no way for an VSL control technique to eliminate the capacity drop.

In this section, we first proposed a lane change (LC) controller which can avoid or relieve the capacity drop at the bottleneck. Two types of VSL controller are designed to combine with the LC controller. The first one is an heuristic local feedback controller with integral action. The second one is a feedback linearization controller which is designed based on the first order cell transmission model. Together with a lane change controller, the feedback linearization VSL controller guarantees stability of the traffic flow and convergence of traffic densities to an equilibrium density with an exponential rate of convergence. In contrast to previous studies which relied on linearized models, our approach is based on feedback linearization and the results obtained are global. Therefore, from the macroscopic point of view the proposed VSL

and lane change control guarantees no capacity drop and maximum flow at the bottleneck. The lane change controller is based on a space model as in this case the control variable is the location of the lane change control commands. This location is found to depend on demand and number of lanes closed. The proposed combined lane change and VSL control design is evaluated using microscopic Monte Carlo simulations under different scenarios. The microscopic results generated are very consistent with the macroscopic ones and demonstrate consistent improvements to traffic mobility and impact on the environment for all the simulated scenarios.

System Modeling

Model of highway bottleneck

Consider a highway segment without on-ramps and off-ramps. A bottleneck is the point with lowest flow capacity. Due to the bottleneck a queue of vehicles forms as traffic demand increases. The flow rate of the bottleneck determines the throughput of the entire highway segment. Therefore, the modeling of the bottleneck traffic flow is crucial to the design of an efficient traffic control strategy. A bottleneck can be introduced by lane drop, incident lane blockage, merge point or other road conditions.

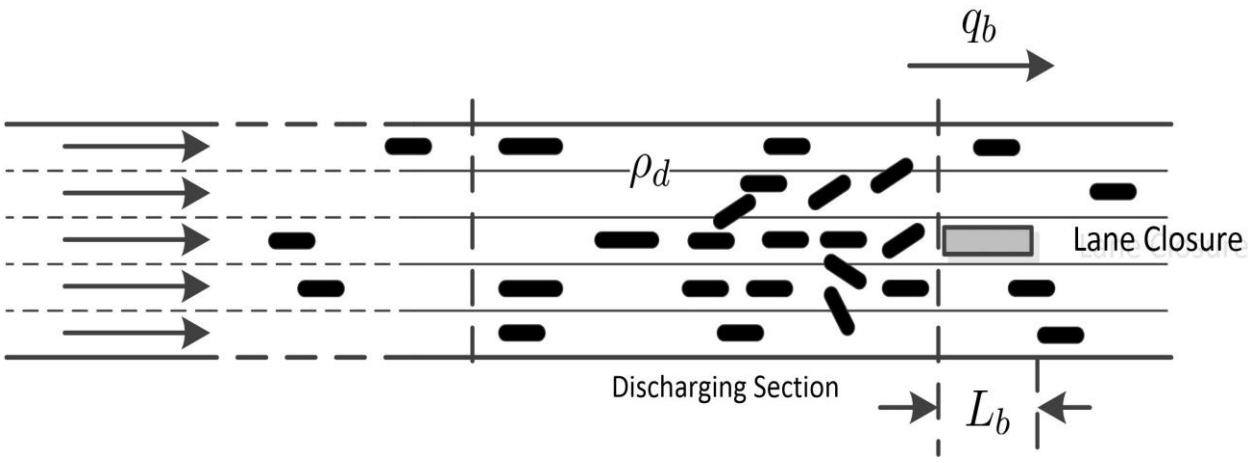


Figure 47. Highway Bottleneck

Figure 47 shows a highway segment with 5 lanes. A bottleneck is introduced by an incident which blocks one lane. The length of the bottleneck is denoted by L_b. We assume that the capacity of the highway segment before the incident is C. Then the ideal capacity of the bottleneck after the incident should be $C_b = \frac{4}{5}C$. As we can see in Figure 47, if L_b is small, the effect of the density within L_b is negligible and will not affect the bottleneck flow. The flow rate q_b at the bottleneck is determined by ρ_d, the vehicle density of the immediate upstream section of the bottleneck, which is referred to as the discharging section in Figure 47. We adopt the assumption of triangular fundamental diagram, that is, when the value of ρ_d is low, $q_b = v_f \rho_d$, where v_f is the free flow speed. However, when ρ_d is higher than some critical value $\rho_{d,c}$, i.e., the demand of the bottleneck is higher than its capacity C_b, a queue forms at the discharging section which propagates upstream. Forced lane changes performed by the vehicles

in the queue reduce the speed of flow in the open lanes. Therefore, the capacity would drop to $C'_{\mathrm{b}} = (1-\epsilon)C_{\mathrm{b}}$ once the queue forms (H.-Y. Jin and Jin 2015; Muralidharan and Horowitz 2015; Kontorinaki et al. 2016). The relationship between ρ_{d} and q_{b} is shown as solid line in Figure 48 and is described by equation (1).

$$q_{\mathrm{b}} = \begin{cases} v_{\mathrm{f}}\rho_{\mathrm{d}}, & \rho_{\mathrm{d}} \le \rho_{\mathrm{d,c}} \\ (1-\epsilon)C_{\mathrm{b}}, & \rho_{\mathrm{d}} > \rho_{\mathrm{d,c}} \end{cases} \quad (1)$$

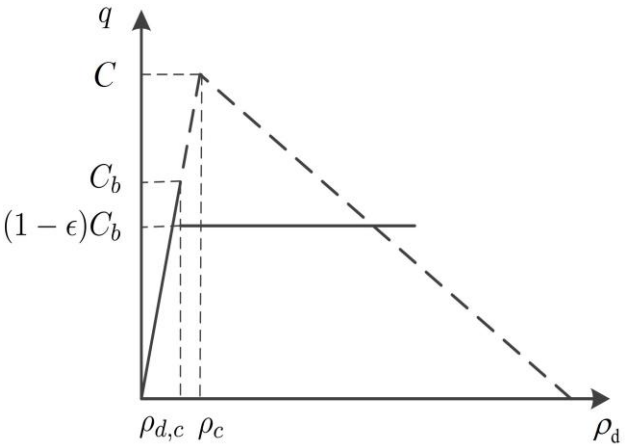


Figure 48. Fundamental Diagram

VSL configuration and cell transmission model

As shown in Figure 49, the upstream highway segment of bottleneck is divided into N sections. The lengths of different sections are expected to be similar but not necessarily identical. VSL signs are installed at the beginning of section 1 through section $N-1$. The speed limit in section N, which functions as the discharging section in Figure 47, is constant and equals v_{f}, the maximum possible speed given by the fundamental diagram, which would let vehicles in open lanes get through the bottleneck as fast as possible, under the assumption of triangular fundamental diagram.

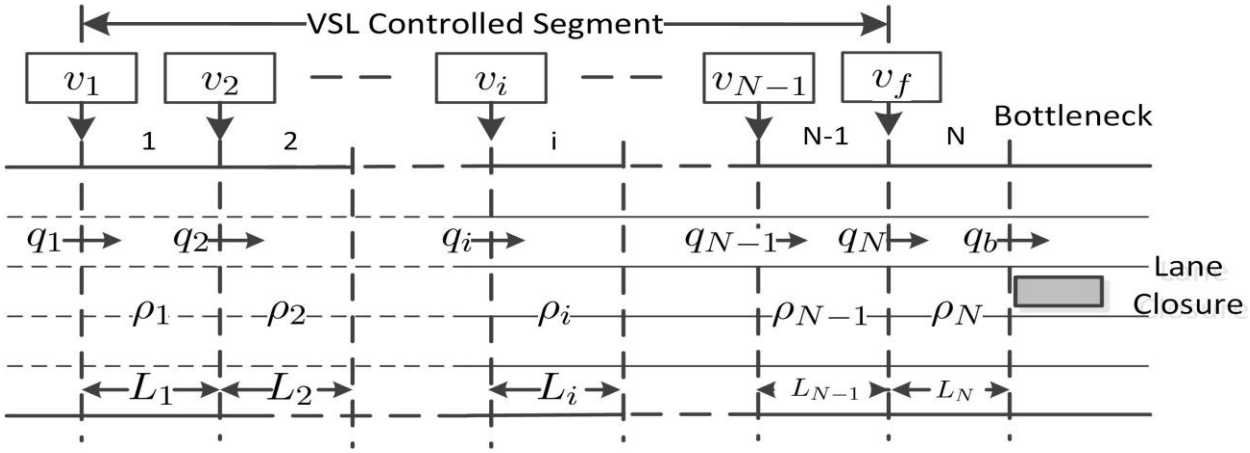


Figure 49. Configuration of VSL Control System

For $i = 1,2, \dots, N$, we denote the length, vehicle density and the inflow rate of section i with L_i, ρ_i and q_i respectively. For $i = 1, \dots, N-1$, we denote the variable speed limit in section i with v_i. The variables $\rho_i,\ q_i,\ v_i$ are all functions of time t. By conservation law, the dynamics of densities ρ_i are described by the differential equations

$$\begin{aligned} \dot{\rho}_i &= (q_i - q_{i+1})/L_i,\ i = 1,2, \dots, N-1 \\ \dot{\rho}_N &= (q_N - q_{\mathrm{b}})/L_N \end{aligned} \tag{2}$$

Under the assumption of triangular fundamental diagram, the flow rate q_i can be found as follows:

$$\begin{aligned} q_1 &= \min\{d, C_1, w_1(\rho_{\mathrm{j},1} - \rho_1)\} \\ q_i &= \min\{v_{i-1}\rho_{i-1}, C_i, w_i(\rho_{\mathrm{j},i} - \rho_i)\},\ i = 2,3, \dots, N \end{aligned} \tag{3}$$

where d is the demand flow of this highway segment assumed to be constant relative to the other variables. $\rho_{\mathrm{j},i}$ is the jam density of section i, at which q_i would be 0. w_i is the backward propagating wave speed in section i, C_i the capacity, i.e., the maximum possible flow rate in section i, given by $C_i = v_i w_i \rho_{\mathrm{j},i}/(v_i + w_i)$. We should note that for $i = N$, C_N and $\rho_{N,\mathrm{c}}$ are not the same as C_{b} and $\rho_{\mathrm{d,c}}$. When ρ_N reaches $\rho_{\mathrm{d,c}}$, q_{b} decreases but section N still has enough space for vehicles in section $N-1$ to flow in. Therefore, $\rho_{N,\mathrm{c}} > \rho_{\mathrm{d,c}}$, $C_N > C_{\mathrm{b}}$. The goal of the VSL controller is to stabilize the system described in (1)-(3) and maximize the flow rate q_{b}. According to (1), maximum q_{b} is obtained at $\rho_N = \rho_{\mathrm{b}}$, which is a discontinuity point of the fundamental diagram. From the macroscopic point of view, it is possible to find a VSL controller to maintain that $\rho_N = \rho_{\mathrm{d,c}}$ (H.-Y. Jin and Jin 2015). However, microscopic simulations in (Y. Zhang and Ioannou 2015) demonstrate that when congestion occurs at the bottleneck, the queue accumulates so fast that VSL control can hardly reduce the density back to $\rho_{\mathrm{d,c}}$, therefore it fails to maintain maximum flow. The reason is explained in the following subsection.

Effects of Lane Change Control

In order to study the effect of lane change control, we build a hypothetical highway segment as shown in Figure 47, which is straight, 8 km long and with 5 lanes, with the microscopic traffic flow simulated using the commercial software VISSIM. The VISSIM model is calibrated with typical freeway road geometry and driving behavior. The bottleneck is formed by an incident which blocks the middle lane. We investigate the relationship between the flow of the bottleneck q_{b} and the density ρ_{d} in the 500 m long discharging section immediately upstream the bottleneck under different levels of traffic demand. Figure 50 shows the relationship between q_{b} and ρ_{d} without any VSL control. The small blue circles describe the fundamental diagram in the case of lane change control. The red asterisks show the corresponding fundamental diagram in the absence of lane change control. The design procedure of LC controller is described in Design of the Lane Change ControllerDesign of the Lane Change Controller.

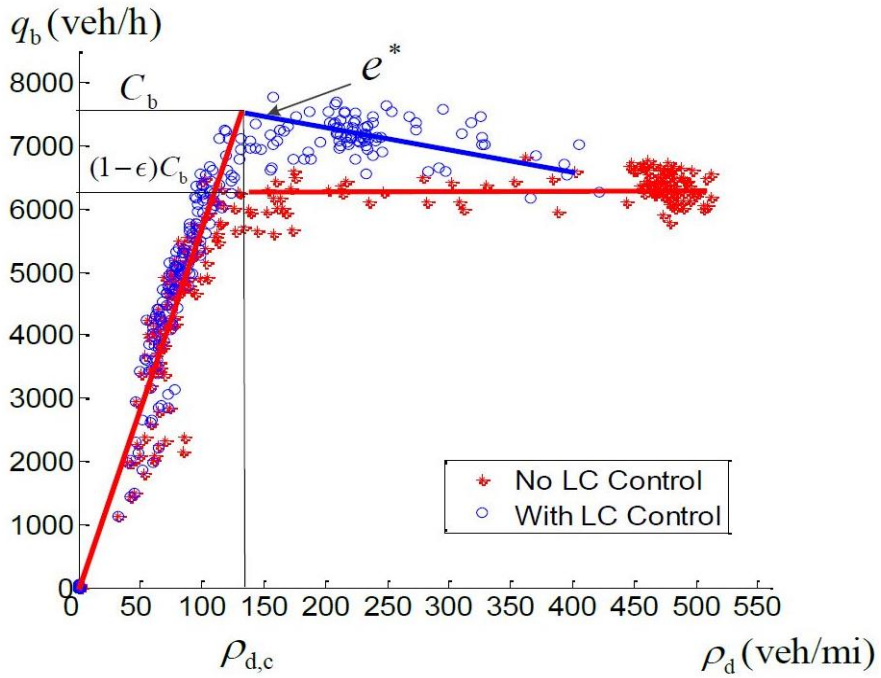


Figure 50. Fundamental Diagram with and without LC Control

Observing Figure 50 ,we can see that when LC control is applied, the capacity of the bottleneck is around 7600 veh/h, which is achieved at $\rho_d \approx 135$ veh/mi. However, when there is no LC control, q_b stops increasing even before ρ_d reaches 135 veh/mi (around ρ_d = 100 veh/mi). The highest flow rate is around 6300 veh/h. The reason why the flow rate in the no control case fails to reach higher level is demonstrated in Figure 47. When vehicles approach the incident spot without being aware that their lane is blocked they are forced to slow down considerably and change lanes. These forced lane changes at low speed cause the traffic to slow down in the open lanes before and after the incident leading to lower volume, while the average density of the discharging section, ρ_d, is still low. Other parts of the fundamental diagram in the no control case fit equation (1) very well. Compared to the fundamental diagram with LC control, we can calibrate the parameters as, $\rho_{d,c} = 135$ veh/mi, $C_b = 7600$ veh/h and $\epsilon = 0.16$. The above stated behavior of the bottleneck makes it difficult for VSL control to increase q_b at the bottleneck, as VSL is only able to regulate the average density ρ_d in the discharging section, but cannot eliminate the forced lane changes at the vicinity of the bottleneck.

On the other hand, with the LC control, we can see that

a) no obvious capacity drop is observed at $\rho_d = \rho_{d,c}$;
b) q_b at $\rho_d > \rho_{d,c}$ is approximately linear with a negative slope w_b, which represents the wave propagation rate;
c) most data points scatter close to $\rho_d = \rho_{d,c}$. The points of high density are rare.

These observations show that the LC controller is able to reduce the number of vehicle stops in the queue at bottleneck and decrease the vehicle density, which makes the system continuous

and easier for the VSL controller to stabilize. As a consequence of the LC control action, in the cell transmission model the relationship between ρ_N and q_b can be modeled as:

$$q_b = \begin{cases} v_f \rho_N, & \rho_N \le \rho_{d,c} \\ w_b(\rho_{j,d} - \rho_N), & \rho_N > \rho_{d,c} \end{cases} \quad (4)$$

where $\rho_{j,d} = v_f \rho_{d,c}/w_b + \rho_{d,c}$.

Although the lane change control is able to recover the triangular shape of the fundamental diagram, when the demand is higher than the capacity C_b, a congestion will still occur at the bottleneck. Now the goal is to design a VSL controller to stabilize (2)-(4) by homogenizing the densities in all sections and have them converge to an equilibrium which corresponds to the maximum possible flow as shown in the following section.

Design of the Lane Change Controller

The design of LC controller includes the pattern of the LC recommendation messages and the length of LC controlled segment. As we will explain below the control variable for LC control is the location of the LC recommendation which depends on a nonlinear spatial model that we developed.

Lane Change Recommendation Messages

Suppose a general highway segment has m lanes, with Lane 1 (Lane m) being the right (left) most lane in the direction of flow. We select the LC recommendation message R_i for lane $i, i = 1,2, \dots, m$ using the following rules:

a) For $1 \le i \le m$, if lane i is open, $R_i =$ "Straight Ahead";

b) For $i = 1 (i = m)$, if lane i is closed, $R_i =$ "Change to Left (Right)";

c) For $1 < i < m$, if lane i is closed, lane $i-1$ and lane $i+1$ are both open, $R_i =$ "Change to Either Side";

d) For $1 < i < m$, if lane i is closed, lane $i-1$ (lane $i+1$) is closed but lane $i+1$ (lane $i-1$) is open, $R_i =$ "Change to Left (Right)";

e) For $1 < i < m$, if lane i is closed, lane $i-1$ and lane $i+1$ are both closed, then we check R_{i-1} and R_{i+1}. If $R_{i-1} = R_{i+1}$, then $R_i = R_{i-1} = R_{i+1}$, else if $R_{i-1} \ne R_{i+1}$, $R_i =$ "Change to Either Side".

Rules (1)-(5) determine the LC recommendation messages depending on the incident location. The 5 rules covers all incident cases and are also mutually disjoint. Therefore they are well-defined and self-consistent.

Length of LC Control Segment

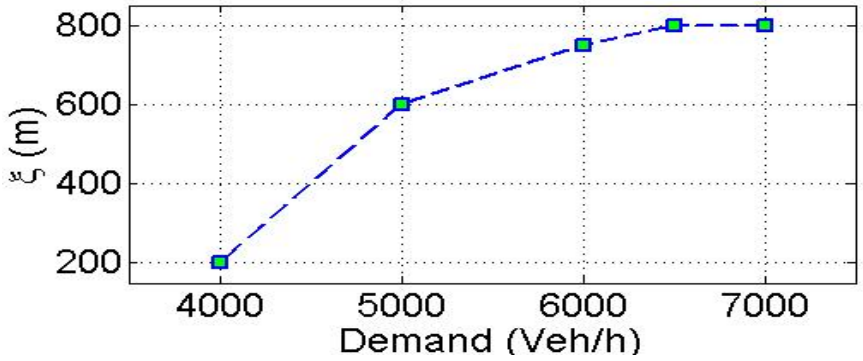


Figure 51. ξ under different traffic demands

The control variables in the LC control case are the length of the LC control segment and the location of the LC recommendation. Within that segment, a LC recommendation is given at each section within the segment. The length of the LC controlled segment need to be long enough in order to provide adequate space and time for upstream vehicles to change lanes. Intuitively, if more lanes are closed at the bottleneck, a longer LC control distance is required. In addition, the capacity of the bottleneck and demand will also affect the LC control distance. On the other hand if the length of LC control segment is too long it may cause other problems as the blocked lane will appear empty to drivers inviting more lane changes in and out of the blocked lane which is going to deteriorate performance in terms of unnecessary maneuvers. We used extensive microscopic simulation studies to develop the following empirical model that allows us to generate the control variable d_{LC} which is the length of the LC controlled section given by the following equation:

$$d_{\text{LC}} = \xi \cdot n, \tag{5}$$

where n is the number of lanes closed at the bottleneck, ξ a design parameter related to the capacity of bottleneck and the traffic demand which in our case is found to have the relationship shown in Figure 51. For a specific highway segment, the minimum value of ξ required under different traffic demands can be found by simulation. Since LC signs are only deployed at the beginning of sections, we choose the number of LC controlled sections M, as $M = \text{argmin} \left| \sum_{i=N-M+1}^{N} l_i - d_{\text{LC}} \right|$, where l_i represents the length of section i. More details can be found in [@zhang2015combined]. Here we assume that the LC controlled segment has no on-ramp or off-ramps. The model (5) is empirical and more spacial than temporal despite the dependence of ξ on demand which may be time varying. The purpose of the LC control is to ask drivers to start changing lanes before the incident. It is an off and on controller i.e change lanes or not required to change lanes. It is different than the VSL controller which is purely dynamic.

Heuristic Variable Speed Limit Controller

Virtual Ramp Metering Strategy

Given the LC controller is applied, VSL controller is expected to work together with LC in order to regulate the vehicle density at the critical value. Our first attempt to design a VSL controller is a non-model based heuristic one, which adopts the idea of ramp metering algorithm, ALINEA. ALINEA adjusts the on-ramp flow rate to keep downstream density at a desired level (Markos Papageorgiou, Hadj-Salem, and Middelham 1997). We generalize it to VSL control by regarding each highway section as the on-ramp of its downstream sections and regulating downstream density with VSLs. Unlike ramp metering, VSL cannot directly control the flow rate by stopping vehicles, therefore a multi-section structure as shown in Figure 49. Configuration of VSL Control System is applied to ensure control effect. The VSL controller in each section is expected to regulate the vehicle density of its downstream sections. The VSL control law is described as follows. Let $\eta_i(k) = \sum_{j=i}^{N} \rho_j\,(k) l_j / \sum_{j=i}^{N} l_i$ denote the average vehicle density of section i through section N at time step k, For each $1 \le i \le N-1$, the VSL command of Section i at time step k can be expressed as:

$$V_i(k) = V_i(k-1) + K_I\left[\rho_{d,c} - \eta_i(k)\right] \tag{6}$$

where $V_i(k)$ is the speed limit command of section i in control period k. K_I is the feedback gain, $\rho_{d,c}$ denote the critical density of the discharging section.
In equation (6), VSL commands respond to the difference to a fixed reference density, in order to suppress the shockwave and keep the density in discharging section.

Constraints on VSL commands

To ensure safety, we apply the following constraints to VSL commands in (6).

f) Finite Command Space. VSL commands would be hard to comply if take value from a continuous space. Hence, we round VSL commands $V_i(k)$ in (6) to multiples of 5 mi/h and apply lower/upper bounds to it. This makes the commands clear for drivers and adds dead-zone characteristics to the controller therefore avoid control chattering.

g) Saturation of Speed Limit Variations. It is dangerous to decrease the speed limit too fast in both time and space. The decrease should be within some threshold $C_v > 0$ between successive control periods and highway sections. We don't bound the speed limit variation if the speed limit increases. In this study, $C_v = 10$ mi/h(16km/h).

The above described constraints can be presented as follows:

$$V_i(k) - V_i(k+1) \le C_v,\ 1 \le i \le N-1 \tag{7}$$

$$V_i(k) - V_{i+1}(k) \le C_v,\ 1 \le i < N-1 \tag{8}$$

$$V_{\min} \le V_i(j) \le V_{\max},\ 1 \le i \le N-1 \tag{9}$$

Hence, the virtual mainline ramp metering VSL controller can be formulated as follows:

$$\bar{V}_i(k) = V_i(k-1) + \left[K_I\left(\rho_{d,c} - \eta_i(k)\right)\right]_5 \tag{10}$$

$$\tilde{V}_i(k) = \max\{\bar{V}_i(k), V_i(k-1) - C_v, V_{i-1}(k-1) - C_v\} \tag{11}$$

$$V_i(k) = \begin{cases} V_{\max}, & \text{if } \tilde{V}_i(k) > V_{\max} \\ V_{\min}, & \text{if } \tilde{V}_i(k) < V_{\min} \\ \tilde{V}_i(k), & \text{otherwise} \end{cases} \tag{12}$$

In (10), $[\cdot]_5$ is the operator which rounds a real number to its closest whole 5 number. In (12), $V_{\max}$ and $V_{\min}$ are the upper and lower bounds of VSL commands respectively.

Combination of VSL Control and LC Control

As described in Design of the Lane Change Controller and Heuristic Variable Speed Limit Controller, the LC controller is designed based on bottleneck layout and traffic demand. The VSL controller takes LC controlled segment as the discharging section and deploys VSL signs at upstream of it to keep desired density and smooth the traffic flow. The effect of LC controller helps the VSL controller to be more effective in generating the desired benefits. The block diagram of combined VSL & LC control system is shown in Figure 52.

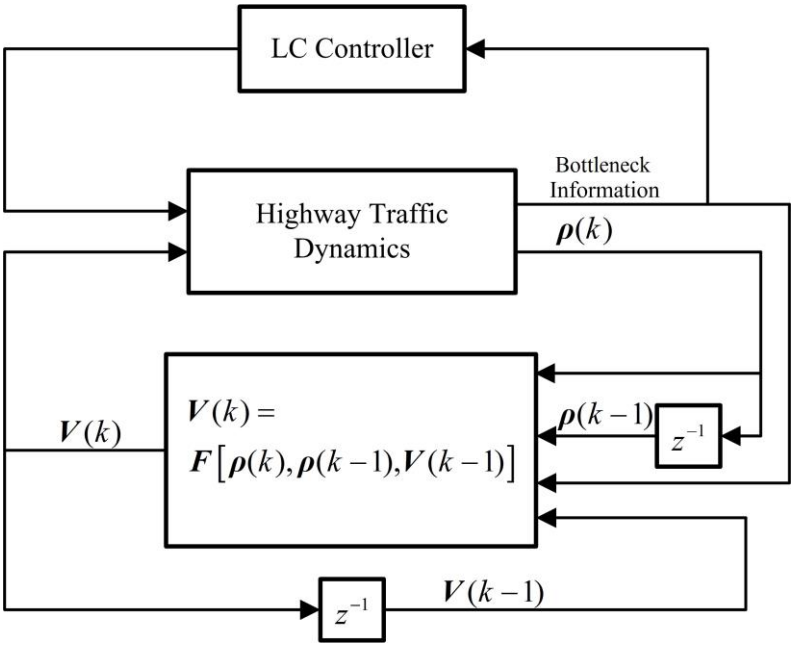


Figure 52. System Block Diagram

Feedback Linearization Variable Speed Limit Controller

In this section, we designed a feedback linearization VSL controller based on the cell transmission model (2)-(4).

Desired Equilibrium Point

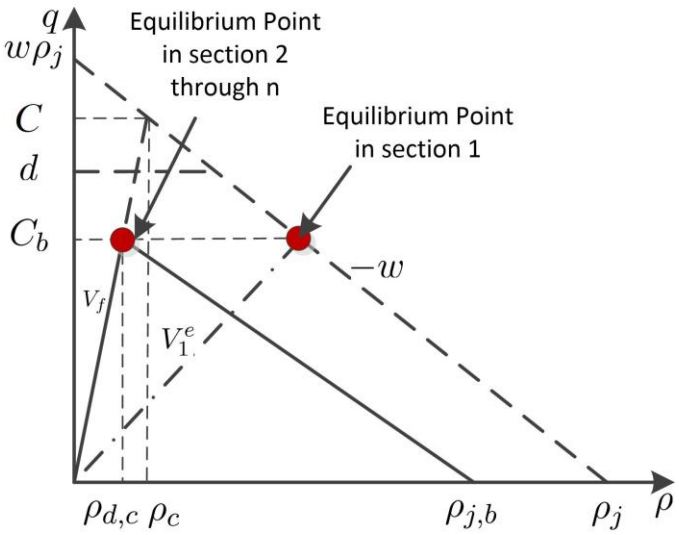


Figure 53. Desired Equilibrium Point

The fundamental diagram under LC control is shown in Figure 53. We consider the demand $d > C_{\mathrm{b}}$, which may introduce congestion at the bottleneck. From the nonlinear system (2)-(4), we calculate the equilibrium point by setting the derivatives in (2)-(4) to be zero. Let $\rho^{\mathrm{e}} = [\rho_1^{\mathrm{e}}, \rho_2^{\mathrm{e}}, \dots, \rho_N^{\mathrm{e}}]^T$ and $v^{\mathrm{e}} = [v_1^{\mathrm{e}}, v_2^{\mathrm{e}}, \dots, v_{N-1}^{\mathrm{e}}]^T$ denote the vector of equilibrium density and the corresponding equilibrium speed limits in each section respectively. The desired equilibrium point should be the one at which maximum possible flow rate C_{b} is achieved and the upstream traffic flow is homogenized. According to the triangular fundamental diagram (4), since the speed limit is constant and equals v_{f} in section N, therefore the optimum equilibrium density for maximum flow is $\rho_N^{\mathrm{e}} = C_{\mathrm{b}}/v_{\mathrm{f}}$. For section 2 through $N-1$, we set

$$\rho_2^{\mathrm{e}} = \cdots = \rho_N^{\mathrm{e}} = C_{\mathrm{b}}/v_{\mathrm{f}},\ v_2^{\mathrm{e}} = \cdots = v_{N-1}^{\mathrm{e}} = v_{\mathrm{f}}. \tag{13}$$

hence at the desired equilibrium point, the densities and speed limits in section 2 through N would be the same and the upstream traffic flow of the bottleneck is homogenized.

Since $d > C_{\mathrm{b}}$, we need to lower the speed limit in section 1 in order to suppress the traffic flow entering the controlled segment. According to (3), the equilibrium point satisfies:

$$v_1^{\mathrm{e}}\rho_1^{\mathrm{e}} = w_1\left(\rho_{\mathrm{j},1} - \rho_1^{\mathrm{e}}\right) = C_{\mathrm{b}}. \tag{14}$$

which gives

$$\rho_1^{\mathrm{e}} = \rho_{\mathrm{j},1} - C_{\mathrm{b}}/w_1,\ v_1^{\mathrm{e}} = C_{\mathrm{b}}w_1/\left(\rho_{\mathrm{j},1}w_1 - C_{\mathrm{b}}\right) \tag{15}$$

The equilibrium point described in (14)-(15) is the desired equilibrium point which maximizes the flow at the bottleneck and homogenizes the upstream traffic. In addition, it minimizes the average travel time according to the fundamental diagram. Without loss of generality, we assume the length of all sections are the same and equal to unit length. The system (2)-(4) can be expressed as follows:

$$\begin{aligned} \dot{\rho}_1 &= w_1\left(\rho_{\mathrm{j},1} - \rho_1\right) - v_1\rho_1 \\ \dot{\rho}_i &= v_{i-1}\rho_{i-1} - v_i\rho_i, \text{ for } i = 2, \dots, N-1 \\ \dot{\rho}_N &= \begin{cases} v_{N-1}\rho_{N-1} - v_{\mathrm{f}}\rho_N, & \rho_N \le \rho_{\mathrm{d,c}} \\ v_{N-1}\rho_{N-1} - w_{\mathrm{b}}\left(\rho_{\mathrm{j,b}} - \rho_N\right), & \rho_N > \rho_{\mathrm{d,c}} \end{cases} \end{aligned} \tag{16}$$

In (16), the only switching point is $\rho_N = \rho_{\mathrm{d,c}}$. This is consistent with real-world, since the capacities of upstream sections are much larger than C_{b}. As long as system (16) converges to the desired equilibrium point, the steady-state bottleneck flow is maximized and upstream traffic is homogenized.

Feedback Linearization VSL Controller

For the design and analysis of the VSL controller, we define the deviations of the state of (16) from the desired equilibrium (13)-(15) by defining the error system as: $e_i = \rho_i - \rho_i^{\mathrm{e}}$ for $i = 1,2,, \dots, N$ and $u_i = v_i - v_i^{\mathrm{e}}$ for $i = 1,2, \dots, N-1$. Substitute into (16), we have

$$\dot{e}_1 = -w_1 e_1 - v_1^e e_1 - u_1\rho_1 \tag{17}$$

$$\dot{e}_i = v_{i-1}^e e_{i-1} + u_{i-1}\rho_{i-1} - v_i^e e_i - u_i\rho_i$$

$$\text{for } i = 2, \dots, N-1$$

$$\dot{e}_N = \begin{cases} v_{N-1}^e e_{N-1} + u_{N-1}\rho_{N-1} - v_{\mathrm{f}} e_N, & e_N \le 0 \\ v_{N-1}^e e_{N-1} + u_{N-1}\rho_{N-1} + w_{\mathrm{b}} e_{N,} & e_N > 0 \end{cases}$$

The transformation of (16) to (17) shifts the non zero equilibrium state of (16) to the zero equilibrium point of (17). The nonlinear terms in (17) are $u_i\rho_i$ for $i = 1,2,\dots,N-1$. Now the problem is to select u_1 through u_{N-1} in order to stabilize system (17) and force all the errors or deviations from the equilibrium state to converge to zero.

We introduce the following feedback controller which 'kills' all nonlinearities and forces the closed loop system to be linear, an approach known as feedback linearization. We choose

$$\begin{aligned} u_i &= \left(-v_i^{\mathrm{e}} e_i - \lambda_i e_{i+1}\right)/\rho_i, \text{for } i = 1, \dots, N-2 \\ u_{N-1} &= \begin{cases} \dfrac{-\lambda_{N-1}e_N - v_{N-1}^{\mathrm{e}} e_{N-1} + v_{\mathrm{f}} e_N}{\rho_{N-1}}, & e_N \le 0 \\ \dfrac{-\lambda_{N-1}e_N - v_{N-1}^{\mathrm{e}} e_{N-1} - w_{\mathrm{b}} e_N}{\rho_{N-1}}, & e_N > 0 \end{cases} \end{aligned} \tag{18}$$

where $\lambda_i > 0$ for $i = 1, \dots, N-1$ are design parameters. This is a switching controller, whose switching logic is based on the value of e_N. Since we avoid the capacity drop by applying the LC control, the controller is continuous at the switching point. With the feedback linearization controller (18) the closed loop system becomes:

$$\begin{aligned} \dot{e}_1 &= -w_1 e_1 + \lambda_1 e_2 \\ \dot{e}_i &= -\lambda_{i-1} e_i + \lambda_i e_{i+1}, \text{for } i = 2 \dots, N-2 \\ \dot{e}_{N-1} &= \begin{cases} -\lambda_{N-2} e_{N-1} - \lambda_{N-1} e_N + v_{\mathrm{f}} e_N, & e_N \le 0 \\ -\lambda_{N-2} e_{N-1} - \lambda_{N-1} e_N - w_{\mathrm{b}} e_N, & e_N > 0 \end{cases} \\ \dot{e}_N &= -\lambda_{N-1} e_N \end{aligned} \tag{19}$$

The stability properties of the closed loop system (19) are described by the following Theorem.

Theorem 1.1. *The equilibrium point $e_i = 0$, $i = 1,2, \dots, N$ of the system (19) is isolated and exponentially stable. The rate of exponential convergence depends on the control design parameters λ_i, $i = 1,2, \dots, N-1$.*

Proof For $i = 1,2, \dots, N$, setting $\dot{e}_i = 0$ in (19), the only equilibrium point is $e_i = 0$. From (19), we can see that the state e_N is decoupled from other states, i.e., $\dot{e}_N = -\lambda_{N-1} e_N$., whose solution is

$$e_N(t) = e_N(0) \exp(-\lambda_{N-1} t), \ \forall t > 0. \tag{20}$$

Since $\exp(-\lambda_{N-1} t) > 0$ for all t, $e_N(t)$ and $e_N(0)$ have the same sign for all $t > 0$, i.e., if $e_N(0) \le 0$, then $e_N(t) \le 0$, if $e_N(0) > 0$, then $e_N(t) > 0$ for all $t > 0$. In other words e_N is either non increasing or non decreasing which means that the state e_N will not switch between $e_N \le 0$ and $e_N > 0$. Therefore, the dynamics of state e_{N-1} can be written as

$$\dot{e}_{N-1} = \begin{cases} -\lambda_{N-2} e_{N-1} - \lambda_{N-1} e_N + v_{\mathrm{f}} e_N, & e_N(0) \le 0 \\ -\lambda_{N-2} e_{N-1} - \lambda_{N-1} e_N - w_{\mathrm{b}} e_N, & e_N(0) > 0 \end{cases}$$

Let us define $e = [e_1, e_2, \dots, e_N]^T$, then the system (19) can be written in the compact form

$$\dot{e} = \begin{cases} A_1 e, & e_N(0) \le 0 \\ A_2 e, & e_N(0) > 0 \end{cases} \tag{21}$$

where

$$A_i = \begin{bmatrix} -w_1 & \lambda_1 & & & \\ & -\lambda_1 & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & & -\lambda_{N-2} & -\lambda_{N-1} + \beta_i \\ & & & & -\lambda_{N-1} \end{bmatrix}, \ i = 1,2$$

and $\beta_1 = -w_{\mathrm{b}}, \beta_2 = v_{\mathrm{f}}$. A_1 and A_2 are both upper triangular matrices with all diagonal entries being negative real numbers, i.e., A_1, A_2 are both Hurwitz. Hence, system (21) is exponentially stable. Therefore (19) is also exponentially stable. In addition, for a given sign of $e_N(0)$ there is no switching taking place in (21).

The rate of convergence to the equilibrium depends on the design parameters $\lambda_i, i = 1,2,\ldots. N-1$ which can be tuned to achieve a desirable convergence rate. It would also depend on the sign of the initial condition $e_N(0)$ as the dynamics that drive the error system depend on whether the initial condition $e_N(0)$ is negative or positive. Q.E.D.

The feedback linearization controller (18) is continuous in time. To apply it on real highway, we discretize the controller and apply the constraints described in Constraints on VSL commands.

Let $u_i(k)$ denotes u_i computed by equation (18) at $t = kT_c$. We have,

$$\bar{v}_i(k) = [v_i^e + u_i(k)]_5 \tag{22}$$

$$\tilde{v}_i(k) = \max\{\bar{v}_i(k), v_i(k-1) - C_v, v_{i-1}(k) - C_v\} \tag{23}$$

$$v_i(k) = \begin{cases} v_{\max}, & \text{if } \tilde{v}_i(k) > v_{\max} \\ v_{\min}, & \text{if } \tilde{v}_i(k) < v_{\min} \\ \tilde{v}_i(k), & \text{otherwise} \end{cases} \tag{24}$$

for $i = 1,2,\ldots,N-1, k = 0,1,2,\ldots.$

The above modifications will influence the ideal performance of the VSL controller described by Theorem 1. Such modifications are necessary in every control application (Y. Wang and Ioannou 2011b; Lu, Varaiya, et al. 2010; Rodrigo Castelan Carlson, Papamichail, and Papageorgiou 2011) and the way to deal with possible deterioration from the ideal performance is to use the design parameters $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}$ to tune the system using intuition and practical considerations. The selection of the feedback gains $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}$ has to consider the trade off between stability and robustness with respect to modeling errors.

Robustness with respect to varying demands

In the analysis above, we assume that the demand d is a constant and $d > C_b$. As explained below, the proposed VSL controller is robust with respect to different demands.

If $d < C_b$, vehicles in the controlled segment would discharge and the densities in each section would be lower than the desired density. The VSLs in each section would increase, but saturated at v_f. This situation is easy as due to the low demand congestion can be avoided or managed very well.

When $d > C_b$ and keeps increasing, according to Theorem 1.1, the controller lowers the speed limit in section 1 and limits the number of vehicles that enter the downstream network. Therefore, a queue would be created whose size will be increasing upstream the flow. It appears, at first glance, as if we are moving congestion from the sections under VSL and LC control to upstream sections. The important question we need to answer is how many vehicles there are in this queue and how fast it grows with and without VSL and LC control in the sections under consideration.

In order to analyze the queue size upstream of section 1, we modify the system (2)-(4) by introducing a new state Q, which represents the number of vehicles in the queue upstream section 1. We assume that $Q = 0$ at steady state flow before the incident. Using the flow conservation equation, we have

$$\dot{Q} = d - q_1 \tag{25}$$

where d is the traffic demand. The inflow rate of section 1, q_1 then becomes

$$q_1 = \begin{cases} \min\{d, C_1, w_1(\rho_{\mathrm{j},1} - \rho_1)\}, & Q \le 0 \\ \min\{C_1, w_1(\rho_{\mathrm{j},1} - \rho_1)\}, & Q > 0 \end{cases} \tag{26}$$

Equation (26) assumes that as long as the queue upstream section 1 is not fully discharged, the inflow rate of section 1 will be as high as the maximum flow rate that section 1 can receive under current ρ_1. Note that the introduction of Q does not make any difference to system (2)-(4) before and during the incident. It only tracks the growth and discharge of the queue upstream section 1. Therefore the stability of the closed-loop system (19) is not affected.

Hence, with the combined VSL and LC controller, the queue size is measured with Q. In the no control case, a queue forms at section N, whose size is denoted by $\hat{Q}$. The following Lemma holds.

Lemma 1.1. *If the demand $d > C_{\mathrm{b}}$, $\hat{Q}$ grows faster than Q at steady state. In particular,*

$$\dot{Q} - \dot{\hat{Q}} = -\epsilon C_{\mathrm{b}} < 0 \tag{27}$$

Proof Similar to Equation (25), we can estimate $\hat{Q}$ with the following equation $\dot{\hat{Q}} = d - \hat{q}_{\mathrm{b}}$, where $\dot{\hat{Q}}$ is the growth rate of $\hat{Q}$, $\hat{q}_{\mathrm{b}}$ is the outflow rate of section N without control. Since $d > C_{\mathrm{b}}$, q_1 converges according to Theorem 1.1 to the desired flow rate C_{b} exponentially with the combined VSL and LC controller. $\hat{q}_{\mathrm{b}}$ would decrease to $\hat{q}_{\mathrm{b}} = (1 - \epsilon)C_{\mathrm{b}}$ due to capacity drop. Substituting the steady state values of q_1 and $\hat{q}_{\mathrm{b}}$ in the above equations we obtain (27), i.e., at steady state, the growth rate of Q is less than that of $\hat{Q}$. Q.E.D.

From the analysis above, it is clear that if the demand d increases from below the bottleneck capacity C_{b} to greater than C_{b} and keeps increasing, the combined VSL and LC controller is able to protect the bottleneck from getting congested by suppressing the speed limit in section 1 therefore ρ_N can be stabilized at the desired value. On the other hand, in the no control case, the bottleneck is directly exposed to the excessive demand, therefore ρ_N increases and leads to capacity drop. Figure 54 plots the steady state bottleneck flow q_{b} with respect to demand d. When $d < C_{\mathrm{b}}$, the bottleneck would not be congested. When $d > C_{\mathrm{b}}$, the bottleneck flow would be stabilized at the maximum value C_{b} by the combined controller in the controlled case. In the no control case, the flow rate would decrease to $(1 - \epsilon)C_{\mathrm{b}}$ due to capacity drop.

Therefore, the combined VSL and LC controller is robust with respect to different levels of traffic demand. The queue of vehicles grows slower in the controlled case than in the case with no control.

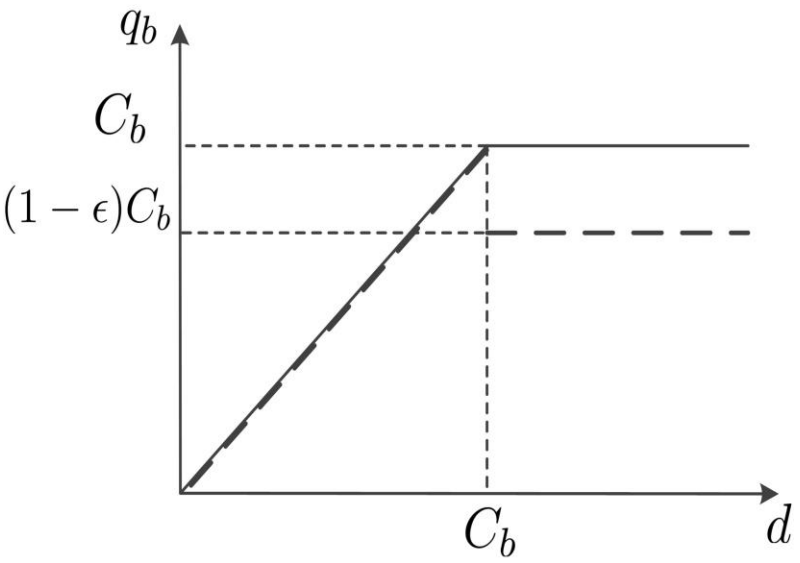


Figure 54. Steady State q_"b" under Different Demands — With Control, - - -Without Control

Numerical Results

Simulation Network

We evaluate the combined VSL & LC control method using a microscopic and macroscopic model of the traffic flow on a 10 mile (16 km)-long southbound segment of I-710 freeway in California, United States (between I-105 junction and Long Beach Port), which has a static speed limit of 65 mi/h (105 km/h). We build this freeway network in VISSIM and calibrate the microscopic model using historical data provided by (Transportation 2015). The car following and lane change behavior of the VISSIM model is calibrated and validated using real measurements under static speed limit of 65 mi/h.

Figure 55. Simulation Network

The studied highway segment has 3-5 lanes at different locations. As shown in Figure 55, we assume the bottleneck is introduced by an incident which blocked one lane. The upstream segment of the bottleneck is divided to 10 500m-600m sections. The bars across the highway in Figure 55 are where VSL signs and LC signs deployed. In VISSIM, incidents are simulated by placing stopped bus in certain lane.

Evaluation of the Heuristic VSL Controller

This section demonstrates the evaluation results of the heuristic VSL controller.

Monte Carlo Simulation and Scenarios

In this case, we load the network with a demand of 6500 veh/h, 30% of which are trucks. This proportion is much higher than that in reality. We use this setup to test the performance on extremely high truck volume. To verify that the proposed control method generates consistent results under different traffic conditions, we set up 3 different scenarios on the highway network to perform a general evaluation of the proposed method and take 10 sets of Monte Carlo simulation for each scenario. The final performance measurements are averages of the Monte Carlo simulation results. In the simulation, all lanes are open at the beginning of simulation. 20 min after simulation begins, certain lane is closed near the incident spot in Figure 55 and the controller is activated. The simulation terminates when 2000 vehicles pass through the bottleneck. We fix the total number of vehicles that passed through the bottleneck in each simulation, so that the measurements are comparable. Other configuration of scenarios are listed in Table 9.

Table 9. Simulation Scenarios

Scenario No.	Total No. of Lanes	Bottleneck Pattern
1	3	Lane 2 Closed
2	3	Lane 3 Closed
3	4	Lane 3 Closed

Performance Measurements

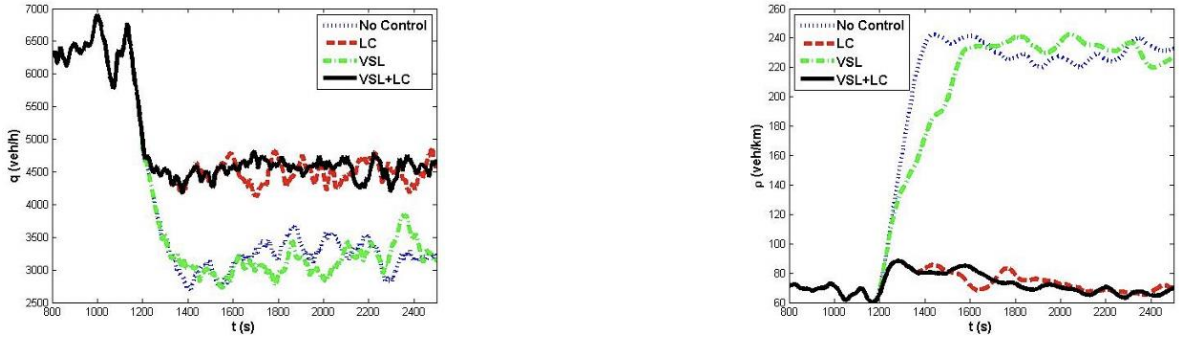
We introduce the following measurements to evaluate the performance of the proposed control method. To be precise, all measures start from the time instant of lane closing and terminate with the simulation.

Control effects on traffic mobility are evaluated by total travel time (TTT) of all vehicles that passed through the highway network (in hours). Let $t_{i,\text{in}}$ and $t_{i,\text{out}}$ denote the time instant vehicle i enters and exits the network respectively. TTT is given by $TTT = \sum_{i=1}^{2000}\left(t_{i,\text{out}} - t_{i,\text{in}}\right)$. Control effects on traffic safety are evaluated by total number of stops $s_{\text{tot}} = \sum_{i=1}^{2000} s_i$ and total number of lane changes $c_{\text{tot}} = \sum_{i=1}^{2000} c_i$,

where s_i, c_i are number of stops and lane changes performed by vehicle i respectively. For environmental impact, we measure fuel consumption rate $fr = \sum_{i=1}^{2000} f_i / (2000 \cdot \sum_{i=1}^{2000} d_i)$ where f_i, d_i are the fuel consumption and distance traveled in the network by vehicle i respectively. The definition of CO_2 emission rate E_{CO_2} and NO_x emission rate E_{NO_x} are similar to fr.

Controller Parameters

In our simulation, the default speed limit when VSL controller is not active is $V_f = 65$mi/h. VSL decrease threshold $C_v = 10$mi/h (16km/h). Bounds of VSL $V_{\min} = 30$mi/h (48km/h), $V_{\max} = 65$mi/h. Feedback gain $K_I = 2$.



(a) Flow Rate **(b) Density**

Figure 56. Traffic Condition in Discharging Section

Simulation Results

In scenario 1-3, we compare the simulation results under the following control modes: 1) No control; 2) LC control only; 3) VSL control only; 4) Combined VSL & LC control.

Figure 56 show the density of discharging section and the bottleneck flow rate during the simulation in scenario 1. After the incident happens at 1200s, the density of discharging section increase dramatically to 250 veh/km and the bottleneck flow rate drops by 50% if LC control is not applied. When VSL control is applied alone, the density of discharging section increases slower however cannot be kept at a lower level. When LC control is applied, the bottleneck flow rate only deceased by about 30%. Since we lose 1 lane out of 3, the flow rate per lane has no drop. LC control ensures a high discharging rate of the bottleneck therefore avoids the congestion. Comparing the flow rate and density curve with and without VSL control, system oscillation is damped by VSL, therefore traffic safety improved. Fuel consumption and emissions also tend to be reduced, which is shown below.

The effects of different control mode on performance measurements defined in Evaluation of the Heuristic VSL Controller are shown in Table 10–Table 12. We can observe that the combined control method provides significant improvement on each measurement, which is also consistent with respect to different scenarios. The combined VSL & LC control strategy

reduces TTT by 26%-32%, s_{tot} by about 90%, c_{tot} by 3%-14%, fr and E_{CO_2} by 16%-24%, E_{NO_x} by 16%-21%.

To study the roles of VSL control and LC control in the combined control strategy respectively, we also analyze the case VSL control and LC control are applied to the traffic system alone. LC control considerably decreases travel time and number of stops, but cannot reduce number of lane changes, since it only spreads forced lane changes along the LC controlled sections, instead of avoiding them. On the other hand, VSL control homogenizes the density and speed in each section. Drivers are not tend to change lane if densities and speeds are similar in all lanes, therefore VSL control reduces number of lane changes in VSL controlled sections. This is very important for traffic safety in truck-dominant highways. Trucks not only take long time and large space to change lane, their large size also blocks the eye sight of drivers, which makes lane change much more dangerous than usual.

Table 10. Performance Measurements of Scenario 1

Performance Measurement	Cars							Trucks						
	No Control	LC	Percentage Changed	VSL	Percentage Changed	VSL+LC	Percentage Changed	No Control	LC	Percentage Changed	VSL	Percentage Changed	VSL+LC	Percentage Changed
Travel Time (min)	29561	20486	-31%	29780	1%	**20574**	**-30%**	9539	6925	-27%	9447	-1%	**7047**	**-26%**
No. of Stops	27503	3007	-89%	25721	-6%	**3099**	**-89%**	6757	719	-89%	6344	-6%	**783**	**-88%**
No. of LC	12344	12089	-2%	11134	-10%	**10630**	**-14%**	1245	1314	6%	1094	-12%	**1142**	**-8%**
Fuel (g/mi/veh)	141.46	120.76	-15%	130.78	-8%	**109.64**	**-22%**	599.24	582.77	-3%	520.60	-13%	**505.71**	**-16%**
CO2 (g/mi/veh)	422.40	354.76	-16%	394.44	-7%	**325.56**	**-23%**	1917.86	1864.23	-3%	1665.80	-13%	**1617.36**	**-16%**
NOx (g/mi/veh)	0.49	0.47	-4%	0.42	-15%	**0.39**	**-20%**	22.10	20.38	-8%	20.03	-9%	**18.65**	**-16%**

Table 11. Performance Measurements of Scenario 2

Performance Measurement	Cars							Trucks						
	No Control	LC	Percentage Changed	VSL	Percentage Changed	VSL+LC	Percentage Changed	No Control	LC	Percentage Changed	VSL	Percentage Changed	VSL+LC	Percentage Changed
Travel Time (min)	29076	19914	-32%	28403	-2%	**19854**	**-32%**	9273	6862	-26%	9280	0%	**6842**	**-26%**
No. of Stops	23889	2541	-89%	22464	-6%	**2321**	**-90%**	7206	573	-92%	6665	-7%	**535**	**-93%**
No. of LC	12404	12944	4%	11254	-9%	**11585**	**-7%**	1354	1543	14%	1233	-9%	**1373**	**1%**
Fuel (g/mi/veh)	141.60	120.67	-15%	128.50	-9%	**109.67**	**-23%**	599.35	582.86	-3%	516.70	-14%	**502.49**	**-16%**
CO2 (g/mi/veh)	421.32	353.24	-16%	386.03	-8%	**323.71**	**-23%**	1918.83	1864.56	-3%	1653.81	-14%	**1607.09**	**-16%**
NOx (g/mi/veh)	0.50	0.48	-5%	0.42	-17%	**0.41**	**-19%**	22.22	20.37	-8%	19.94	-10%	**18.56**	**-16%**

Table 12. Performance Measurements of Scenario 3

Performance Measurement	Cars							Trucks						
	No Control	LC	Percentage Changed	VSL	Percentage Changed	VSL+LC	Percentage Changed	No Control	LC	Percentage Changed	VSL	Percentage Changed	VSL+LC	Percentage Changed
Travel Time (min)	30033	20378	-32%	30033	0%	**20426**	**-32%**	9524	6938	-27%	9650	1%	**6914**	**-27%**
No. of Stops	27544	2797	-90%	25763	-6%	**2681**	**-90%**	6729	695	-90%	6568	-2%	**650**	**-90%**
No. of LC	12475	12380	-1%	11295	-9%	**11084**	**-11%**	1276	1331	4%	1152	-10%	**1162**	**-9%**
Fuel (g/mi/veh)	143.37	120.71	-16%	132.38	-8%	**110.05**	**-23%**	601.58	583.50	-3%	523.66	-13%	**506.20**	**-16%**
CO2 (g/mi/veh)	427.21	354.32	-17%	398.29	-7%	**326.01**	**-24%**	1925.31	1866.57	-3%	1675.56	-13%	**1618.96**	**-16%**
NOx (g/mi/veh)	0.51	0.47	-6%	0.43	-15%	**0.40**	**-21%**	22.16	20.40	-8%	20.12	-9%	**18.67**	**-16%**

The environmental evaluation is interesting. VSL and LC control has different performance on different measurements and vehicle types. For trucks, fr and E_{CO_2} are highly sensitive to accelerations. Large portion of fuel consumption and CO_2 emission are produced by speeding up and down in shock waves. Therefore, although LC control reduced the travel time of trucks by 26%-27%, fr and E_{CO_2} of trucks are only reduced by 3%. On the other hand, VSL control suppresses the shockwave and smooth the speed of all vehicles, which reduce fr and E_{CO_2} of trucks by 13%-14%.

For cars, fr and E_{CO_2} are not as sensitive to accelerations as those of trucks. Engine efficiency, which increases with speed, is also a major factor. LC Control significantly increases the average speed and engine efficiency of cars, therefore decreasefr and E_{CO_2} of cars by 15%-17%. In the meantime, VSL control also reduce fr and E_{CO_2} of cars by 7%-9%.

NO_x is major toxic road traffic emission. Since we assume cars are all gasoline-based, the NO_x emission of cars is very small comparing to that of trucks. Both VSL control and LC control have contributions on reduction of NO_x.

From the simulation results and analysis above, combined VSL & LC control method can improve the bottleneck flow rate, smooth and homogenize the traffic flow simultaneously, hence is able to provide significant and consistent improvement on traffic mobility, safety and environmental impact in truck-dominant highway networks.

Evaluation of the Feedback Linearization VSL Controller

In this section, we design and evaluate a combined VSL and LC controller for the simulation of a real world highway segment. We use both macroscopic and microscopic traffic flow models and carry out Monte Carlo simulations for different incident scenarios in order to evaluate consistency with respect to performance improvements.

Simulation Network and Scenarios

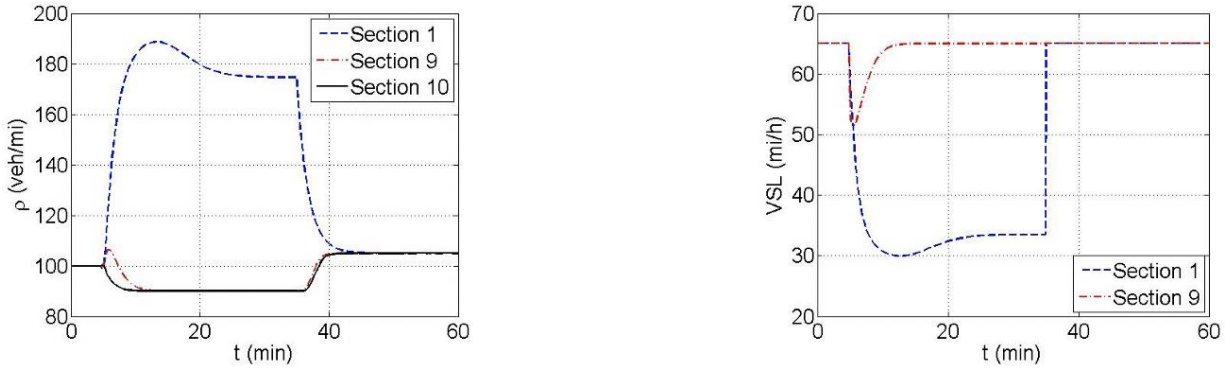
We use the same network in Figure 55 to evaluate the performance of the feedback linearization VSL controller. To demonstrate the performance, robustness and consistency of the proposed controller under different incident conditions, we consider 3 different scenarios with different incident durations. We simulate each scenario under different demand flows. In each scenario, the incident occurs 5 minutes after simulation begins and lasts for **30 min** in scenario 1, which simulates the case of an incident of moderate duration which may be due to an accident; for **10 min** in scenario 2 which simulates the case of a short incident due to a vehicle breakdown or minor accident. The incident is **not removed** after occurrence in scenario 3, which simulates a long time lane closure or a construction site or a physical bottleneck. We evaluate the combined VSL and LC control performance for each scenario with constant demand flows of 6000 veh/h and 6500 veh/h which is higher than the capacity of the bottleneck. 5% of the demand are trucks.

Macroscopic Simulation

In this section, we use a macroscopic model to evaluate the performance of the proposed VSL controller. Since the macroscopic model used does not take into account lane changes and their effect close to the incident, we apply the LC controller to the corresponding microscopic model and use the microscopic model data to validate the macroscopic cell transmission model. The desired equilibrium point of the I-710 highway segment is calculated to be

$$\rho_1^{\mathrm{e}} = 174.6 \text{ veh/mi},\ \rho_2^{\mathrm{e}} = \rho_3^{\mathrm{e}} = \cdots = \rho_{10}^{\mathrm{e}} = 90 \text{ veh/mi}$$
$$v_1^{\mathrm{e}} = 33.5 \text{ mi/h},\ v_2^{\mathrm{e}} = v_3^{\mathrm{e}} = \cdots = v_9^{\mathrm{e}} = 65 \text{ mi/h}$$

The LC recommendation sign is deployed at the beginning of section 9 and section 10 in Figure 55, and recommends vehicles to change lanes by moving to the open lanes on either side. For the VSL controller, the following parameters are used: $C_{\mathrm{v}} = 10$ mi/h, $v_{\max} = 65$ mi/h, $v_{\min} = 10$ mi/h, $T_{\mathrm{c}} = 30$ s. We choose $\lambda_1 = \lambda_2 = \cdots = \lambda_9 = 20$. We should note that as mentioned in Simulation Network, the capacity of the bottleneck with incident is 4500 veh/h. However, in the macroscopic model, we are assuming a strict triangular fundamental diagram and the capacity C_{b} is calibrated to be $v_{\mathrm{f}} \times \rho_{10}^{\mathrm{e}}$ = 5850 veh/h. The reasons for this difference are explained in the following section. Since the logic of our VSL controller is to stabilize the density at the critical value, the accurate value of equilibrium density is more important than the value of flow rate. The densities and variable speed limits for the case of scenario 1 with demand $d = 6500$ veh/h are plotted in Figure 57. For clarity of presentation, we only plot the densities in section 1, 9 and 10 and VSL commands in section 1 and 9.

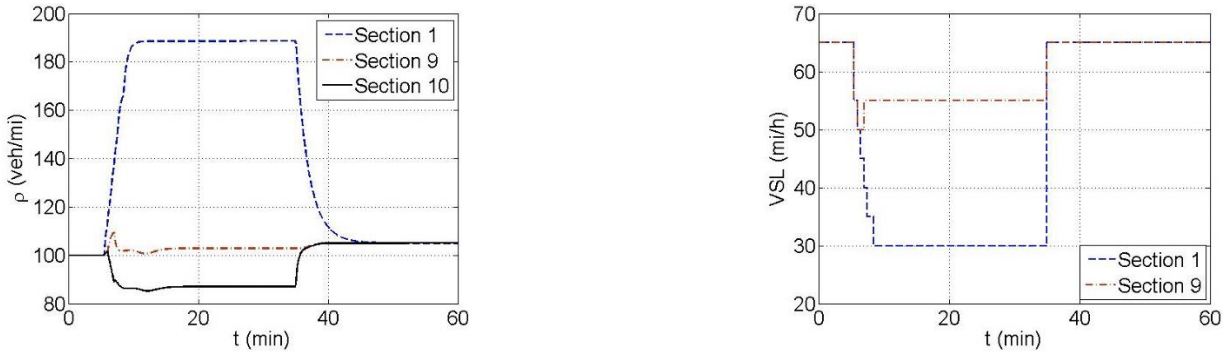


(a) Vehicle Density **(b) VSL Command**

Figure 57. Controller Performance without Constraints

Figure 57 demonstrates what is predicted by theory. That is the density in section 1 converges to the desired density of 174.6 veh/mi and the densities in sections 9, 10 to the desired density of 90 veh/mi till the incident is removed at $t = 35$ min, in which case the densities converge to 105 veh/h, which is higher than the pre-incident value. This is because the queue formed at section 1 during the incident needs to discharge, therefore the temporary demand of the bottleneck after the incident is higher than the demand of the overall network.

We then apply the constraints (22)-(24) to the VSL controller. The densities and VSL commands with constraints are shown in Figure 58. Figure (a) in Figure 58 demonstrates that the density in the discharging section converges to $\rho_{10} = 85$ veh/mi, which is lower than $\rho_{10}^{e} = 90$ veh/mi. According to the fundamental diagram in Figure 53, the steady state flow would be a bit lower than the desired flow rate. However, the difference is negligible. The VSL command in section 1 converges to $v_1 = 30$ mi/h and the VSL command in section 9 converges to $v_9 = 55$ mi/h, which are not exactly the same as the desired values due to the application of the constraints.



(a) Vehicle Densities **(b) VSL Commands**

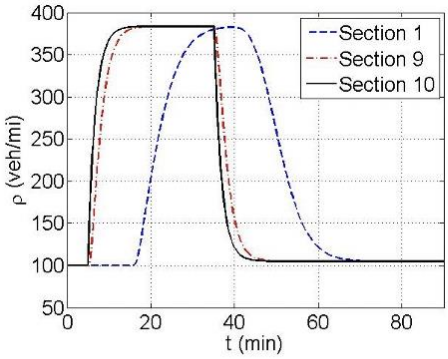
Figure 58. Controller Performance with Constraints

In Figure 57, ρ_9 and ρ_{10} converge to the corresponding equilibrium point in less than 10 min while ρ_1 converges to ρ_1^{e} much slower (in about 20 min). The reason of this phenomenon is the different values of ρ_1^{e} and ρ_9^{e}. As discussed in [@carlson2011local], a low value of speed limit would suppress the capacity of the section. After the incident occurs, v_1 decreases to a low value and ρ_1 increases rapidly, since because of the outflow of section 1, q_2 is suppressed by v_1. Then the process of adjusting ρ_1 from the overshoot to ρ_1^{e} takes long time due to the low level of q_2.

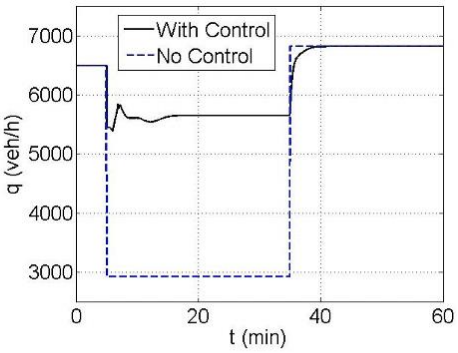
On the other hand, from Figure 58, we can see that with the constrained VSL, ρ_1 converges fast and no overshoot is observed. This is because v_1 is constrained (22)-(24) thus fails to adjust ρ_1 back to ρ_1^{e} after overshooting, however, as stated before, the difference is negligible. Similarly, in Figure (b) in Figure 58 , the VSL command v_1 converges to 30 mi/h in less than 10 min and stays at that value. Since the VSL commands only take whole 5 mi/h values due to (22), small variation of v_1 in the continuous case are all rounded up. Therefore, in the constrained case, there are no variations of v_1 around 30 mi/h.

Figure 59 (a) demonstrates how vehicle densities evolve in scenario 1 without any control. The density increases dramatically in the discharging section to 370 veh/h and propagates upstream. Even after the incident is removed at $t = 35$ min, the shockwave continues propagating backwards and takes longer time to discharge. Figure 59 (b) shows the flow rate at the bottleneck with and without control. During the incident, the flow rate decreases to less than 3000 veh/h due to capacity drop in the case of no control, while the bottleneck flow converges to 5600 veh/h with the combined VSL and LC controller. Again, the flow rate under

control is higher than the real capacity of the bottleneck due to the assumption of triangular fundamental diagram.



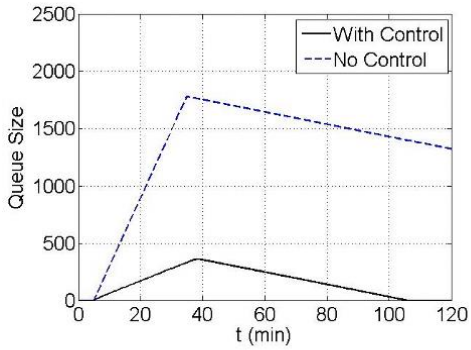
(a) Vehicle Densities without Control



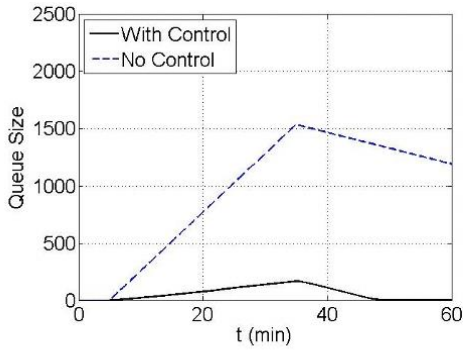
(b) Bottleneck Flow with and without Control

Figure 59. System Behavior without Control

We use scenario 1 to examine the growth of the queue at the entrance to the controlled network. The numbers of vehicles in the queues are plotted in Figure 60 with respect to the time t. When the demand $d = 6500$ veh/h, the maximum number of vehicles in the queue is 1700 in the case of no control, while the number is less than 500 in the control case, which demonstrates that the combined VSL and LC controller reduces the queue size significantly. The queues grow slower and discharge faster with lower demand, as less vehicles arrive at the tail of the queue.



(a)



(b)

Figure 60. Growth and Discharge of the Queue

Microscopic Simulation

In this section, we use a microscopic traffic model that is closer to the real environment in order to confirm the improvements predicted by theory and demonstrated by the macroscopic model. In addition, the microscopic model allows us to evaluate additional performance criteria such as number of stops and lane changes that affect safety as well as the environmental impact of VSL and LC controllers. We simulate the I-710 traffic flow network shown in Figure 55

for the above mentioned 3 traffic scenarios. The simulated demand consists of 85% light duty passenger vehicles and 15% trucks. This ratio represents the highest truck ratio at peak hours on I-710, therefore shows the worst traffic condition (Transportation 2015). To show consistency of the results, we conducted 10 sets of Monte-Carlo simulations with different random seeds for each scenario. The curves in Figure 61 are generated from a single simulation. The evaluation results in Table 13–Table 15 are the average of 10 simulations.

Consistency between microscopic and macroscopic models

Figure 61 shows the density and flow rate of the discharging section in both microscopic and macroscopic simulations. We can see that the density curve in macroscopic and microscopic simulations match each other. The microscopic flow rates in the no control cases are very similar and consistent with those in macroscopic simulations. However, when the combined VSL and LC controller is applied, the flow rates in microscopic simulations are lower than those in macroscopic simulations, which means that the flow speed in the discharging section in microscopic simulations is lower than what we get from the macroscopic model.

The deviation in speed is due to the following factors:

1. *Modeling error.* In the macroscopic model, we use a simplified triangular fundamental diagram to model the discharging section, which implies that the flow speed at the desired density is v_{f}. However, the actual speed would be lower than v_{f}. Especially when the LC controller is applied, drivers are usually conservative when merging to the open lanes.
2. *Speed limit following delay*. In the macroscopic model, we assume that the flow speed follows the speed limit exactly with no delay. However, in the microscopic model, the traffic flow needs time and space to accelerate to the desired speed limit. When vehicles change lanes, they do not adjust to new speeds instantaneously.
3. *Friction effect*. The friction effect reflects the empirically observed drivers' fear of moving fast in the open lanes when an incident or slowly moving vehicles exist in neighboring lanes [@wright2015new]. In microscopic simulation, this phenomenon is captured and has an effect when compared with the macroscopic simulations.

Figure 62 demonstrates the relationship between ρ_{10} and q_b at the equilibrium state under the combined VSL and LC controller in microscopic simulations. In Figure 62, the negative slope part, i.e., the congested part of the fundamental diagram is not observed even when the demand d is higher than the capacity, since the controller protects the bottleneck from getting congested. For different levels of demand, the data points concentrate in different clusters which shows that the controller homogenizes the traffic flow. Furthermore, when $d \leq 3000$ veh/h, the data points stay close to the line with the slope $v_{\mathrm{f}} = 65$ mi/h. When d keeps increasing, the data points move to the right side of the line due to the factors we explained above.

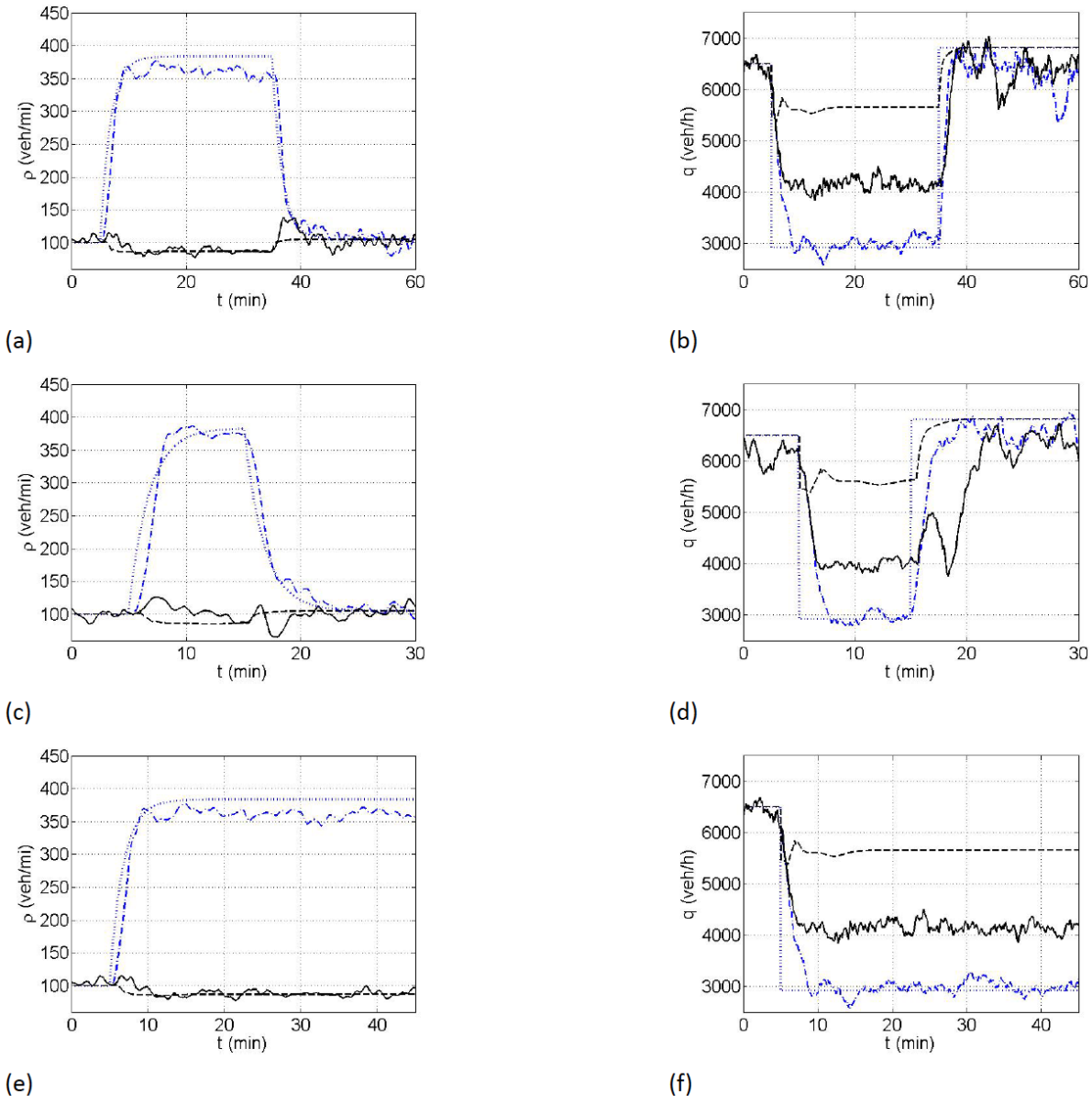


Figure 61. Comparison of Macroscopic and Microscopic Models

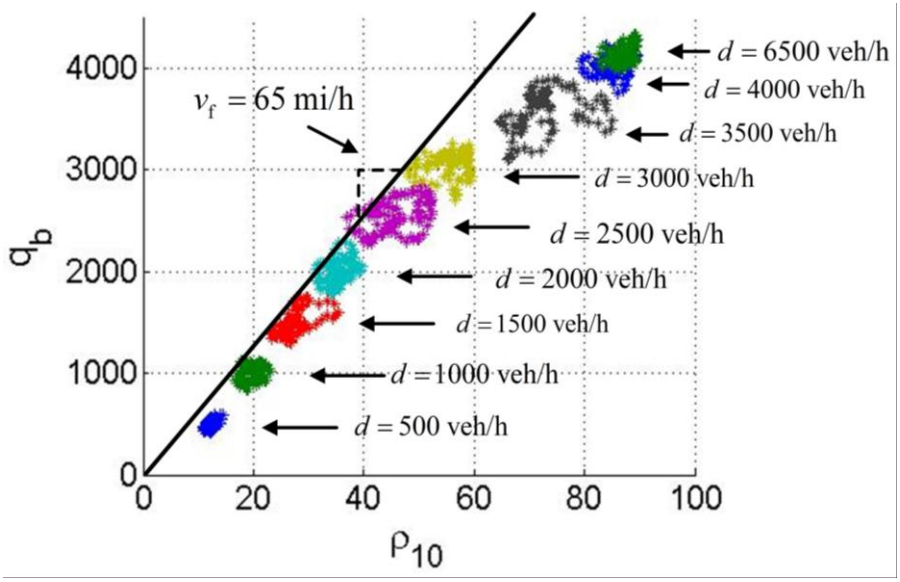


Figure 62. Fundamental Diagram with Combined Controller

Performance Measurement and Criteria

We use the following measurements to evaluate the performance of the proposed controller. To be precise, in scenario 1 and 2, the measurements start at the time instant that the incident begins ($t = 5$ min) and terminate at the time instant 10 minutes after the incident ends ($t = 45$ min in scenario 1 and $t = 25$ min in scenario 2), so that the traffic states can achieve steady state. In scenario 3, where the incident is not removed, the measurements start at the time instant that the incident begins ($t = 5$ min) and terminate at $t = 45$ min. In each scenario, we collect the data of all vehicles that pass through the bottleneck during the above defined measuring periods and calculate the following values: (a)Average travel time $\bar{T}_t$. (b)Average number of stops $\bar{s}$. (c)Average number of lane changes $\bar{c}$. (d)Average fuel consumption rate. (e)Average CO2 emission rate. (f)Average NOx emission rate. (g)Average PM25 emission rate. Control effects on traffic mobility are evaluated using the average travel time. Let N_v denote the number of vehicles pass through the bottleneck during the measuring period. Average travel time $\bar{T}_t$ is defined as

$$\bar{T}_t = \sum_{i=1}^{N_v} \left(t_{i,\text{out}} - t_{i,\text{in}}\right) / N_v$$

where $t_{i,\text{in}}$ and $t_{i,\text{out}}$ denote the time instant vehicle i enters and exits the network respectively. Note that our simulation network has enough space upstream of the controlled segment, therefore the time waiting in the queue is also counted.

Control effects on traffic safety are evaluated by the average number of stops and average number of lane changes. Less stops and lane changes indicate smoother traffic flow and lower probability of crash [@ioannou2012dynamic]. $\bar{s}$ and $\bar{c}$ are defined as

$$\bar{s} = \sum_{i=1}^{N_v} s_i / N_v, \ \bar{c} = \sum_{i=1}^{N_v} c_i / N_v$$

where s_i, c_i are number of stops and lane changes performed by vehicle i respectively. For environmental impact, we measure the average fuel consumption rate and the average emission rates of CO2, NOx, and PM25. These rates are uniformly defined as:

$$R = \sum_{i=1}^{N_v} E_i / \sum_{i=1}^{N_v} d_i$$

where E_i denotes the fuel consumed or a certain type of emission generated by vehicle i in the highway network, d_i represents the distance traveled by vehicle i in the network, and R denotes the fuel consumption rate or the tailpipe emission rate of CO2, NOx, or PM25. The fuel consumption rate and emission rates are calculated using the MOVES model of the Environment Protection Agency (EPA) based on the speed and acceleration profile of each vehicle [@epa2010motor].

Evaluation Results

Table 13, Table 14 and Table 15 demonstrate the results of microscopic evaluation of all 3 scenarios under different traffic demands. From the results, we can see that the combined VSL & LC controller is able to provide significant improvements in traffic mobility, safety and environment. For traffic mobility, the proposed controller reduces the average travel time of each vehicle by 6.25% - 22.13%.

For traffic safety, the combined VSL and LC controller dramatically decreases the average number of stops by 83% - 88.75% in different scenarios, therefore drastically reduces the instances of the stop-and-go traffic, smooths the traffic flow and damps the shockwave. Average number of lane changes is also decreased by 5.6% - 10.48%. The combined VSL and LC controller homogenizes the density and speed in each section. Drivers tend to not change lane if densities and speeds are similar in all lanes, therefore the VSL control reduces the number of lane changes in the network under consideration. This is highly important for traffic safety in highway segments with high truck ratio. Trucks not only take long time and large space to change lane, their large size also blocks the eye sight of drivers, which makes lane changes of trucks much more dangerous than other vehicles.

The proposed controller reduces the fuel consumption rate and tailpipe emission rate from two perspectives. First, it reduces the travel time of vehicles, therefore decreases the emission levels of vehicles waiting in the queue. Second, it smooths the traffic flow and suppresses the acceleration and deceleration, therefore decreases the emission in these transient states. In the simulation, fuel consumption rate is decreased by 4.26% - 8.82%. The improvement in CO_2 emission rate is approximately proportional to the improvement of fuel consumption rate,

since CO_2 is the main product of fuel burnt. The proposed controller reduces NO_x emission rate by about 3.54% - 6.71%. The emission rate of PM25 is also decreased by 3.74% - 7.73%. Therefore, the combined VSL and LC controller is able to bring environmental benefits.

The question how much of these improvements is due to VSL and LC controller alone is also answered using these simulation studies. From Table 13–Table 15, we can see that when the LC controller is applied alone, all evaluation criteria improve except for the average number of lane changes. The improvements on T_t and $\bar{s}$ are significant, while other criteria are only improved slightly. As discussed in Effects of Lane Change Control, the LC controller is able to recommend upstream vehicles to make lane changes before stopping at the queue and avoid the capacity drop therefore reduce the average travel time and average number of stops. Improvements on environmental criteria are results of improvements of traffic mobility. However, for the average number of lane changes, the LC controller only makes the lane changes take place in advance, instead of avoiding them, thus fails to reduce $\bar{c}$. Furthermore, when the VSL controller is applied alone, only the average number of stops is reduced. Other criteria are not improved and in some cases are even deteriorated by the VSL controller. This is because the VSL controller (18) is designed based on the assumption that the capacity drop has already been removed by the LC controller. When the LC controller is absent, VSL is not able to improve the bottleneck flow and reduce the vehicle density. But when the VSL controller is applied together with the LC controller, all criteria are further improved since the VSL stabilizes the vehicle densities at the desired equilibrium point and homogenizes the traffic flow. When the traffic flow is homogenized in each section and lane, the drivers do not tend to change lanes frequently, hence the average numbers of lane changes are also reduced. Comparing the three scenarios, the improvement on each measurement criteria in scenario 2 appears to be less significant than the other 2 scenarios. The reason is that the incident duration in scenario 2 is very short.

Table 13. Evaluation Results of Scenario 1

Demand	6000 veh/h					6500 veh/h				
Control	No Control	LC Only	VSL Only	Control	Improvement	No Control	LC Only	VSL Only	Control	Improvement
T_t	18.85	17.12	18.95	16.85	-10.59%	20.72	17.67	21.21	16.83	-18.76%
$\bar{s}$	11.16	2.45	3.61	1.90	-83.00%	12.10	2.55	3.78	1.91	-84.21%
$\bar{c}$	4.00	4.75	4.74	3.78	-5.60%	4.67	5.54	5.88	4.31	-7.71%
NOx	1.56	1.49	1.61	1.49	-4.43%	1.64	1.58	1.60	1.53	-6.71%
CO2	558.56	543.22	577.59	536.01	-4.04%	589.46	556.47	605.59	537.21	-8.86%
Energy	178.65	173.67	184.76	171.40	-4.06%	186.78	177.93	193.73	170.31	-8.82%
PM25	0.049	0.048	0.047	0.050	0.66%	0.054	0.054	0.053	0.050	-7.73%

Table 14. Evaluation Results of Scenario 2

Demand	6000 veh/h					6500 veh/h				
Control	No Control	LC Only	VSL Only	Control	Improvement	No Control	LC Only	VSL Only	Control	Improvement
T_t	12.41	11.87	13.46	11.63	-6.25%	13.58	12.62	15.02	12.42	-8.54%
$\bar{s}$	5.16	0.75	2.16	0.65	-87.37%	5.72	1.58	2.33	0.91	-84.09%
$\bar{c}$	3.68	3.80	3.90	3.52	-4.31%	4.27	4.81	5.01	3.91	-8.33%
NOx	1.42	1.41	1.44	1.39	-2.48%	1.48	1.49	1.51	1.42	-4.05%
CO2	483.37	479.17	497.81	470.16	-2.73%	508.13	504.16	524.36	487.18	-4.12%
Energy	154.53	151.65	159.18	150.36	-2.70%	161.04	161.15	167.66	154.18	-4.26%
PM25	0.041	0.041	0.041	0.041	-0.77%	0.046	0.047	0.047	0.045	-2.17%

Table 15. Evaluation Results of Scenario 3

Demand	6000 veh/h					6500 veh/h				
Control	No Control	LC Only	VSL Only	Control	Improvement	No Control	LC Only	VSL Only	Control	Improvement
T_t	19.84	17.25	18.16	16.69	-15.89%	21.25	16.75	20.45	16.55	-22.13%
$\bar{s}$	15.46	2.13	4.00	1.74	-88.75%	16.12	2.54	3.72	1.83	-88.65%
$\bar{c}$	4.61	4.55	5.11	4.21	-8.60%	4.58	5.36	6.36	4.10	-10.48%
NOx	1.58	1.51	1.58	1.50	-4.95%	1.58	1.55	1.66	1.50	-4.95%
CO2	570.72	538.41	564.54	529.76	-7.18%	568.96	550.32	597.94	523.25	-8.04%
Energy	182.55	172.17	180.58	169.39	-7.21%	182.85	175.99	191.26	168.11	-8.06%
PM25	0.052	0.047	0.047	0.050	-3.74%	0.052	0.053	0.053	0.050	-3.74%

Coordinated Variable Speed Limit, Ramp Metering, and Lane Change Controller

Section based on the publication:

Y. Zhang and P.A. Ioannou, "Coordinated variable speed limit, ramp metering and lane change control of highway traffic," IFAC-PapersOnLine 50.1 (2017): 5307-5312.

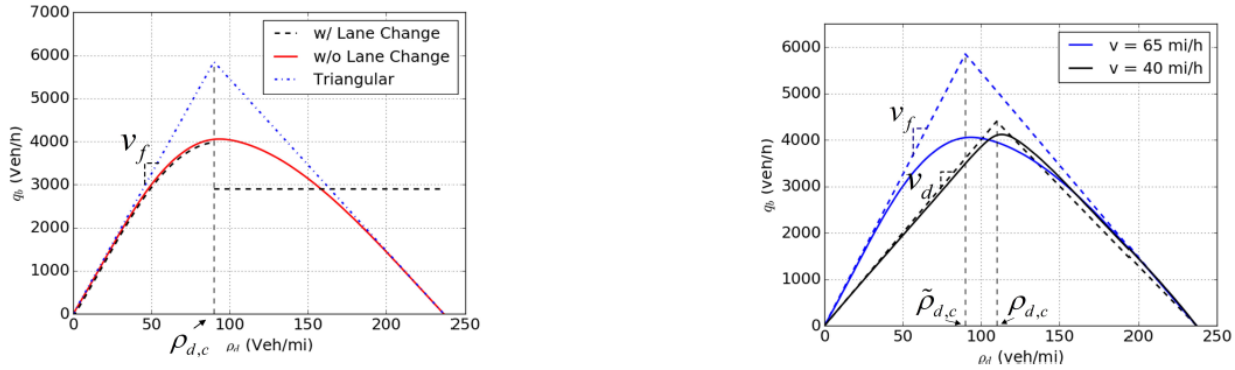
Introduction

The coordination of RM and VSL considers network mobility, on-ramp queues and fairness between the mainline and the ramps. The objective is to keep a balanced delay time between vehicles on the mainline and the ramps and avoid queues on the ramps from spilling back to the urban roads. In this section, we use an analytical method to design a coordinated VSL and RM controller based on a cell transmission macroscopic model with triangular fundamental diagram which together with a lane change controller guarantees stability of the traffic flow and convergence of traffic density to the desired equilibrium point exponentially fast. Considering the fact that RM controllers have been widely deployed in the United States, we assume that the RM control command is determined before the VSL and design the VSL controller to coordinate with the RM and stabilize the traffic flow. The coordinated VSL and RM controller with lane change is evaluated using Monte Carlo microscopic simulations and shows significant improvement in traffic mobility, safety and the environment impact.

System Modeling

Effect of VSL on the Fundamental Diagram

Consider the highway bottleneck shown in Figure 55. A bottleneck is introduced by an incident that blocks one lane. The speed limit upstream the bottleneck is the free flow speed $v_f = 65$ mi/h. As discussed in Effects of Lane Change Control, the lane change controller can avoid the capacity drop. However, as shown in Figure 63 (a), in the fundamental diagram with lane change control, the low ρ_d part is very close to its triangular approximation, which means that the flow speed is close to v_f, while the flow speed decreases as ρ_d approaches $\rho_{d,c}$. In last section we attribute the reduction of speed to modeling error, delay of speed limit following and driver's caution when passing the incident site. This deviation of speed will not harm the benefit of VSL with respect to traffic mobility when designing the VSL controller based on the triangular fundamental diagram as long as ρ_d is stabilized at $\rho_{d,c}$. However, if the speed limit upstream the bottleneck is v_f, vehicles need to decelerate when approaching the bottleneck, which leads to shock waves that propagate upstream.



(a) w/ and w/o LC

(b) w/ and w/o VSL

Figure 63. Effects of LC and VSL on Fundamental Diagrams

If we decrease the speed limit upstream the bottleneck to v_d, such that $0 < v_d < v_f$, according to (Markos Papageorgiou, Kosmatopoulos, and Papamichail 2008), the critical density in the fundamental diagram will be shifted to higher value and the slope of the under-critical part of the fundamental diagram will be decreased and made closer to a straight line. Our microscopic simulations confirm this statement. The black solid line in Figure 63 (b) shows the fundamental diagram under a speed limit of 40 mi/h. Compared to the one under 65 mi/h, which is shown as the blue solid line in Figure 63 (b), the capacity of the bottleneck is not decreased despite under a lower speed limit as the critical density is increased from $\tilde{\rho}_{d,c}$ to $\rho_{d,c}$. As we can see in the figure, this fundamental diagram is very close to its triangular approximation, that is, the speed deviation at $\rho_{d,c}$ is very small. If we design the coordinated VSL and RM controller based on this fundamental diagram and let the VSL command converge to v_d at the equilibrium state, the shockwave upstream the bottleneck will be attenuated. We demonstrate this with microscopic simulations in Numerical Simulations. To conclude, under speed limit of v_d, the highway bottleneck can be modeled with high accuracy as equation (4).

Cell Transmission Model with Ramp Flows

The highway segment to be controlled by the coordinated VSL and RM controller is shown in Figure 64. The bottleneck is introduced by a lane closure. The highway segment upstream the bottleneck is divided into $N + 1$ sections, which are indexed as section 0 through section N. For $i = 0,1,\dots,N$, ρ_i, q_i, r_i, s_i represent the vehicle density, mainline in-flow rate, on-ramp flow rate and off-ramp flow rate in section i respectively, where ρ_i, s_i are measurable, r_i are determined by the RM controller, therefore also measurable. For $i = 0,1,\dots,N-1$, v_i denote the variable speed limit in section i. In section N, the speed limit is a constant denoted by v_d. q_b denotes the flow rate through the bottleneck. Let $R_i = r_i - s_i$ be the net ramp flow and L_i the length of section i, for $i = 0,1,\dots,N$. According to the flow conservation law, we have

$$\begin{aligned} \dot{\rho_i} &= \frac{1}{L_i}(q_i - q_{i+1} + R_i), \quad \text{for } i = 0,1,\dots,N-1 \\ \dot{\rho_N} &= \frac{1}{L_N}(q_N - q_b + R_N) \end{aligned} \tag{28}$$

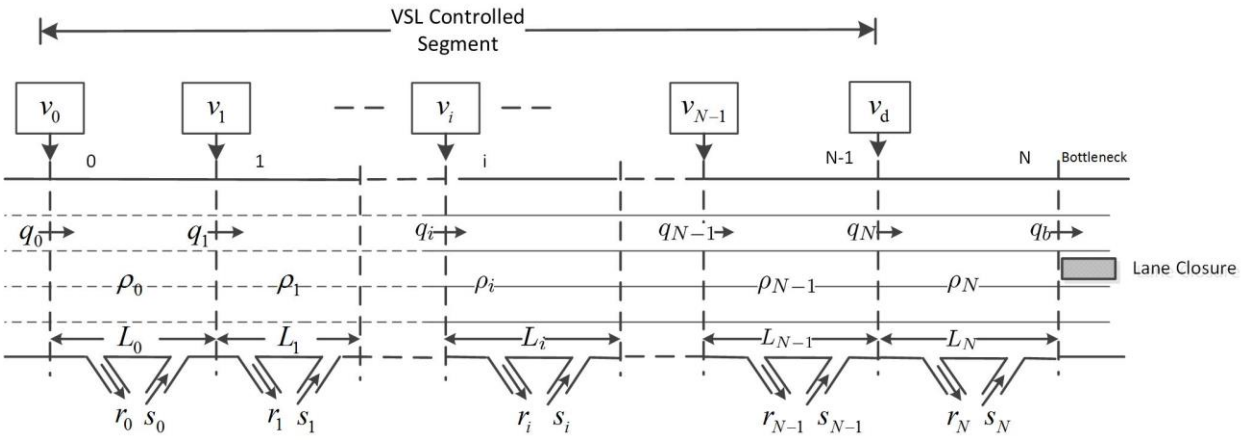


Figure 64. Configuration of the Highway Segment

The flow rate and bottleneck model is the same as (3) and (4). For the sake of completeness, we write the equations here.

$$\begin{aligned} q_0 &= \min\{d, C_0, w_0(\rho_{\mathrm{j},0} - \rho_0)\} \\ q_i &= \min\{v_{i-1}\rho_{i-1}, C_i, w_i(\rho_{\mathrm{j},i} - \rho_i)\},\ i = 1, \dots, N \end{aligned} \tag{29}$$

$$q_{\mathrm{b}} = \begin{cases} v_{\mathrm{d}}\rho_N, & \rho_N \le \rho_{\mathrm{d,c}} \\ w_{\mathrm{b}}(\rho_{\mathrm{j,d}} - \rho_N), & \rho_N > \rho_{\mathrm{d,c}} \end{cases} \tag{30}$$

Controller Design

In this section, the coordinated VSL and RM controller is designed. We first design the VSL controller by assuming that the RM control command is given. Then we choose the ramp metering strategy, ALINEA/Q, to manage the ramp flows and the queue lengths on ramps.

Design of VSL

The goals of designing the VSL controller include: (1) Given any type of RM controller, the VSL controller should be able to coordinate with it and stabilize the density ρ_N in the discharging section at the critical value $\rho_{d,c}$, in order to keep q_b at the highest level. (2) Homogenize the traffic flow upstream the bottleneck in order to improve the traffic safety and bring environmental benefits. Consider the subsystem which includes section 1 through section N. Define the error states

$$e_i = \rho_i - \rho_{d,c}, \text{for } i = 1,2,\dots,N$$

We have

$$\begin{aligned} \dot{e}_i &= \frac{1}{L_i}(v_{i-1}\rho_{i-1} - v_i\rho_i + R_i), \text{ for } i = 1,2,\dots,N-1 \\ \dot{e}_N &= \begin{cases} \dfrac{v_{N-1}\rho_{N-1} - v_d\rho_N + R_N}{L_N}, & \rho_N \le 0 \\ \dfrac{v_{N-1}\rho_{N-1} - w_b(\rho_{j,b} - \rho_N) + R_N}{L_N}, & \rho_N > 0 \end{cases} \end{aligned} \tag{31}$$

Let

$$\begin{aligned} v_i &= \frac{-\lambda_i L_{i+1} e_{i+1} + v_d\rho_{d,c} - \sum_{j=i+1}^{N} R_j}{\rho_i}, \text{ for } i = 0,1,\dots,N-2 \\ v_{N-1} &= \begin{cases} \dfrac{-\lambda_{N-1} L_N e_N + v_d\rho_N - R_N}{\rho_N - 1}, & \rho_N \le \rho_{d,c} \\ \dfrac{-\lambda_{N-1} L_N e_N + w_b(\rho_{j,b} - \rho_N) - R_N}{\rho_N - 1}, & \rho_N \le \rho_{d,c} \end{cases} \end{aligned} \tag{32}$$

Substitute the controller (32) into the open-loop system (31), we have the following closed-loop system:

$$\begin{aligned} \dot{e}_i &= -\lambda_{i-1} e_i + \frac{L_{i+1}}{L_i}\lambda_i e_{i+1}, \text{ for } i = 1,2,\dots,N-2 \\ \dot{e}_{N-1} &= \begin{cases} -\lambda_{N-2} e_{N-1} + \dfrac{L_N}{L_{N-1}}(\lambda_{N-1} - v_d)e_N, & \rho_N \le 0 \\ -\lambda_{N-2} e_{N-1} + \dfrac{L_N}{L_{N-1}}(\lambda_{N-1} + w_b)e_N, & \rho_N > 0 \end{cases} \\ \dot{e}_N &= -\lambda_{N-1} e_N \end{aligned} \tag{33}$$

Theorem 1.2. *$e_i = 0$, for $i = 1,2,\dots,N$ is the unique and isolated equilibrium point of the closed-loop system (33) and is guaranteed to be globally exponentially stable. The rate of exponential convergence depends on the control design parameters λ_i, $i = 0,1,\dots,N-1$.*

The proof of Theorem 1.2 is similar to the proof of Theorem Theorem 1.1. According to Theorem 1.2, the steady state value of ρ_i is $\rho_{i,ss} = \rho_{d,c}$, $i = 1,\dots,N$. The steady state value of v_i is $v_{i,ss} = v_d - \sum_{j=i+1}^{N} R_i / \rho_{d,c}$, $i = 1,\dots,N-1$. Therefore, by applying the coordinated VSL and RM controller, ρ_1 through ρ_N are stabilized and homogenized. The effect of a ramp flow is compensated by its upstream VSL and does not affect downstream traffic. If $R_i = 0$, then $v_{i,ss} = v_d$, for $i = 1,\dots,N-1$. That is the upstream speed limit converges to v_d. By adjusting the value of v_d, we can guarantee that the shockwave resulted by speed deviation between actual traffic flow and the triangular fundamental diagram is eliminated.

Now let us consider the dynamics of ρ_0 and v_0. Since q_1 converges to $v_d\rho_{d,c}$, if the demand $d > v_d\rho_{d,c}$, ρ_0 will increase. Once $\rho_0 > \rho_{j,0} - d/w_0$, we have

$$\dot{\rho}_0 = \frac{1}{L_0}\left(w_0(\rho_{j,0} - \rho_0) - v_0\rho_0 + R_0\right) \tag{34}$$

Substitute (32) into (34), we have

$$\dot{\rho}_0 = \frac{1}{L_0}\left(w_0(\rho_{j,0} - \rho_0) - v_d\rho_{d,c} + \sum_{j=0}^{N} R_j\right)$$

Assume that $\sum_{j=0}^{N} R_j$ is constant, then

$$\rho_0 = \rho_{j,0} + \frac{\sum_{j=0}^{N} R_j - v_d\rho_{d,c}}{w_0}$$

is a stable equilibrium point. As long as $\sum_{j=0}^{N} R_j < v_d\rho_{d,c}$, ρ_0 will not exceed the jam density $\rho_{j,0}$ and v_0 will not go negative, thus the VSL controller is feasible.

For driver's acceptance and safety, we as well apply the constraints (22)-(24) to the VSL controller (32).

Design of the RM Controller

According to Theorem 1.2, the VSL controller (32) can stabilize the system and improve the mobility as long as the net ramp flow is lower than the bottleneck capacity. It seems that RM control is unnecessary. However, if no RM is applied and large ramp flows flush into the mainline, the merging of ramp flows will severely disturb the mainline flow. Furthermore, when the net ramp flow is high, the VSL controller (32) will suppress the mainline flow in order to spare the capacity for the ramp flows. That is, without RM control, the ramp flow will always have priority which may harm the fairness between the ramp flows and the mainline flow, or even make the VSL controller infeasible. Furthermore, the RM controller should be able to manage the queue on the ramps so that the queues do not spill backwards to the urban road network. We adopt the ALINEA/Q, which modifies the classic ALINEA ramp metering strategy with queue adjustment. The original ALINEA/Q method proposed in (Smaragdis and Papageorgiou 2003) includes the downstream occupancy and the queue length in the feedback loop. In this paper, to be consistent with the VSL controller, we use the downstream density instead of occupancy.

For an on-ramp i, two RM rates, $r_i^d(k)$ and $r_i^q(k)$, are decided respectively based on the downstream density and the queue length on the ramp at each time step $t = kT_c$. The final RM rate $r_i(k)$ is the maximum of the two. i.e.,

$$\begin{aligned} r_i^d(k) &= r(k-1) + \beta_d\left[\left(\rho_{d,c} - \rho_i(k)\right)\right] \\ r_i^q(k) &= \beta_q\left(w_i^r - w_i(k)\right) + d_i(k-1) \\ r_i(k) &= \max\{r_i^d(k), r_i^q(k)\} \end{aligned} \tag{35}$$

where $\rho_i(k)$ is the density in the highway section that connects to ramp i, $w_i(k)$ is the queue length on ramp i at time step k, $d_i(k-1)$ is the demand from ramp i within time step $k-1$, w_i^r is the reference queue length of ramp i. $r_i^d(k)$ is an integral feedback controller that regulates $\rho_i(k)$ to be close to $\rho_{d,c}$, which helps maintain the vehicle density on mainline at the desired equilibrium value. $r_i^q(k)$ adjusts the RM rate in order to prevent the queue length from being too large, i.e., if $w_i(k)$ is larger than w_i^r, the RM rate will increase to discharge excessive vehicles in the queue and newly arrived vehicles. Since the final RM rate is the maximum of the two, the ramp flow will get the priority to pass the bottleneck if the ramp queue is large, while the mainline flow will get the priority if the vehicle density on the mainline is high. In this way, the ALINEA/Q strategy maintains the fairness between the ramp flows and the mainline flow and avoids the ramp queues from piling up towards the urban road.

Numerical Simulations

In this section, we use the microscopic simulator VISSIM to carry out Monte Carlo simulations to evaluate the performance of the coordinated VSL, RM and lane change control on traffic mobility, safety and the environment.

Scenario Setup

We evaluate the proposed controller on the highway segment in Figure 65. To coordinate with the ramps, we divide the highway segment in to 8 sections, the VSL signs are deployed at the beginning of section 0 through 6. An incident blocks the middle lane at the end of section 7 and creates a bottleneck. 4 on-ramps, which are equipped with RM, and 5 off-ramps are connected to the highway segment. The lane change control is deployed at the beginning of section 7. The incident occurs at 5 minutes after simulation starts, and lasts for 30 min. The capacity of the highway segment is 6800 veh/h without incident. During the incident, the ideal bottleneck capacity is about 4500 veh/h. We load the network with the real demand at 5pm on Monday, which is a peak hour. The mainline demand is 4500 veh/h, the on-ramp demand from upstream to downstream are 400 veh/h, 500 veh/h, 300 veh/h, 300 veh/h respectively.

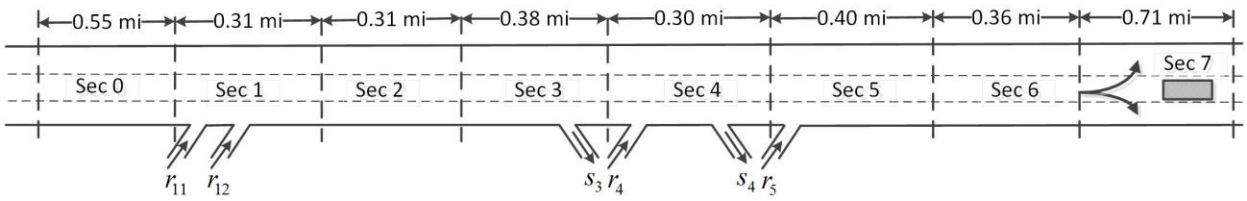


Figure 65. Geometry of Simulation Network

Simulation Results

Figure 66 shows the bottleneck flow with and without the coordinated VSL, RM and lane change control. When there is no control, the flow rate decreases immediately to around 3000 veh/h due to the lane blockage and capacity drop, and increases right away after the incident is removed as the queue in the bottleneck area flushes downstream. When the controller is applied, the flow rate decreases to around 4200 veh/h, which is higher than that in the no control case since the capacity drop is avoided by the lane change control and VSL stabilizes the vehicle densities. The bottleneck flow starts increasing about 10 min after the incident is removed as the high density area is held in section 0 by the VSL controller. The high density wave moves forward from section 0 and the flow rate q_b starts increasing once the wave front reach the bottleneck.

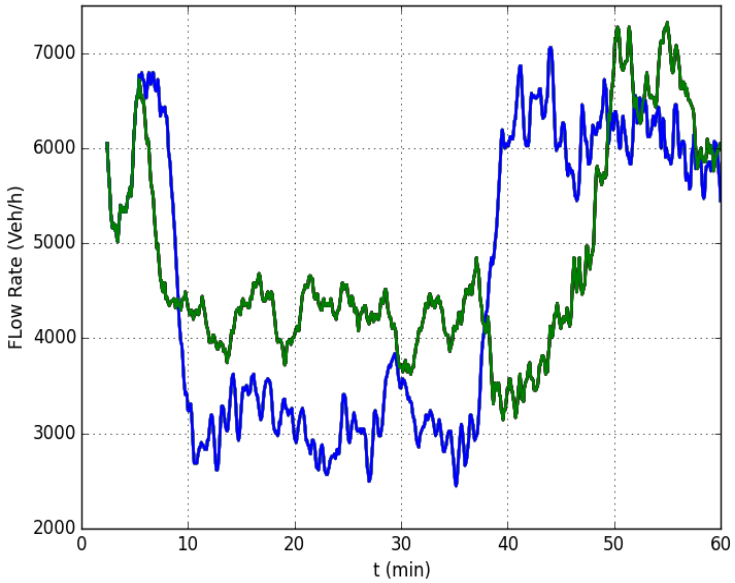
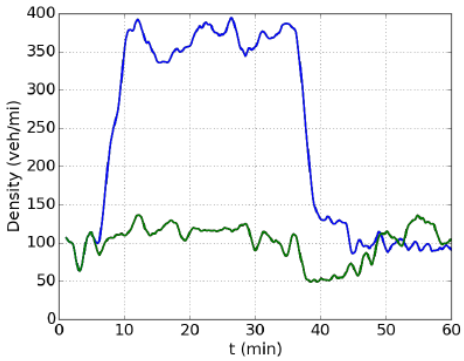
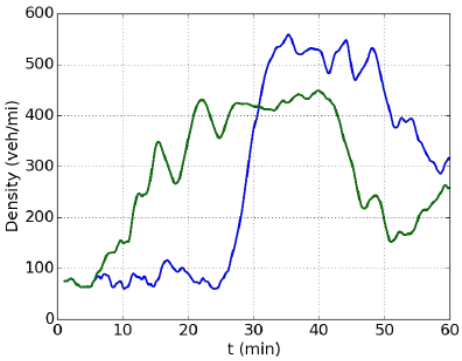


Figure 66. Bottleneck Flow —with control, —no control

Figure 67 shows the curve of ρ_7 and ρ_0, which are the vehicle density of the discharging section and the first VSL controlled section, respectively. When there is no control, ρ_7 starts increasing immediately as the incident occurs at $t = 5$ min. In addition the shockwave propagates upstream, which makes ρ_0 starts increasing at $t = 25$ min and reaches 500 veh/mi. The high density in section 0 does not discharge until 15 min after the incident is removed. When the coordinated controller is applied, ρ_7 increases slightly and is stabilized at 110 veh/mi. ρ_0 increases immediately after the incident since v_0 decreases to reduce the flow into downstream sections and is stabilized at around 400 veh/h which is lower than that without control.



(a) Density in section 7

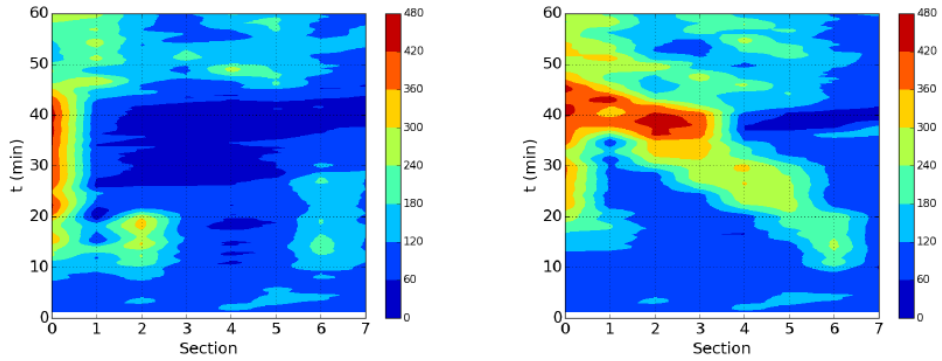


(b) Density in section 0

Figure 67. Vehicle Densities w/ and w/o Control —with control, —no control

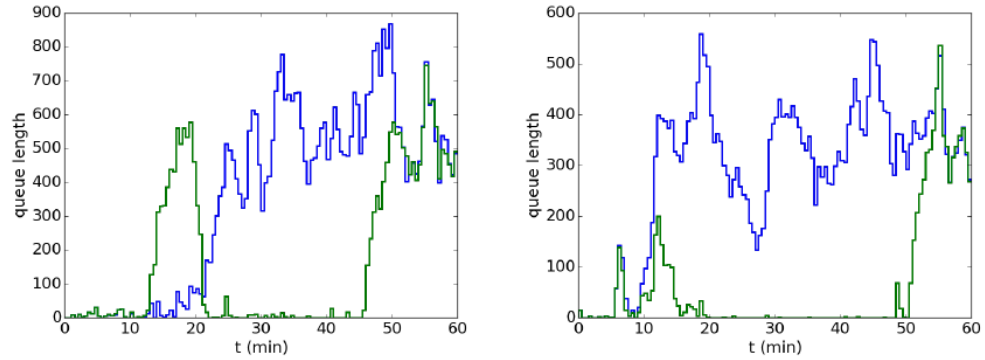
Figure 68 demonstrates the contour plot of vehicle densities with respect to time and space with different values of v_d. When $v_d = 40$ mi/h, high density is held in section 0 during the incident, while downstream sections are highly homogenized. ρ_2 is higher than $\rho_{d,c}$ at the beginning of the incident as the ramp flows r_{11} and r_{12} flush in but then discharged under control. The density in section 6 is slightly higher than $\rho_{d,c}$ as vehicles receive the lane change recommendations and make lane changes thus slightly disturbs upstream flow. When $v_d = 65$ mi/h, a shockwave propagates upstream. After the incident is removed, the vehicles in section 0 flush downstream and meet with the shockwave, which leads to a high density area in section 2. However, in this case, the discharging section is still well protected. As the shockwave propagates upstream, vehicle densities converge to $\rho_{d,c}$ gradually from downstream section to upstream section. This is because we use the cascade structure of VSL controller in Figure 49, which attenuates the shockwave section by section. Thus the controller is robust to parameter selection.

Figure 69 shows the queue length on ramp r_{11} and r_3, with RM control alone and with the coordinated controller. With RM control alone, the queues pile up fast as the densities in mainline increase. Due to the queue adjustment mechanism of ALINEA/Q, the queue lengths are maintained around the reference value. With the coordinated controller, the queue lengths increase in the transient process when the incident begins and the mainline density is being adjusted to the desired level and then discharge fast. After the incident is removed, large flow flushes downstream, the RM controller decrease the rate to give priority to the mainline, therefore the queue lengths increase.



(a) $v_d = 40$ mi/h (b) $v_d = 65$ mi/h

Figure 68. Density Contours



(a) Queue length on r_{11} (b) Queue length on r_4

Figure 69. Queue Length w/ and w/o Control

We use the following metrics to evaluate the performance of the coordinated controller. To evaluate traffic mobility, we use: (a) average travel time $\bar{T}_t$; for traffic safety, use (b) average number of stops $\bar{s}$ and (c) average number of lane changes $\bar{c}$; for the environment, we use (d) average emission of CO2 and (e) average fuel consumption. The detailed definition of the above metric can be found in (Y. Zhang and Ioannou 2017a).

Table 16 shows the evaluation results. The improvement in traffic mobility, safety and the environment is significant. The average travel time is reduced by about 27% as the bottleneck throughput is increased. For traffic safety, the number of stops dramatically decreased by 81% as the lane change control prevented vehicles from stopping at the bottleneck and waiting for lane changes. The 10% reduction in number of lane changes is contributed by both homogenization of mainline flow and the regulated merging behavior of ramp flows. For the environment metrics, the reductions of CO2 emission and energy consumption are usually proportional to each other, which are both around 8% in this case.

Table 16. Evaluation Results

Control Type	No Control	RM + VSL	Improvement
$\bar{T}_t$(min)	15	11	27%
$\bar{s}$	23	4	82%
$\bar{c}$	5.1	4.6	10%
CO2 (g/veh/mi)	585	538	8%
Fuel (g/veh/mi)	187	172	8%

Comparison of Feedback Linearization and Model Predictive Strategies for VSL Control

Section based on the publication:

Y. Zhang, I.I. Sirmatel, F. Alasiri, P.A. Ioannou, and N. Geroliminis, "Comparison of feedback linearization and model predictive techniques for variable speed limit control," 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018.

Introduction

Given the fact that LC control is able to relieve or eliminate the capacity drop, one important question arising at this point is that if other VSL control strategies are combined with the LC control, will the system performance exceed the performance under the FL controller? Intuitively, since MPC control follows an optimization based routine, it should provide the 'optimal' performance to some extent. However, FL controller guarantees exponential stability of the equilibrium point with highest bottleneck flow rate. Therefore, by tuning the feedback gain, the FL controller should be able to force the system to converge as fast as possible, only limited by the saturation of control input.

In this section, we propose FL and MPC schemes for VSL-actuated highway traffic, where we assume that an LC controller is active just upstream of the bottleneck. Both controllers are designed with a CTM-based model representing the ideal system. TTS performance and robustness with respect to perturbations on model parameters and measurement noise of the proposed controllers are evaluated via simulation studies. Results show both VSL controller is able to improve the total time spent under different levels of perturbation and measurement noise. Furthermore, feedback linearization VSL can provide better performance than model predictive VSL with much less computational effort.

Nonlinear Model Predictive Control

Model predictive control strategy generates the control command at each control step by solving a finite horizon optimal control problem in a receding horizon manner. Here, we formulate the cost function of the MPC problem as the quadratic error of the states of system (17). To take into consideration the vehicles that are blocked upstream the VSL controlled segment, we augment the system by add a new state Q, that is

$$\dot{Q} = d - q_0, \tag{36}$$

with $Q = 0$ at $t = 0$. Therefore, if the number of vehicles upstream of section 0 is greater than the number at time 0, $Q > 0$, otherwise $Q \leq 0$. We should note here that the introduction of Q is only for the purpose of evaluating the TTS. Both the FL and MPC controllers are implemented based on system (17). The performance metric TTS is defined as follows:

$$\text{TTS} = \int_0^T Q\,(t) + \sum_{i=0}^{N} \rho_i\,(t) L_i\,dt \tag{37}$$

The open-loop highway system (17) can be implicitly expressed as

$$\dot{e} = f(e, u) \tag{38}$$

Here we formulate the problem of finding the VSL commands $u(\cdot)$ that try to maintain system (38) at the equilibrium point as the following finite-horizon constrained optimal control problem (OCP):

$$\begin{aligned} \underset{u(\cdot)}{\text{minimize}} \quad & \int_{kT_c}^{kT_c+T_p} e(\tau)^T \tilde{Q} e(\tau) + u(\tau)^T \tilde{R} u(\tau) d\tau \\ \text{subject to} \quad & e(kT_c) = \hat{e}(kT_c) \\ & \dot{e} = f(e, u),\ \forall \tau \in [t, t + T_p] \\ & v_{\min} - v_e \le u(\tau) \le v_{\max} - v_e, \end{aligned} \tag{39}$$

where t is the current control sampling instant in time, $\hat{e}(t)$ is the measurement on error states taken at that instant, $\tilde{Q}$ and $\tilde{R}$ are weighting matrices on error and control input, respectively, whereas T_p is the prediction horizon. The optimization problem is solved at the beginning of each control step kT_c, with $\hat{e}(kT_c)$ as the initial condition. Constraint (12) has already been included in the constraints of the optimization problem. (10) and (11) are also applied to the MPC VSL commands before applied to the system.

Due to the continuous-time dynamics, the OCP (39) is an infinite dimensional optimization problem. We resort to approximating it as a finite dimensional nonlinear program (NLP) via the direct multiple shooting method [@bock1984multiple]. Details on direct methods from numerical optimal control literature can be found in (Diehl et al. 2006).

Numerical Simulation

In this section, macroscopic simulation is used to evaluate the performance and robustness of the FL and NMPC schemes combined with LC.

Scenario setup

The FL and MPC controllers have evaluated on the network shown in Figure 55 In our simulation, the incident happens 5 minutes after the simulation starts, and it lasts for 30 min. The nominal demand is 6000 veh/h. The desired equilibrium point of this network is calibrated to be:

$$\begin{aligned} &\rho_0^{\mathrm{e}} = 278 \text{ veh/mi} \\ &\rho_1^{\mathrm{e}} = \rho_2^{\mathrm{e}} = \cdots = \rho_7^{\mathrm{e}} = 110 \text{ veh/mi} \\ &v_0^{\mathrm{e}} = 15.8 \text{ mi/h} \\ &v_1^{\mathrm{e}} = v_2^{\mathrm{e}} = \cdots = v_7^{\mathrm{e}} = 40 \text{ mi/h} \end{aligned}$$

For the FL controller, we choose $\lambda_i = 50$ for $i = 0,1, \ldots,6$. The NMPC controller is implemented using the direct multiple shooting method via the CasADi toolbox (Andersson 2013) in MATLAB 8.5.0 (R2015a), on a 64-bit Windows PC with 3.4-GHz Intel Core i7 processor and 8-GB RAM, where IPOPT (Wächter and Biegler 2006) is used for solving the NLPs. In our simulation, we

choose the prediction horizon $T_p = 10$ min, which is much greater than the control time step $T_c = 30$ s. Weight matrices are chosen as $\tilde{Q} = \mathbf{I}$ and $\tilde{R} = 0.1\mathbf{I}$, with $\mathbf{I}$ denoting the identity matrix of appropriate dimensions. The computation time of NMPC is around 0.35 seconds, whereas it is negligible for FL. The NMPC scheme is still computationally tractable, as its computation time of 0.35 s per step is negligible with respect to the control time step of 30 s.

Performance and Robustness Analysis with Macroscopic Simulations

To compare the performance and robustness of the FL and MPC VSL controllers, we evaluate the following criteria for the two controllers: 1) Total time spent (TTS) as defined in (37), and sensitivity of TTS with respect to 2) perturbation on traffic demand, 3) perturbation on model parameters and 4) measurement noise. In the simulation, the FL and MPC controllers are synthesized with the ideal model (38), but the control command are applied on a perturbed model. The structure of the simulation system is shown in Figure 70. For the traffic demand, we add up to $\pm 20\%$ perturbation on the nominal demand 6000 veh/h. For the model parameters, as shown in Figure 71, we respectively add up to $\pm 20\%$ perturbation on the nominal value of $\rho_{d,c}$ and C_b, which directly alter the shape of the fundamental diagram of the bottleneck section. For the measurement noise, we use Gaussian white noise with different levels of standard deviation up to $\sigma = 0.1\rho_{cb}$ to match the scale of the density measurements.

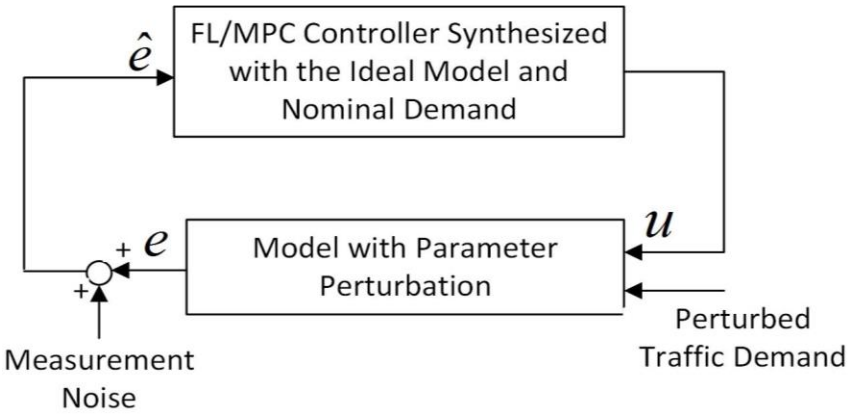


Figure 70. Simulation System

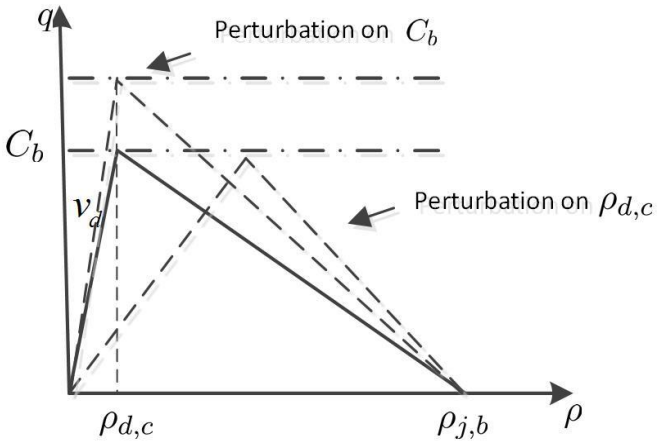


Figure 71. Perturbation on the Model Parameters

Figure 72 shows the behavior of the vehicle density in the discharging section under FL and MPC controller. Both controllers are able to maintain the density around the desired value $\rho_7^e = 110$ veh/h after the incident occurs at $t = 5$ min. The oscillation is introduced by the roundup-to-5 constraint. However, the MPC controller introduces higher frequency chattering and a sharp decrease at the beginning of the incident.

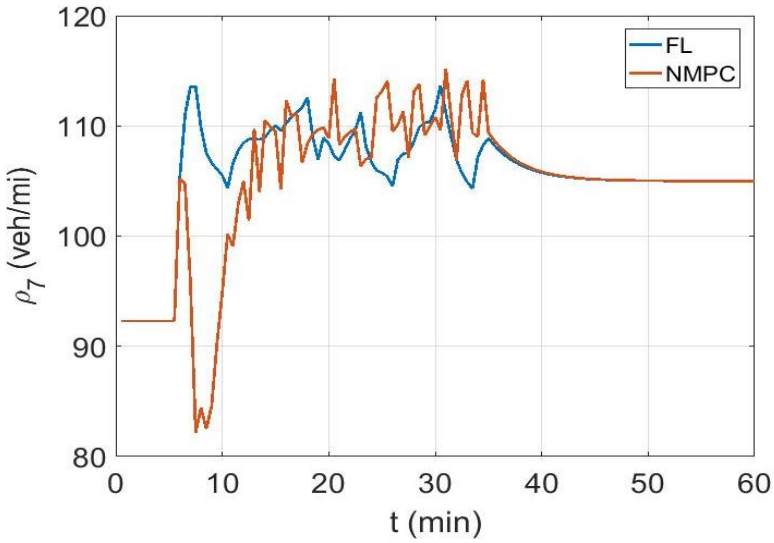


Figure 72. ρ_7 with FL and MPC

A series of simulation experiments are conducted with different levels of perturbation and measurement noise. Figure 73 shows how TTS varies with varying demand levels. The figure showcases that both controllers are able to function properly under various levels of demand, the TTS increases and decreases approximately linear with the demand. This demonstrates that both MPC and FL VSL controllers are robust with respect to the variation of demand, which is due to the selection of the desired equilibrium point (13)-(15). At the equilibrium point, the speed limit in section 0 is decreased to block excessive traffic demand at upstream of the entire control segment, therefore the bottleneck flow is not affected. Furthermore, under different

levels of perturbation, the performance of FL and MPC controller are similar. But the TTS of FL is always slightly lower than that of MPC, which shows that MPC fails to beat FL in TTS although the control commands are generated by solving the optimization problem in receding horizon fashion.

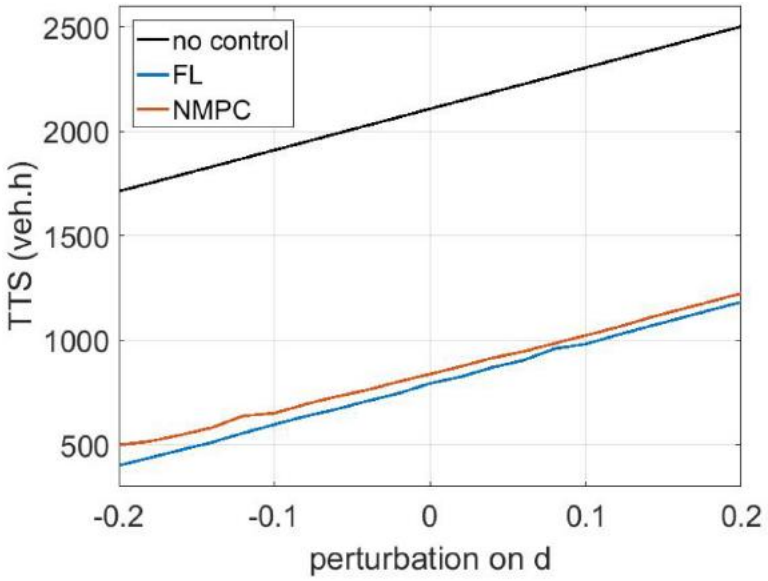


Figure 73. Performance sensitivity of no control (black), FL (blue), and NMPC (red) to perturbations on demand d.

In Figure 74 and Figure 75, the change in TTS is plotted with respect to different values of perturbation on C_b and $\rho_{d,c}$, respectively. These results show that both controllers achieve significant improvements over the no control case and are able to operate properly even under situations with high amount of uncertainty in these model parameters. With perturbation on C_b, the TTS under FL and MPC are increased by 45% and 43% in the worst case, respectively. Considering the fact that in this case the bottleneck capacity is decreased by 20% as a baseline, the TTS does not increase too much due to the modeling error and is still much lower than that in the no control case. The worst case for the perturbation on $\rho_{d,c}$ is 27% worse than the non-perturbed value for FL, and 16% for NMPC.

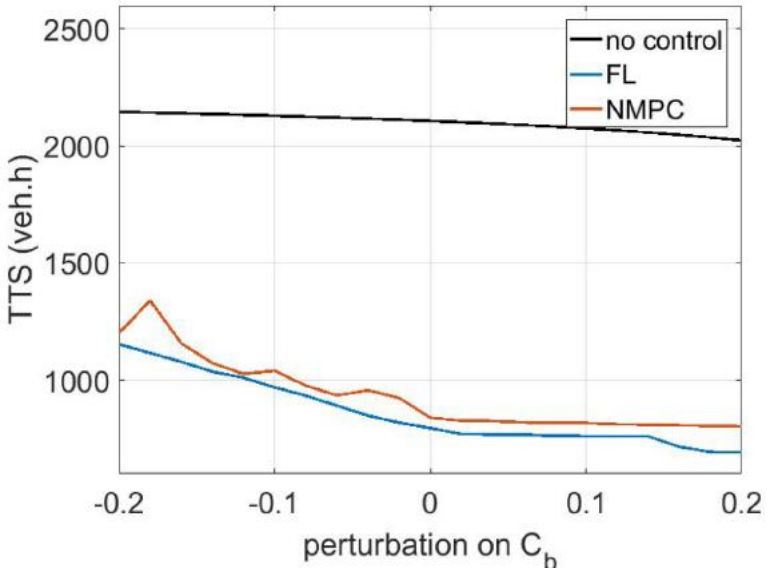


Figure 74. Performance sensitivity of no control (black), FL (blue), and NMPC (red) to perturbations on C_b.

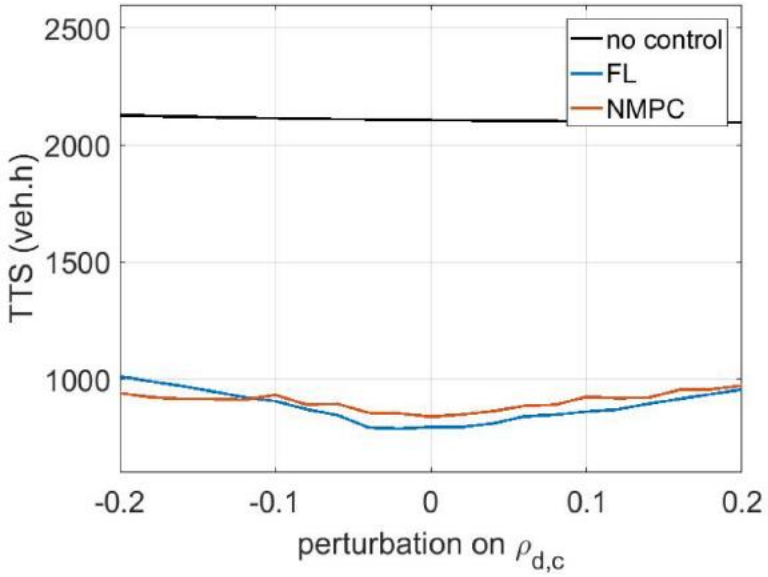


Figure 75. Performance sensitivity of no control (black), FL (blue), and NMPC (red) to perturbations on ρ_"d,c" .

The sensitivity of TTS performance in the case of varying levels of standard deviation in measurement noise is given in Figure 76, which shows that the TTS under both controllers increases with the standard deviation of measurement noise. However, the system does not diverge as the no control case. The performance of FL is always better than that of NMPC in this case.

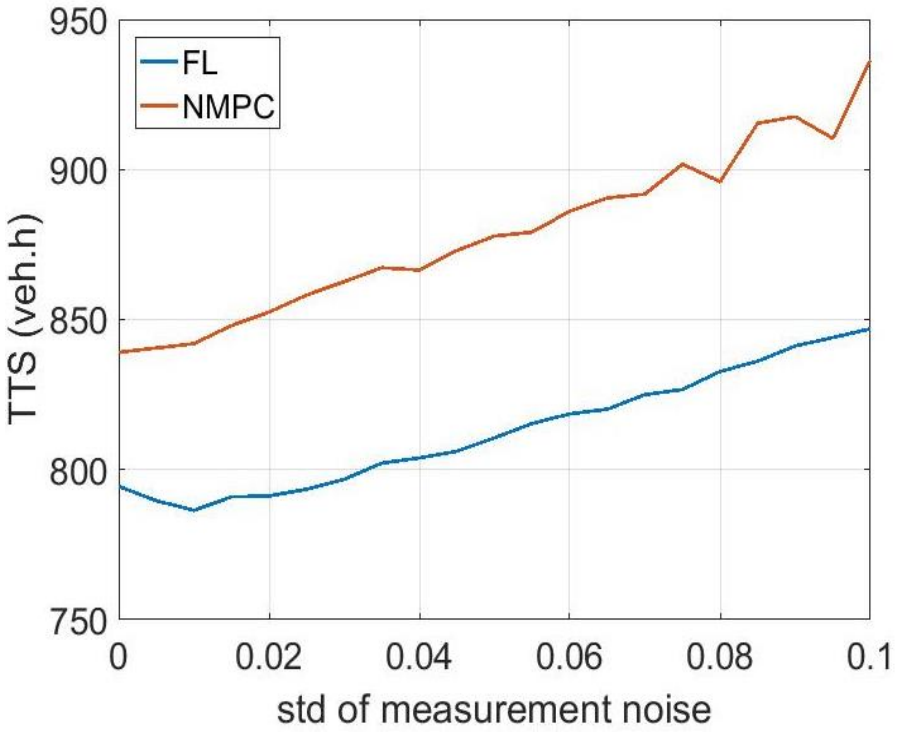


Figure 76. Performance sensitivity of FL (blue) and NMPC (red) to increasing levels of standard deviation in measurement noise.

Performance and Robustness Analysis with Microscopic Simulations

Table 17 shows the microscopic simulation results with calibrated model parameter set:

$$w_1 = 14 \text{ mi/h},\ w_b = 40 \text{ mi/h},\ \rho_{cb} = 110 \text{ veh/mi}$$

The performance of the MPC controller is similar to that of the FL controller.

Table 18-Table 20 demonstrate the simulation results of MPC and FL controller under different values of model parameters. From the result, we can see that the FL controller is robust with respect to the perturbations on w_1, w_b and ρ_{cb}. As to MPC, the mobility performance is significantly adversed by the perturbations on w_1 and ρ_{cb}, which both change the value of the equilibrium point. But MPC is robust with respect to the perturbations on w_b which does not change the equilibrium point and can be compensated by the control input.

Table 17. Evaluation Results with Original Parameters

		TTT (hr)	Stops	LC	CO (g/veh/mi)	Nox (g/veh/mi)	CO2 (g/veh/mi)	Energy (g/veh/mi)
No Control	mean ± std	1270 ± 42	23.2 ± 1.3	6.6 ± 0.2	3.4 ± 0.1	1.8 ± 0.1	605 ± 20	194 ± 6
	Improvement	-	-	-	-	-	-	-
LC Only	mean ± std	1075 ± 40	10.5 ± 0.9	5.9 ± 0.3	3.4 ± 0.1	1.7 ± 0.1	552 ± 16	176 ± 5
	Improvement	15%	55%	11%	0%	6%	9%	9%
FL	mean ± std	1036 ± 36	9.9 ± 1.3	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	529 ± 13	169 ± 4
	Improvement	18%	57%	17%	12%	11%	13%	13%
MPC	mean ± std	1018 ± 41	8.7 ± 1.2	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	525 ± 15	168 ± 5
	Improvement	20%	63%	17%	12%	11%	13%	13%

Table 18. Evaluation Results under Different w1

			TTT (hr)	Stops	LC	CO (g/veh/mi)	Nox (g/veh/mi)	CO2 (g/veh/mi)	Energy (g/veh/mi)
FL	w_1=9	mean ± std	1036 ± 36	9.9 ± 1.3	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	529 ± 13	169 ± 4
		Improvement	18%	57%	17%	12%	11%	13%	13%
	w_1=14	mean ± std	1036 ± 36	9.9 ± 1.3	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	529 ± 13	169 ± 4
		Improvement	18%	57%	17%	12%	11%	13%	13%
	w_1=6	mean ± std	1036 ± 36	9.9 ± 1.3	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	529 ± 13	169 ± 4
		Improvement	18%	57%	17%	12%	11%	13%	13%
MPC	w_1=9	mean ± std	1096 ± 55	12.3 ± 2.4	5.5 ± 0.2	3.1 ± 0.1	1.6 ± 0.1	533 ± 16	170 ± 5
		Improvement	14%	47%	17%	9%	11%	12%	12%
	w_1=14	mean ± std	1018 ± 41	8.7 ± 1.2	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	525 ± 15	168 ± 5
		Improvement	20%	63%	17%	12%	11%	13%	13%
	w_1=6	mean ± std	1226 ± 61	12.1 ± 1.9	5.6 ± 0.3	3.1 ± 0.1	1.6 ± 0.1	546 ± 20	174 ± 6
		Improvement	3%	48%	15%	9%	11%	10%	10%

Table 19. Evaluation Results under Different ρcb

			TTT (hr)	Stops	LC	CO (g/veh/mi)	NOx (g/veh/mi)	CO2 (g/veh/mi)	Energy (g/veh/mi)
FL	ρ_{cb} = 100	mean ± std	1024 ± 44	8.8 ± 2	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	528 ± 14	169 ± 5
		Improvement	19%	62%	17%	12%	11%	13%	13%
	ρ_{cb} = 110	mean ± std	1036 ± 36	9.9 ± 1.3	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	529 ± 13	169 ± 4
		Improvement	18%	57%	17%	12%	11%	13%	13%
	ρ_{cb} = 120	mean ± std	1031 ± 43	9.4 ± 2.2	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	526± 15	168 ± 4
		Improvement	19%	59%	17%	12%	11%	13%	13%
MPC	ρ_{cb} = 100	mean ± std	1236 ± 41	11.4 ± 0.3	5.5 ± 0.2	3.1 ± 0.1	1.6 ± 0.1	544 ± 16	174 ± 5
		Improvement	3%	51%	17%	9%	11%	10%	10%
	ρ_{cb} = 110	mean ± std	1018 ± 41	8.7 ± 1.2	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	525 ± 15	168 ± 5
		Improvement	20%	63%	17%	12%	11%	13%	13%
	ρ_{cb} = 120	mean ± std	1242 ± 35	11.6 ± 1.0	5.5 ± 0.2	3.1 ± 0.1	1.6 ± 0.1	542 ± 17	173 ± 6
		Improvement	2%	50%	17%	9%	11%	10%	11%

Table 20. Evaluation Results under Different wb

			TTT (hr)	Stops	LC	CO (g/veh/mi)	Nox (g/veh/mi)	CO2 (g/veh/mi)	Energy (g/veh/mi)
FL	w_b=20	mean ± std	1025 ± 36	9.6 ± 1.0	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	527 ± 13	169 ± 4
		Improvement	19%	59%	17%	12%	11%	13%	13%
	w_b=40	mean ± std	1036 ± 36	9.9 ± 1.3	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	529 ± 13	169 ± 4
		Improvement	18%	57%	17%	12%	11%	13%	13%
	w_b=60	mean ± std	1042 ± 34	10.2 ± 1.8	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	526 ± 15	168 ± 4
		Improvement	18%	56%	17%	12%	11%	13%	13%
MPC	w_b=20	mean ± std	1098 ± 58	12.4 ± 2.4	5.5 ± 0.2	3.1 ± 0.1	1.6 ± 0.1	533 ± 16	170 ± 5
		Improvement	14%	47%	17%	9%	11%	12%	12%
	w_b=40	mean ± std	1018 ± 41	8.7 ± 1.2	5.5 ± 0.2	3.0 ± 0.1	1.6 ± 0.1	525 ± 15	168 ± 5
		Improvement	20%	63%	17%	12%	11%	13%	13%
	w_b=60	mean ± std	1092 ± 53	12.3 ± 2.2	5.5 ± 0.2	3.1 ± 0.1	1.6 ± 0.1	529 ± 15	169 ± 5
		Improvement	14%	47%	17%	9%	11%	13%	13%

Stability Analysis and Variable Speed Limit Control of a Traffic Flow Model

Section based on the publication:

Y. Zhang and P.A. Ioannou, "Stability analysis and variable speed limit control of a traffic flow model," Transportation Research Part B: Methodological 118 (2018): 31-65.

Stability Analysis of Cell Transmission Model under All Operating Conditions

In Combined Variable Speed Limit and Lane Change Control and Coordinated Variable Speed Limit, Ramp Metering, and Lane Change Controller, we designed a coordinated variable speed limit, ramp metering and lane change control based on the first-order cell transmission model. However, the analysis of dynamical behavior and stability properties of the open-loop cell transmission model which takes capacity drop into consideration is missing from the previous work, which makes it difficult for us to perform an analytical comparison of the open-loop and closed-loop performance of the VSL controlled cell transmission model. In addition, the analysis of the closed-loop behavior in In Combined Variable Speed Limit and Lane Change Control and Coordinated Variable Speed Limit, Ramp Metering, and Lane Change Controller is performed with a simplified CTM, i.e., consider only the region in the state space near the desired equilibrium point (13) and under the assumption that the demand is higher than the bottleneck capacity. It remains unclear whether the global stability of the desired equilibrium point is still valid with the complete CTM and in other operating scenarios.

In (Gomes et al. 2008), Gomes et al. performed a thorough analysis of the equilibrium points and their stability properties of the CTM model. However, the authors did not take the capacity drop phenomenon into consideration. Reference (Lovisari et al. 2014) developed sufficient conditions for the stability of the equilibrium points of CTM in terms of connectivity of a graph associated with the traffic network. The results of (Gomes et al. 2008) and (Lovisari et al. 2014) are established based on the monotonicity of CTM. However, if the CTM is modified to account for capacity drop and the fact that the discharging flow rate of a congested road section decreases with density density (Y. Zhang and Ioannou 2017a; Srivastava, Jin, and Lebacque 2015; H.-Y. Jin and Jin 2015; Kontorinaki et al. 2016), then the CTM is no longer monotone.

Therefore, in this section, We use the CTM which take into consideration the effect of capacity drop which is due to microscopic phenomena such as forced lane changes at a bottleneck (Y. Zhang and Ioannou 2017a) and the decreasing discharging flow of the road section, then consider all possible traffic flow scenarios, identify all equilibrium points and analyze their stability properties for a single road section, then extend the results to arbitrary number of sections under different traffic demand levels and capacity constraints as well as under all initial density conditions, based on which the design of the VSL controller which guarantees global stability of the closed-loop system with complete CTM and under all possible operating scenarios is perform in the next section.

Stability of Traffic Flow in a Single-Section Road Segment

Consider a single road section of unit length with an inflow q_1 and outflow q_2, expected to meet a demand of flow d as shown in Figure 77.

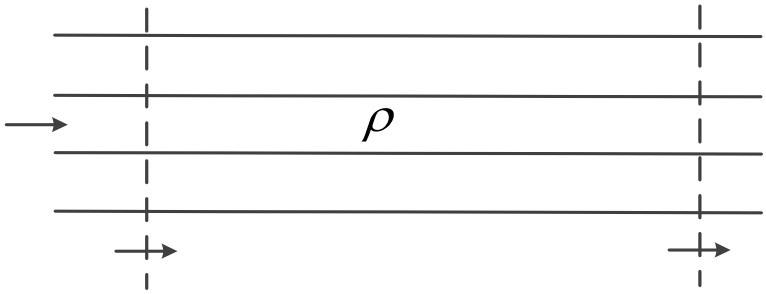


Figure 77. Single Road Section

We assume that the vehicle density ρ is uniform along the section, i.e., it is independent of distance from the entrance to the exit of the section and does not vary across the lanes in the vertical direction. Under these assumptions, the evolution of ρ with respect to time is given by the following differential equation:

$$\dot{\rho} = q_1 - q_2, 0 \le \rho(0) \le \rho^j, \tag{40}$$

Where

$$\begin{aligned} q_1 &= \min\{d, C, w(\rho^j - \rho)\}, \\ q_2 &= \begin{cases} \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), (1-\epsilon(\rho))C_d\} & \text{if } C_d < C \\ \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), C_d\} & \text{otherwise} \end{cases}, \\ v_f\rho_c &= w(\rho^j - \rho_c) = \tilde{w}(\tilde{\rho}^j - \rho_c) = C, \\ 0 &< \rho_c < \rho^j, 0 < \tilde{w} < w, v_f > 0, \\ \epsilon(\rho) &= \begin{cases} 0 & \text{if } 0 \le \rho \le \dfrac{C_d}{v_f} \\ \epsilon_0 & \text{otherwise} \end{cases}, \end{aligned} \tag{41}$$

and the constants in equation (40)-(41)are defined as follows:

- C: the capacity of the road section.
- w: the back propagation speed.
- ρ^j: jam density, the highest density possible, at which $q_1 = 0$.
- v_f: free flow speed of the road section.
- $\tilde{w}$: the rate that the outflow q_2 decreases with ρ, when $\rho \ge \rho_c$.
- $\tilde{\rho}^j$: the jam density associated with outflow q_2.
- ρ_c: the critical density of the road section, at which $v_f\rho_c = w(\rho^j - \rho_c) = \tilde{w}(\tilde{\rho}^j - \rho_c) = C$.
- C_d: the downstream capacity.

In equation (41), the inflow q_1 is dictated by the upstream demand d as well as the potential ability of the section to absorb traffic flow, which is the value $\min\{C, w(\rho^j - \rho)\}$. If $\rho \leq \rho_c$, the section can absorb as much flow as the capacity C, however if $\rho > \rho_c$, the section's ability to absorb upstream flow decreases with ρ at a rate w. When $\rho = \rho^j$, $q_1 = 0$ as the section is completely congested. The outflow q_2 is dictated by the ability of the section to send traffic flow to downstream and the downstream capacity. When $\rho \leq \rho_c$, the section's ability to send traffic flow increases with ρ, but when $\rho > \rho_c$, this ability decreases with ρ at a rate $\widetilde{w}$ (Lebacque 2003; Claudio Roncoli, Papageorgiou, and Papamichail 2015; Srivastava, Jin, and Lebacque 2015; Srivastava and Jin 2016). Since $w > \widetilde{w}$, we have $\widetilde{w}(\tilde{\rho}^j - \rho) > w(\rho^j - \rho)$ for all $\rho > \rho_c$, which captures the phenomenon that if the downstream segment has enough capacity, the density in a congested road section upstream will eventually decrease to a value less than or equal to ρ_c. The capacity of the downstream segment is C_d. If $C_d < C$ and $\rho \leq \frac{C_d}{v_f}$, then the outflow $q_2 = v_f \rho$ can increase up to C_d. However, when $\rho > \frac{C_d}{v_f}$, the section generates more flow than C_d, a queue will form at the outlet, which may cause forced lane changes which in turn reduce the flow speed leading to the reduction of flow to lower than the capacity C_d i.e., to $(1 - \epsilon_0)C_d$ (Y. Zhang and Ioannou 2017a; H.-Y. Jin and Jin 2015). This phenomenon is known as capacity drop. The original CTM is modified to include the capacity drop effect as shown in equation (41). The model (40)-(41) with $\epsilon_0 = 0$ is the CTM of (Daganzo 1994). The $\epsilon_0 > 0$ denotes the level of capacity drop, in which case, despite the availability of flow, q_2 is restricted from reaching the capacity C_d. Note that capacity drop can only occur when the downstream capacity C_d is lower than the capacity of the section C. In system (40)-(41), we model the capacity drop using a reduction in the downstream capacity C_d which has been verified by microscopic simulations using VISSIM in (Y. Zhang and Ioannou 2017a). The modeling of capacity drop has been discussed in (Kontorinaki et al. 2016) more extensively where different models are considered. These models do not change the methodology and results of this section, which can be easily extended to different capacity drop models.

The purpose of this subsection is to analyze the stability properties of the model (40)-(41). Since these properties will depend on the characteristics of the road section defined by the constants C, C_d, the demand d which could vary and the magnitude of capacity drop ϵ_0 which may depend on microscopic effects (Y. Zhang and Ioannou 2017a; Kontorinaki et al. 2016), the following five possible operating scenarios are identified and represented by the sets Ω_i, $i = 1,2,\ldots,5$. The union of these sets $\bigcup_{i=1}^{5} \Omega_i$, as shown in Figure 78, covers all possible situations. Let $I = (C_d, C, d, \epsilon_0)$ be the state of the road section. We analyze the stability properties of the dynamical model (40)-(41) when $I \in \Omega_i$, $i = 1,2,\ldots,5$. Theorem 1.3 presents the results of the analysis.

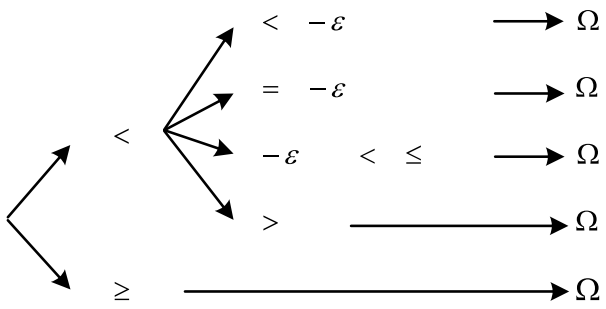


Figure 78. All Possible Operating Scenarios

Theorem 1.3. *For constant but otherwise arbitrary demand d, we have the following results:*

a) *Let $I \in \Omega_1$. Then $\forall \rho(0) \in [0, \rho^j]$, $\rho(t)$ converges exponentially fast to $\frac{d}{v_f}$.*

b) *Let $I \in \Omega_2$. Then*

- $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$, *$\rho(t)$ converges exponentially fast to $\frac{d}{v_f} = \frac{(1-\epsilon_0)C_d}{v_f}$.*
- $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j - \frac{d}{w}]$, *$\rho(t) = \rho(0), \forall t \geq 0$.*
- $\forall \rho(0) \in (\rho^j - \frac{d}{w}, \rho^j]$, *$\rho(t)$ converges exponentially fast to $\rho^j - \frac{d}{w} = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$.*

c) *Let $I \in \Omega_3$. Then*

- $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$, *$\rho(t)$ converges exponentially fast to $\frac{d}{v_f}$.*
- $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$, *$\rho(t)$ converges exponentially fast to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$.*

d) *Let $I \in \Omega_4$. Then $\forall \rho(0) \in [0, \rho^j]$, $\rho(t)$ converges exponentially fast to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$.*

e) *Let $I \in \Omega_5$. Then $\forall \rho(0) \in [0, \rho^j]$, $\rho(t)$ converges exponentially fast to $\frac{\min\{d,C\}}{v_f}$.*

Proof. a) When $I \in \Omega_1$, we plot the relationship of q_1, q_2 given by equation (41) in Figure 79.

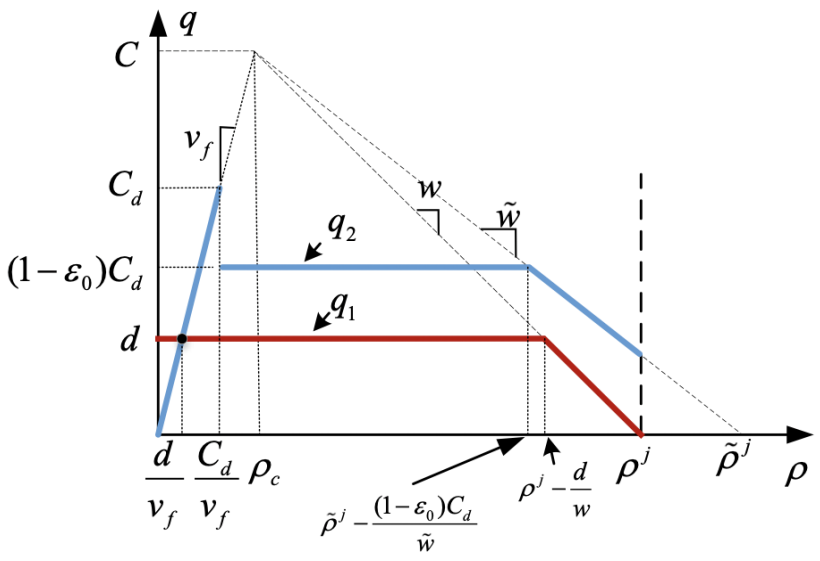


Figure 79. Fundamental Diagram for I ∈ Ω_1

From the density equation (40), the equilibrium points of the system are the values of ρ for which $\dot{\rho} = 0$, which happens when $q_1 = q_2$. It is clear from Figure 79 that the only intersection of q_1 and q_2 is the point $\rho^e = \frac{d}{v_f}$, which implies that this is the only equilibrium of ρ in the region $[0, \rho^j]$ of feasible values of ρ. We define the Lyapunov function

$$V(\rho) = \frac{(\rho - d/v_f)^2}{2},$$

whose time derivative

$$\dot{V}(\rho) = \left(\rho - \frac{d}{v_f}\right)\dot{\rho} = -\left(\rho - \frac{d}{v_f}\right)(q_2 - q_1).$$

We show in Appendix C that

$$\dot{V} \leq -\alpha\left(\rho - \frac{d}{v_f}\right)^2,$$

where $\alpha = \min\{v_f, \frac{(1-\epsilon_0)C_d - d}{\rho^j - d/v_f}, \frac{(\tilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\} > 0$. Hence ρ converges exponentially fast to $\frac{d}{v_f}$ with a rate greater than or equal to α for all possible initial conditions in $[0, \rho^j]$ (P. A. Ioannou and Sun 2012). The rate of convergence is guaranteed to be greater than or equal to α as it is clear from the value of V and $\dot{V}$.

b) When $I \in \Omega_2$, the plot of q_1, q_2 generated from equation (41) is given in Figure 80.

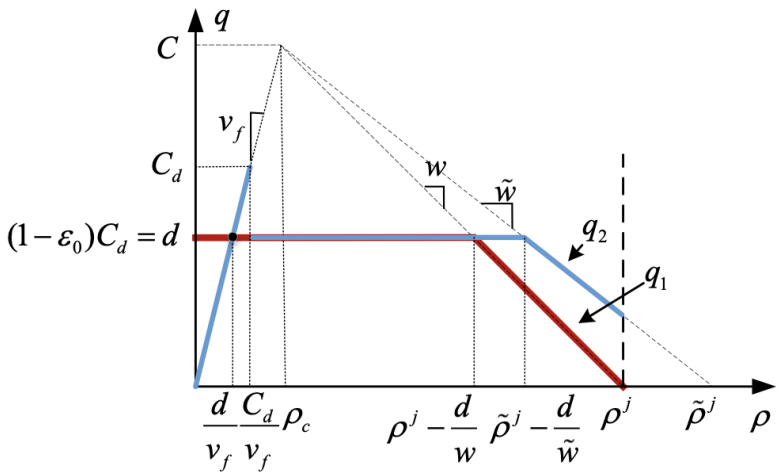


Figure 80. Fundamental Diagram for I ∈ Ω_2

In this case, q_1 and q_2 intersect at one point $\rho = \frac{d}{v_f}$ and $q_1 = q_2$ for all $\rho \in (\frac{C_d}{v_f}, \rho^j - \frac{d}{w}]$. Therefore, we have one isolated equilibrium point $\rho_1^e = \frac{d}{v_f}$ and an equilibrium manifold which is the interval $(\frac{C_d}{v_f}, \rho^j - \frac{d}{w}]$.

From Figure 80, we know that $\forall \rho \in \left[0, \frac{C_d}{v_f}\right]$, $q_1 = (1-\epsilon_0)C_d = d$ and $q_2 = v_f\rho$ which gives

$$\dot{\rho} = -v_f\rho + d, \ \forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right],$$

whose solution is

$$\rho(t) = \frac{d}{v_f} + \left(\rho(0) - \frac{d}{v_f}\right)e^{-v_f t} \leq \frac{C_d}{v_f}.$$

Hence $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$ we have $\rho(t) \in \left[0, \frac{C_d}{v_f}\right]$, $\forall t \geq 0$ and according to the solution above, $\rho(t)$ converges exponentially fast to $\frac{d}{v_f} = \frac{(1-\epsilon_0)C_d}{v_f}$.

For $\rho(0) \in (\frac{C_d}{v_f}, \rho^j - \frac{d}{w}]$, we have $q_1 = q_2$, therefore $\dot{\rho} = 0$, which implies that $\rho(t) = \rho(0), \ \forall t \geq 0$, for all $\rho(0) \in (\frac{C_d}{v_f}, \rho^j - \frac{d}{w}]$.

If $\rho(0) \in (\rho^j - \frac{d}{w}, \rho^j]$, it is clear from Figure 80 that $q_2 > q_1$ which implies that $\dot{\rho} < 0$ until $\rho(t) = \rho^j - \frac{d}{w}$ at which time $\dot{\rho} = 0$. This implies that for all $\rho(0) \in (\rho^j - \frac{d}{w}, \rho^j]$, $\rho(t)$ converges at least asymptotically with time to $\rho^j - \frac{d}{w}$. In Appendix C we show that this rate of convergence is exponential, i.e.,

$$\left|\rho(t) - \left(\rho^j - \frac{d}{w}\right)\right| \leq c_0 e^{-\alpha t}, \forall \rho(0) \in (\rho^j - \frac{d}{w}, \rho^j],$$

where $c_0 > 0$ and $\alpha = \min\{w, w - \tilde{w}\} > 0$.

c) When $I \in \Omega_3$, q_1 and q_2 are plotted in Figure 81.

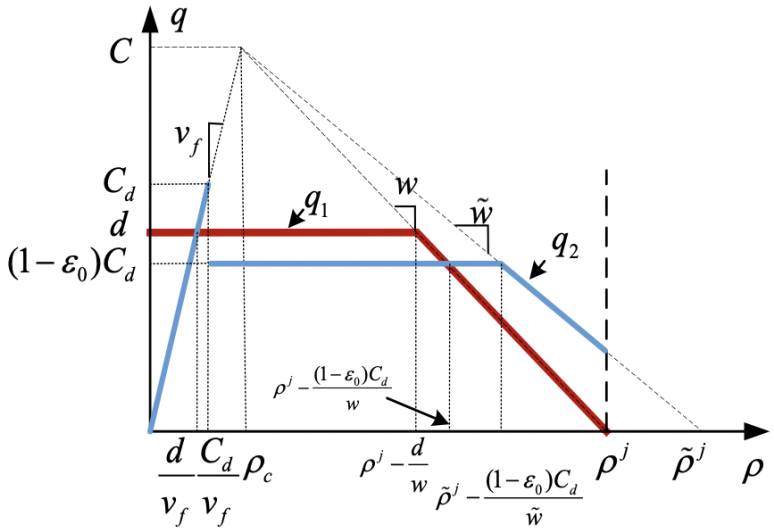


Figure 81. Fundamental Diagram for I ∈ Ω_3

From Figure 81, it is clear that the only values of ρ for which $q_1 = q_2$ are $\frac{d}{v_f}$ and $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$, which implies that the system has two isolated equilibrium points $\rho_1^e = \frac{d}{v_f}$ and $\rho_2^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$ when $I \in \Omega_3$. We show below that $\rho_1^e = \frac{d}{v_f}$ is exponentially stable with a region of attraction $\left[0, \frac{C_d}{v_f}\right]$ and $\rho_2^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$ is exponentially stable with a region of attraction $(\frac{C_d}{v_f}, \rho^j]$.

For $\rho(0) \in \left[0, \frac{C_d}{v_f}\right]$, we have $q_1 = d, q_2 = v_f\rho$, therefore $\dot{\rho} = -v_f\rho + d, \forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$, whose solution is

$$\rho = e^{-v_f t}\left(\rho(0) - \frac{d}{v_f}\right) + \frac{d}{v_f},$$

which implies that $\rho(t) \in \left[0, \frac{C_d}{v_f}\right], \forall t \geq 0$ and $\rho(t)$ converges exponentially fast to $\rho_1^e = \frac{d}{v_f}$.

Consider the equilibrium point ρ_2^e and choose the Lyapunov function

$$V(\rho) = \frac{(\rho - \rho_2^e)^2}{2},$$

then $\dot{V} = -(\rho - \rho_2^e)(q_2 - q_1)$. We show in Appendix C that

$$\dot{V} \leq -\alpha(\rho - \rho_2^e)^2,$$

where $\alpha = \min\{\frac{d-(1-\epsilon_0)C_d}{\rho_2^e - C_d/v_f}, w, (w - \tilde{w})\} > 0,\ \forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$ which implies exponential convergence to the equilibrium point $\rho_2^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$.

d) When $I \in \Omega_4$, q_1 and q_2 described by equation are plotted in Figure 82.

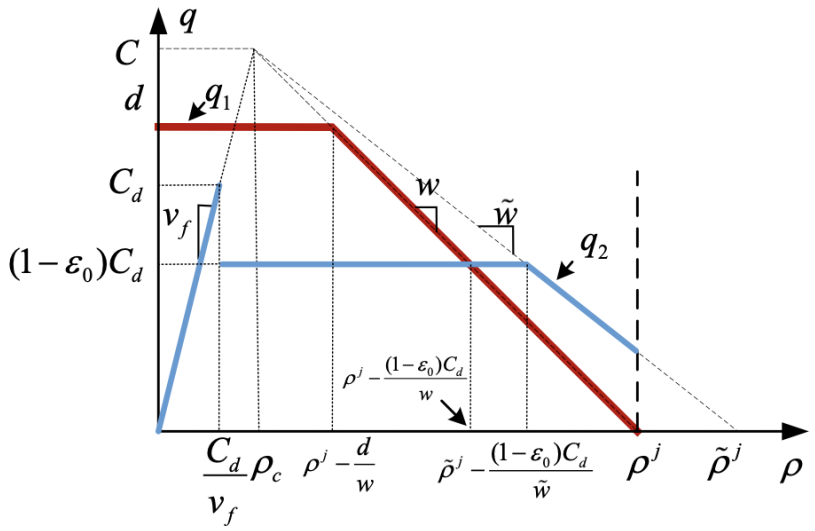


Figure 82. Fundamental Diagram for I ∈ Ω_4

It is clear that $q_1 = q_2$ when $\rho = \rho^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, which is a unique equilibrium when $I \in \Omega_4$. Choose the Lyapunov function

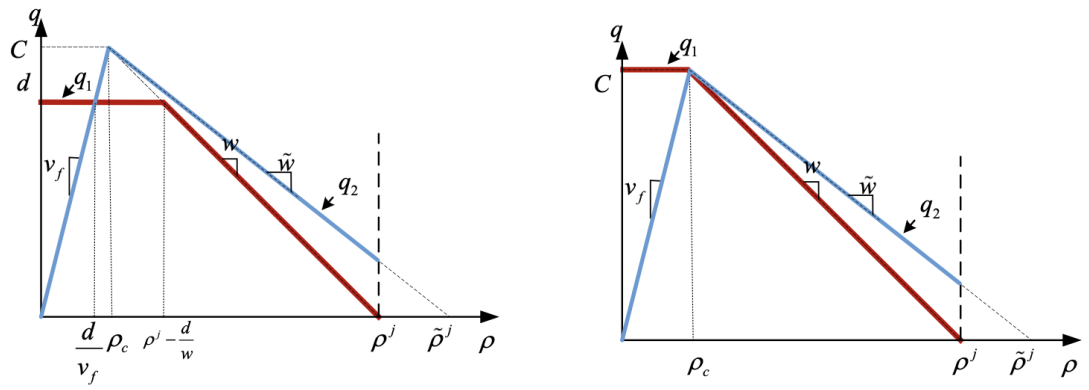
$$V(\rho) = \frac{(\rho - \rho^e)^2}{2},$$

then $\dot{V} = -(\rho - \rho^e)(q_2 - q_1)$. We show in Appendix C that

$$\dot{V} = -\alpha(\rho - \rho^e)^2,$$

where $\alpha = \min\{\frac{d-C_d}{\rho^e}, \frac{d-(1-\epsilon_0)C_d}{\rho^e - C_d/v_f}, w, (w - \tilde{w})\} > 0, \forall \rho \in [0, \rho^j]$, which implies exponential convergence to the equilibrium point $\rho^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, $\forall \rho(0) \in [0, \rho^j]$.

e) When $I \in \Omega_5$, q_1 and q_2 described by equation (41) are plotted in Figure 83.



(a) $d < C$ (b) $d \geq C$

Figure 83. Fundamental Diagram for I ∈ Ω_5

In this case it is clear that there is only one equilibrium point $\rho^e = \frac{\min\{d,C\}}{v_f}$, depending whether the demand $d < C$ or $d \geq C$. We choose the Lyapunov function

$$V(\rho) = \frac{(\rho - \rho^e)^2}{2}$$

and show in Appendix C that

$$\dot{V} = -\alpha(\rho - \rho^e)^2,$$

where $\alpha = \min\{v_f, \frac{(\tilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\} > 0$ if $d < C$ and $\alpha = \min\{v_f, (w - \tilde{w})\} > 0$ if $d \geq C$, $\forall \rho \in [0, \rho^j]$, which implies exponential convergence to the equilibrium point $\rho^e = \frac{\min\{d,C\}}{v_f}$, $\forall \rho(0) \in [0, \rho^j]$.

Stability of Traffic Flow in a Multi-Section Road Segment

The equilibrium points and their stability analysis of the single section CTM can be extended to the general N section case. Consider a road segment which is divided into N ($N \geq 2$) sections as in Figure 84. Without loss of generality, we assume that the geometry of all sections is identical and each section has unit length. In the single section case, we assume the density ρ to be the same along the section. We extend this to the case of multiple sections 1 to N where each section has its own density. The capacity of all sections remains the same constant C and the capacity at the outlet is C_d whereas the demand d appears at the entrance of section 1 as shown in Figure 84. It is well-known that the CTM in the multiple section case may include discontinuities in the values of densities when transitioning from one section to another. The control objective to be achieved via VSL, will require all section densities to converge to the same value in order to have smooth flow.

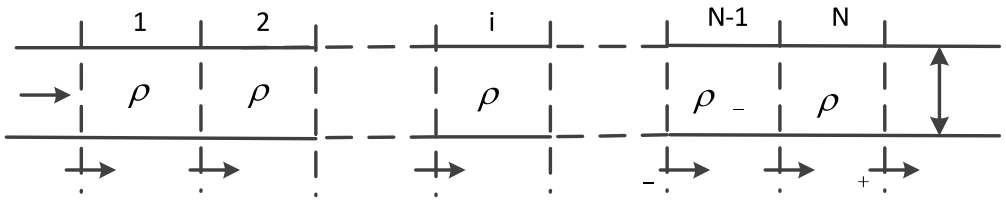


Figure 84. Multiple Section Road Network

Let $\rho = [\rho_1, \rho_2, \ldots, \rho_N]^T$ be the state vector of the traffic flow system, where ρ_i represents the density in section i. Section i can absorb the flow $\min\{C, w(\rho^j - \rho_i)\}$ from upstream and can generate the flow $\min\{v_f \rho_i, \tilde{w}(\tilde{\rho}^j - \rho_i)\}$ into the downstream section. Therefore, the dynamics of the vehicle densities in each section are formulated as:

$$\begin{aligned} \dot{\rho}_i &= q_i - q_{i+1}, 0 \le \rho_i(0) \le \rho^j, \text{for } i = 1,2,\ldots,N, \\ q_1 &= \min\{d, C, w(\rho^j - \rho_1)\}, \\ q_i &= \min\{v_f \rho_{i-1}, \tilde{w}(\tilde{\rho}^j - \rho_{i-1}), C, w(\rho^j - \rho_i)\}, i = 2,\ldots,N, \\ q_{N+1} &= \begin{cases} \min\{v_f \rho_N, \tilde{w}(\tilde{\rho}^j - \rho_N), (1 - \epsilon(\rho_N))C_d\} & \text{if } C_d < C \\ \min\{v_f \rho_N, \tilde{w}(\tilde{\rho}^j - \rho_N), C_d\} & \text{otherwise} \end{cases}, \end{aligned} \tag{42}$$

where

$$\epsilon(\rho_N) = \begin{cases} 0 & \text{if } 0 \le \rho_N \le \dfrac{C_d}{v_f} \\ \epsilon_0 & \text{otherwise} \end{cases}$$

and $0 < \epsilon_0 < 1$ denotes the level of capacity drop at the outlet of the Nth section. Since we assume that the capacities of all sections 1 to N have the same value C, the capacity drop can only happen at the outlet of section N, when $C_d < C$, which affects the value of q_{N+1}. We know that $\forall t \ge 0$, the density vector $\rho(t)$ belongs to the feasible set

$$S = \{\rho | 0 \le \rho_i \le \rho^j, \text{for } i = 1,2,\ldots,N\}.$$

Let $\rho^e = [\rho_1^e, \rho_2^e, \ldots, \rho_N^e]^T$ be the equilibrium vector of system (42) obtained by setting $\dot{\rho}_i = 0$, for $i = 1,2,\cdots,N$. Let q_i^e denote the value of q_i when $\rho = \rho^e$, then the equilibrium condition of system (42) is given by

$$q_1^e = q_2^e = \ldots = q_{N+1}^e, \tag{43}$$

due to $\dot{\rho}_i = q_i - q_{i+1} = 0$, for $i = 1,2,\cdots,N$.

Define the vector of initial condition $\rho(0) = [\rho_1(0), \rho_2(0), \ldots, \rho_N(0)]^T$ and the parameter vector $I = (C_d, C, d, \epsilon_0)$, whose partition sets are the same as in the case of a single section and are shown in Figure 78. Then the equilibrium states of (42) for all possible I in the sets Ω_1 to Ω_5 and corresponding stability properties are given by the following theorem.

Theorem 1.4. *Let* $\mathbf{1} = [1,1,\ldots,1]^T$ *be a vector with* N *elements each equal to 1. For constant but otherwise arbitrary demand* d*, we have the following results:*

a) *Let* $I \in \Omega_1$*. The equilibrium state of (42) is equal to* $\rho^e = \frac{d}{v_f} \times \mathbf{1}$ *and it is exponentially stable, i.e for all* $\rho(0) \in S$*,* $\rho(t)$ *converges exponentially fast to* $\rho^e = \frac{d}{v_f} \times \mathbf{1}$*.*

b) *Let* $I \in \Omega_2$*. System (42) has an isolated equilibrium state* $\rho^e = \frac{d}{v_f} \times \mathbf{1}$*, which is locally exponentially stable, and an infinite number of equilibrium states defined by the set*

$$S^e = \{\left(\rho^j - \frac{d}{w}\right) \times \mathbf{1}\} \cup \{\rho | \rho_i = \frac{d}{v_f}, i = 1,2,\ldots,N-1, \frac{C_d}{v_f} < \rho_N < \rho^j - \frac{d}{w}\}$$
$$\cup \left[\cup_{i=1}^{N-1} \{\rho | \frac{d}{v_f} \leq \rho_i < \rho^j - \frac{d}{w}, \rho_k = \frac{d}{v_f}, 1 \leq k < i, \rho_r = \rho^j - \frac{d}{w}, i < r \leq N\}\right].$$

All equilibrium states $\rho^e \in S^e$ *are stable in the sense that for any* $\mu > 0, \exists \eta > 0$*, such that* $\forall \rho(0)$ *that satisfy* $\| \rho(0) - \rho^e \| < \eta$*,* $\rho(t)$ *converges to a* $\bar{\rho}^e \in S^e$ *that satisfies* $\| \bar{\rho}^e - \rho^e \| < \mu$*. Furthermore,* $\forall \rho(0) \in \{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\ldots,N\}$*,* $\rho(t)$ *converges to* $\rho^e = \frac{d}{v_f} \times \mathbf{1}$ *exponentially fast, and* $\forall \rho(0) \notin \{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\ldots,N\}$*,* $\exists \rho^e \in \{\frac{d}{v_f} \times \mathbf{1}\} \cup S^e$*, such that* $\rho(t)$ *converges to* ρ^e *asymptotically with time.*

c) *Let* $I \in \Omega_3$*. System (42) has two isolated equilibrium states* $\rho^{e1} = \frac{d}{v_f} \times \mathbf{1}$ *and* $\rho^{e2} = \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$*, which are both locally exponentially stable. Furthermore,* $\forall \rho(0) \in \{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\ldots,N\}$*,* $\rho(t)$ *converges to* ρ^{e1} *exponentially fast and* $\forall \rho(0) \notin \{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\ldots,N\}$*,* $\rho(t)$ *converges to either* ρ^{e1} *or* ρ^{e2} *exponentially fast.*

d) *Let* $I \in \Omega_4$*. The equilibrium state (42) is equal to* $\rho^e = \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$ *and is exponentially stable, i.e for all* $\rho(0) \in S$*,* $\rho(t)$ *converges exponentially fast to* $\rho^e = \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$*.*

e) *Let* $I \in \Omega_5$*. The equilibrium state of (42) is equal to* $\rho^e = \frac{\min\{d,C\}}{v_f} \times \mathbf{1}$ *and is exponentially stable, i.e for all* $\rho(0) \in S$*,* $\rho(t)$ *converges exponentially fast to* $\rho^e = \frac{\min\{d,C\}}{v_f} \times \mathbf{1}$*.*

The proof of Theorem 1.4 is given in Appendix D.

The above stability properties show that depending on the situation classified by the operating scenarios Ω_1 to Ω_5 and initial density value in the section, the density will reach an equilibrium that is not always the one that corresponds to maximum flow rate. In fact, when $I \in \Omega_2$ there are an infinite number of equilibrium points and when $I \in \Omega_3$, there are two equilibrium points. One in the free flow region and one in the congested region depending on the initial density condition. The objective of feedback is to close the loop so that the system converges to a single equilibrium point for the density which also corresponds to the maximum possible flow

rate and speed. The feedback control variable is variable speed limit that provides speed commands to the upstream section in order to control the inflow to the section in a way that guarantees the maximum possible outflow from the downstream section. Such a design is presented in the next section.

VSL Control of the Cell Transmission Model under All Operating Conditions

VSL Control: Single Section

The stability analysis of the flow in Stability of Traffic Flow in a Single-Section Road Segment shows that if $C_d \geq C$, i.e., the downstream capacity is higher than the capacity of the section, i.e., $I \in \Omega_5$ then the density $\rho(t)$ converges exponentially fast to a unique equilibrium point $\frac{\min\{d,C\}}{v_f}$, which corresponds to the maximum possible flow. The steady state speed of flow in the section is v_f and the steady state section flow will be at the maximum possible value $q = q_1 = q_2 = \min\{d, C\}$ according to the model (40)-(41). In this case no control action is needed. When $C_d < C$ and $d < (1-\epsilon_0)C_d$, i.e., $I \in \Omega_1$, the demand is lower than the dropped capacity of the downstream segment and therefore the density converges exponentially fast to $\frac{d}{v_f}$ and the steady state flow speed and flow rate in the section will be v_f and d respectively. In this case, no control action is needed as the section operates at the maximum possible flow rate level dictated by the demand d. The problem arises when $C_d < C$ and $d \geq (1-\epsilon_0)C_d$. where we have the following control problem cases:

6. $(1-\epsilon_0)C_d = d < C_d < C$, i.e., $I \in \Omega_2$.
7. $(1-\epsilon_0)C_d < d \leq C_d < C$, i.e., $I \in \Omega_3$.
8. $C_d < d, C_d < C$, i.e., $I \in \Omega_4$.

In case (i) we showed in previous section that a maximum flow of $d = (1-\epsilon_0)C_d$ can be maintained at an infinite number of density equilibrium points specified by an isolated point and an equilibrium manifold, which include low and high density values with steady state speeds $v_{ss} \leq v_f$. In this case, the control objective is to maintain the maximum flow of $d = (1-\epsilon_0)C_d$ with a lowest possible density which in this case is $\frac{d}{v_f} = \frac{(1-\epsilon_0)C_d}{v_f}$ with free flow speed v_f.

In case (ii), we showed that we have two stable equilibrium points for density. One at low density which is equal to $\frac{d}{v_f}$ and one at high density equal to $\rho^j - \frac{(1-\epsilon_0)C_d}{v_f}$. In this case, maximum flow in the section corresponds to the density equilibrium point $\rho = \frac{d}{v_f}$ therefore the control objective is to choose the VSL in a way that the density converges to $\frac{d}{v_f}$ for all possible initial density conditions.

In case (iii), there is only one equilibrium point for density which is in the high density region and corresponds to the steady state flow of $(1-\epsilon_0)C_d$. In this case, the maximum possible

flow is C_d and corresponds to the density of $\frac{C_d}{v_f}$. However, the convergence of ρ to $\frac{C_d}{v_f}$ does not guarantee that q_1 and q_2 converge to C_d due to the capacity drop. From equation (41) and Figure 82, we know that q_2 is a function of ρ. For $\rho \in \left[0, \frac{C_d}{v_f}\right]$, $q_2 = v_f \rho$, and for $\rho \in (\frac{C_d}{v_f}, \tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\tilde{w}}]$, $q_2 = (1-\epsilon_0)C_d$. Therefore, we have

$$\lim_{\rho \to \left(\frac{C_d}{v_f}\right)_-} q_2(\rho) = \lim_{\rho \to \frac{C_d}{v_f}} v_f \rho = C_d \text{ and } \lim_{\rho \to \left(\frac{C_d}{v_f}\right)_+} q_2(\rho) = \lim_{\rho \to \frac{C_d}{v_f}} (1-\epsilon_0)C_d = (1-\epsilon_0)C_d,$$

i.e., if ρ converges to $\frac{C_d}{v_f}$ from the left side, then q_2 converges to the maximum value C_d. However, if ρ converges to $\frac{C_d}{v_f}$ from the right side, q_2 converges to $(1-\epsilon_0)C_d$. Therefore, the control objective in this case is to choose the VSL so that $\rho(t)$ satisfies the following conditions: $\exists t_0 > 0$, such that $\forall t \geq t_0,\ \rho(t) \leq \frac{C_d}{v_f}$ and $\lim_{t\to\infty} \rho(t) = \frac{C_d}{v_f}$.

Therefore for all cases (i), (ii) and (iii), the control objective is to choose the VSL control so that $\rho(t)$ converges to the desired equilibrium point $\frac{\min\{d, C_d\}}{v_f}$, and the flow rate q_1 and q_2 converge to the maximum possible level which is equal to $\min\{d, C_d\}$.

A reasonable control action is to use VSL control to restrict the incoming flow q_1 to the level that is within the capacity constraints of the section at the bottleneck so that the density and flow rate converge to the desired possible values.

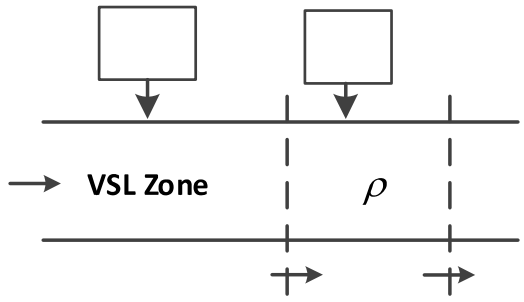


Figure 85. Road Section with VSL Control

As shown in Figure 85, we apply the VSL command v in the upstream segment of the section under consideration, which is referred to as the VSL zone. All vehicles are asked to follow the speed limit v in the VSL zone and follow the free flow speed limit v_f inside the section. Decreasing the speed limit leads to lower flow q_1 from the VSL zone to the section as shown in Figure 85. Figure 86 shows how the changing of the speed limit v can control the flow rate q_1 entering the section through a nonlinear relationship. Suppose the VSL zone has similar characteristics as the road section under consideration. If the VSL command is set to $v < v_f$, the fundamental diagram of the VSL zone is distorted such that the parameters $\rho^j, w, \tilde{w}$ remain

unchanged, while the maximum possible flow is decreased to $\frac{vw\rho^j}{v+w}$, as shown in Figure 86, obtained by simple geometric considerations (H.-Y. Jin and Jin 2015; Hadiuzzaman and Qiu 2013; Csikós and Kulcsár 2017).

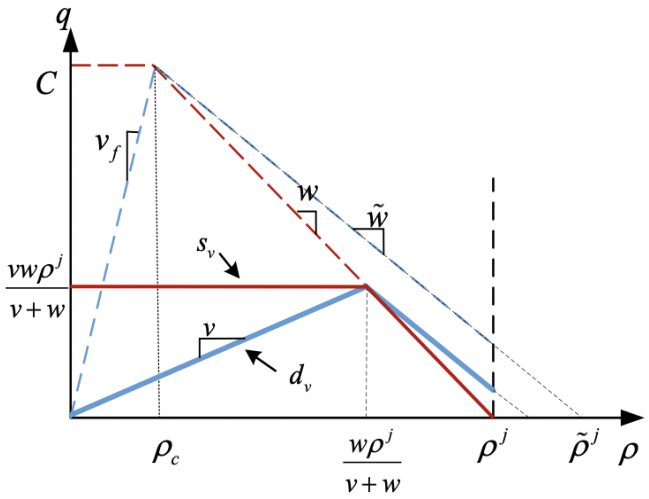


Figure 86. Fundamental Diagram of the VSL Zone

In Figure 86, the red line s_v denotes the flow rate that the VSL zone can absorb from upstream under different densities in the VSL zone and the blue line d_v denotes the flow rate that the VSL zone sends to the section under consideration. However, since the single section model does not include the density in the VSL zone, the flow into the road section from the VSL zone is assumed to be $\min\{d, \frac{vw\rho^j}{v+w}\}$, where $\frac{vw\rho^j}{v+w}$ is the maximum possible flow in the VSL zone under speed limit v. Then the density ρ in the section is given by the following equation:

$$\begin{aligned} \dot{\rho} &= q_1 - q_2,\ 0 \le \rho(0) \le \rho^j, \\ q_1 &= \min\{d, \frac{vw\rho^j}{v+w}, C, w(\rho^j - \rho)\}, \\ q_2 &= \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), (1-\epsilon(\rho))C_d\}. \end{aligned} \tag{44}$$

We design a VSL controller to overcome capacity drop and achieve the control objectives in all cases, by first considering the most complicated case $I \in \Omega_4$, in which $d > C_d$. Since in equation (44), $\frac{vw\rho^j}{v+w}$ is the only term in q_1 that depends on v, we derive the VSL controller using feedback linearization under the assumption that $q_1 = \frac{vw\rho^j}{v+w}$. Then we show in Theorem Theorem 1.5 that, for the general equation where $q_1 = \min\{d, \frac{vw\rho^j}{v+w}, C, w(\rho^j - \rho)\}$, the derived controller can still guarantee that ρ converges to $\frac{C_d}{v_f}$ and q_1, q_2 converge to the maximum value C_d. Furthermore, we also show in Theorem Theorem 1.5 below that, when $I \in \bigcup_{i=1}^{3} \Omega_i$, i.e., $d \le$

C_d, the same controller guarantees the convergence of ρ to the desired equilibrium point $\frac{d}{v_f}$ and the convergence of q_1, q_2 to the maximum level d.

As discussed above, when $I \in \Omega_4$, the desired equilibrium point is $\rho^e = \frac{C_d}{v_f}$. Define the error state $x = \rho - \frac{C_d}{v_f}$ and recall that the control objective is to force ρ to converge to $\frac{C_d}{v_f}$, i.e., x converge to 0 from the left side ($\rho \leq \frac{C_d}{v_f}$). If $x(0) \leq 0$, that is $\rho(0) \leq \frac{C_d}{v_f}$, we choose v so that

$$q_1 = q_2 - \lambda x, \tag{45}$$

where $\lambda > 0$ is a design constant to be selected. Thus, we have

$$\dot{x} = \dot{\rho} = q_1 - q_2 = -\lambda x, \tag{46}$$

which implies that $\forall x(0) \leq 0$ and $t \geq 0$, $x(t) \leq 0$ and x converges to 0 exponentially fast. Since we assume that $q_1 = \frac{vw\rho^j}{v+w}$, solving equation (45) for v, we have,

$$v = \frac{w(q_2 - \lambda x)}{w\rho^j - (q_2 - \lambda x)}, \tag{47}$$

whose denominator is guaranteed to be greater than 0 as we show in detail in the proof of Theorem Theorem 1.5.

If $x(0) > 0$, i.e., $\rho(0) > \frac{C_d}{v_f}$ we choose v such that

$$q_1 = q_2 - \lambda(x + \delta_1), \tag{48}$$

where $\delta_1 > 0$ is a design constant. Then we have $\forall x(0) > 0$

$$\dot{x} = \dot{\rho} = q_1 - q_2 = -\lambda(x + \delta_1).$$

Thus x will decrease exponentially toward the value $-\delta_1 < 0$. At some finite time $t = t_0 > 0$, $x(t_0) = -\delta_2$, i.e., $\rho(t_0) = \frac{C_d}{v_f} - \delta_2$, where $0 < \delta_2 < \min\{\delta_1, \frac{C_d}{v_f}\}$, thus $\rho(t_0)$ is in the region of (45), (47). At the time instant $t = t_0$, we have $x(t) \leq 0$ and controller (47) is switched on which guarantees as shown above that $x(t)$ will converge to zero exponentially fast. Assuming that $q_1 = \frac{vw\rho^j}{v+w}$ and solving (48) for v, we have

$$v = \frac{w\big(q_2 - \lambda(x + \delta_1)\big)}{w\rho^j - \big(q_2 - \lambda(x + \delta_1)\big)}.$$

The use of the design constant δ_1 is to reduce the incoming flow via VSL so that the density of the section reduces to be within the set $\left[0, \frac{C_d}{v_f}\right]$, which guarantees convergence to the equilibrium point which corresponds to maximum flow and speed. The choice of δ_1 will depend

on how aggressively we want the density to move to the "good" free speed region. Using the above VSL controller derivation and assuming that the speed is not allowed to go below zero or exceed the speed limit v_f, the following equations summarize the VSL controller for the section under the assumption that $q_1 = \frac{vw\rho^j}{v+w}$, which we will relax subsequently. When $I \in \bigcup_{i=1}^{4} \Omega_i$, the VSL control is generated as follows:

$$
\begin{aligned}
\bar{v}_1 &= \frac{w[q_2 - \lambda(x+\delta_1)]}{w\rho^j - [q_2 - \lambda(x+\delta_1)]}, \\
\bar{v}_2 &= \frac{w(q_2 - \lambda x)}{w\rho^j - (q_2 - \lambda x)}, \\
v_i &= \text{med}\{0, \bar{v}_i, v_f\}, i = 1,2, \\
v &= \begin{cases} v_1 & \text{if } \rho(0) > \frac{C_d}{v_f} \text{ and } \rho(t) > \frac{C_d}{v_f} - \delta_2 \\ v_2 & \text{if } \rho(0) > \frac{C_d}{v_f} \text{ and } \rho(t) = \frac{C_d}{v_f} - \delta_2 \\ v_2 & \text{if } \rho(0) \le \frac{C_d}{v_f} \text{ and } \rho(t) \le \frac{C_d}{v_f} \end{cases},
\end{aligned}
\tag{49}
$$

where $x = \rho - \frac{C_d}{v_f}$, and

$$
\delta_1 > 0, 0 < \delta_2 < \min\{\delta_1, \frac{C_d}{v_f}\},\ 0 < \lambda < \frac{v_f w \rho^j}{C_d}
$$

are design constants and med$\{\cdot\}$ denotes the median of the numbers, which indicates that the VSL command saturates at the upper bound v_f and the lower bound 0. The upper bound $\frac{v_f w \rho^j}{C_d}$ of λ guarantees that the denominator of v is not 0, which we will show in the proof of Theorem 1.5. The shape of the function v as it varies with ρ is shown in Figure 87.

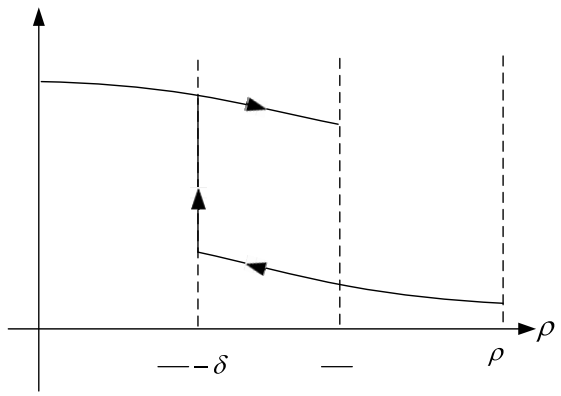


Figure 87. Switching Logic of VSL Controller

For $I \in \Omega_5$, the VSL control is

$$v = v_f. \tag{50}$$

In Theorem 1.5 below, we show that the above controller also works for any value of $q_1 = \min\{d, \frac{vw\rho^j}{v+w}, C, w(\rho^j - \rho)\}$, and guarantees the exponential convergence of the density to the desired equilibrium point and the exponential convergence of the flow rate to the maximum possible value of $q_1 = q_2 = C_d$. Furthermore, when $I \in \bigcup_{i=1}^{3} \Omega_i$, i.e., $d \leq C_d$, controller (49) guarantees the exponential convergence of ρ to the desired equilibrium point $\frac{d}{v_f}$ and the convergence of q_1, q_2 to the maximum level d.

Theorem 1.5. *For $q_1 = \min\{d, \frac{vw\rho^j}{v+w}, C, w(\rho^j - \rho)\}$, we have the following:*

a) *Let $I \in \bigcup_{i=1}^{4} \Omega_i$, i.e., $C_d < C$, and consider the VSL controller (49). The closed-loop system (44), (49) has a unique equilibrium point $\rho^e = \frac{\min\{d, C_d\}}{v_f}$. In addition, $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$, $\rho(t)$ converges to ρ^e exponentially fast and $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$, $\rho(t)$ decreases to $\frac{C_d}{v_f} - \delta_2$ exponentially fast which brings it to the region where $\rho(t)$ converges to ρ^e exponentially fast. The flow rate and speed converge to the desired values of $\min\{d, C_d\}$ and v_f respectively with the same rate.*

b) *Let $I \in \Omega_5$, i.e., $C_d \geq C$, and consider the VSL controller (50). System (44), (50) has a unique equilibrium point $\rho^e = \frac{\min\{d, C\}}{v_f}$. In addition, $\forall \rho(0) \in \left[0, \rho^j\right]$, $\rho(t)$ converges to ρ^e exponentially fast. The flow rate and speed converge exponentially fast to the desired values of $\min\{d, C\}$ and v_f respectively.*

The proof of Theorem 1.5 is given in Appendix E. Theorem 1.5 shows that the VSL controller guarantees that for all cases $I \in \bigcup_{i=1}^{5} \Omega_i$, the density, flow rate and flow speed converge exponentially fast to unique values that correspond to maximum possible flow through the section for all initial density conditions within the set $\left[0, \rho^j\right]$. Theorem 1.5 shows in an analytically rigorous manner that VSL control can stabilize the flow in the section and force it to converge to the maximum possible flow under any situation. This maximum flow depends on the characteristics and relationships between demand d and capacities C, C_d as well as capacity drop level ϵ_0. It is also clear from the analysis of the open-loop system that without the VSL control the flow can reach steady states that do not correspond to maximum possible flow.

From equation (49), we can see that the logic of the VSL controller is to deactivate the capacity drop with v_1 by suppressing the inflow sufficiently and then force the system state to converge to the desired equilibrium point with v_2. This logic and the feedback linearization technique can always be used to design a VSL controller if different capacity drop models such as those presented in (Kontorinaki et al. 2016) are included in the CTM.

VSL Control: Multiple Sections

The analysis in Stability of Traffic Flow in a Multi-Section Road Segment shows that the stability properties of the open-loop N-section system are similar to those of the single-section system. For the cases $I \in \Omega_1$ and $I \in \Omega_5$, $\rho(t)$ converges exponentially fast to the unique equilibrium state $\rho^e = \frac{d}{v_f} \times \mathbf{1}$ and $\rho^e = \frac{\min\{d,C\}}{v_f} \times \mathbf{1}$ respectively, which corresponds to the maximum possible flow rate. In these two cases no control action is needed.

When $I \in \Omega_2 \cup \Omega_3$, the control objective is to stabilize the system at the equilibrium state $\rho^e = \frac{d}{v_f} \times \mathbf{1}$, at which the maximum possible flow rate d is achieved and the densities in each section are stabilized at the lowest possible value whereas the speed of flow converges to the free flow speed v_f.

When $I \in \Omega_4$, the maximum possible flow rate is C_d, which corresponds to the equilibrium state $\rho^e = \frac{C_d}{v_f} \times \mathbf{1}$. From equation (42), we know that due to capacity drop

$$\lim_{\rho_N \to \left(\frac{C_d}{v_f}\right)_-} q_{N+1}(\rho_N) = \lim_{\rho_N \to \frac{C_d}{v_f}} v_f \rho_N = C_d$$

and

$$\lim_{\rho_N \to \left(\frac{C_d}{v_f}\right)_+} q_{N+1}(\rho_N) = \lim_{\rho_N \to \frac{C_d}{v_f}} (1-\epsilon_0) C_d = (1-\epsilon_0) C_d.$$

Therefore, in this case, in order to achieve the maximum possible flow rate C_d, we want to choose the VSL control so that there exits $t_0 \geq 0$ such that $\forall t \geq t_0$, $\rho_N(t) \leq \frac{C_d}{v_f}$ and $\rho_i(t)$ converges to $\frac{C_d}{v_f}$, for $i = 1,2,\ldots,N$. Furthermore, we want to achieve a steady state flow speed v_f in all sections.

Similar to the single section case, the VSL controller is applied to the N-section road segment as shown in Figure 88.

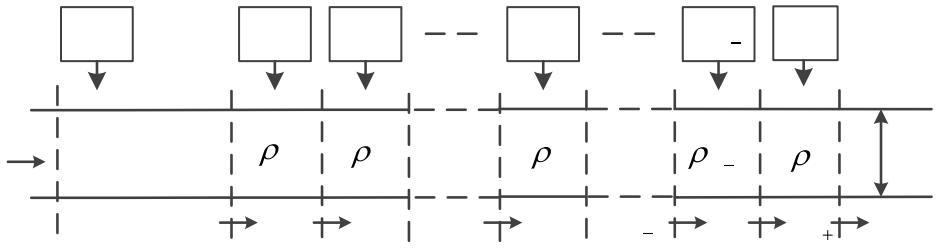


Figure 88. VSL Controlled Road Segment

All vehicles in the upstream segment of section 1 are asked to follow the VSL command v_0 and all vehicles in section i follow the VSL command v_i, for $i = 1,2,\dots,N-1$. The speed limit in section N is set to the constant free flow speed v_f.

If the speed limit of section i is set to be $v_i \le v_f$, $i = 1,2,\dots,N-1$, then the fundamental diagram of section i is distorted as shown in Figure 89.

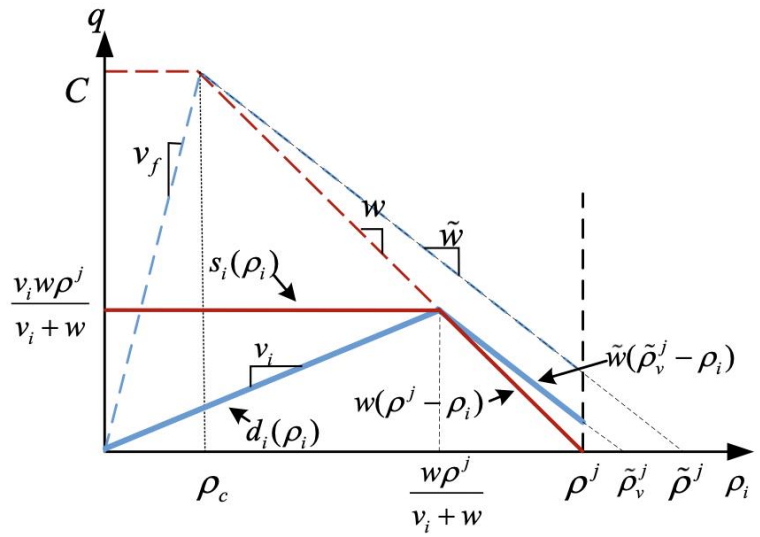


Figure 89. Fundamental Diagram of Section i

In Figure 89, $s_i(\rho_i)$ denotes the ability of section i to absorb traffic flow from section $i-1$. We have $s_i(\rho_i) = \min\{\frac{v_i w \rho^j}{v_i + w}, w(\rho^j - \rho_i)\}, i = 1,2,\dots,N-1$. $d_i(\rho_i)$ denotes the traffic flow generated by section i to go into section $i+1$. We have $d_i(\rho_i) = \min\{v_i \rho_i, \frac{v_i w \rho^j}{v_i + w}\}, i = 1,2,\dots N-1$. Therefore, we have

$$q_i = \min\{d_{i-1}(\rho_{i-1}), s_i(\rho_i)\} = \min\{v_{i-1}\rho_{i-1}, \frac{v_{i-1} w \rho^j}{v_{i-1} + w}, \frac{v_i w \rho^j}{v_i + w}, w(\rho^j - \rho_i)\}, i = 2,\dots,N-1.$$

For the road segment upstream section 1, i.e., the segment with speed limit v_0, whose density is not included in system (42), we assume the flow rate generated by this segment to be $d_0 = \min\{d, \frac{v_0 w \rho^j}{v_0 + w}\}$, which is independent of the density in the section with speed limit v_0, therefore

$$q_1 = \min\{d_0, s_1(\rho_1)\} = \min\{d, \frac{v_0 w \rho^j}{v_0 + w}, \frac{v_1 w \rho^j}{v_1 + w}, w(\rho^j - \rho_1)\}.$$

The speed limit in section N is constant v_f, therefore section N can absorb a flow of $s_N(\rho_N) = \min\{C, w(\rho^j - \rho_N)\}$, therefore

$$q_N = \min\{d_{N-1}(\rho_{N-1}), s_N(\rho_N)\} = \min\{v_{N-1}\rho_{N-1}, \frac{v_{N-1}w\rho^j}{v_{N-1}+w}, C, w(\rho^j - \rho_N)\}.$$

For the sake of simplicity, we omit the term $\tilde{w}(\tilde{\rho}_v^j - \rho_i)$ from $d_i(\rho_i)$, where $\tilde{\rho}_v^j$ is $\tilde{\rho}^j$ distorted by the VSL. As shown n Figure 89, for $i = 1, \dots, N-1$, if the outflow $q_{i+1} = \tilde{w}(\tilde{\rho}_v^j - \rho_i)$, then the inflow $q_i \le s_i(\rho_i) = w(\rho^j - \rho_i) < q_{i+1}$ will force ρ_i to decrease until $q_{i+1} \ne \tilde{w}(\tilde{\rho}_v^j - \rho_i)$. Therefore, this simplification does not affect the results. The system model with VSL control inputs can be formulated as follows:

$$\begin{aligned}
\dot{\rho}_i &= q_i - q_{i+1}, 0 \le \rho_i(0) \le \rho^j, \text{for } i = 1,2,\dots,N, \\
q_1 &= \min\{d, \frac{v_0 w\rho^j}{v_0+w}, \frac{v_1 w\rho^j}{v_1+w}, w(\rho^j - \rho_1)\}, \\
q_i &= \min\{v_{i-1}\rho_{i-1}, \frac{v_{i-1}w\rho^j}{v_{i-1}+w}, \frac{v_i w\rho^j}{v_i+w}, w(\rho^j - \rho_i)\}, i = 2,3,\dots,N-1, \\
q_N &= \min\{v_{N-1}\rho_{N-1}, \frac{v_{N-1}w\rho^j}{v_{N-1}+w}, C, w(\rho^j - \rho_N)\}, \\
q_{N+1} &= \min\{v_f\rho_N, (1-\epsilon(\rho_N))C_d, \tilde{w}(\tilde{\rho}^j - \rho_N)\}.
\end{aligned} \tag{51}$$

Similar to the single section system, the objective is to design a VSL controller that can overcome the capacity drop and achieve the control objectives in all cases. We derive the VSL controller using feedback linearization for the case of $I \in \Omega_4$ then show in Theorem 1.6 that the controller also works for all other scenarios.

When $I \in \Omega_4$, i.e., $d > C_d$, we need to decrease v_0 to suppress q_1 so that the flow from upstream can be handled by the downstream capacity C_d. We start by assuming that $q_1 = \frac{v_0 w\rho^j}{v_0+w}$, which is the only term in the equation of q_1 that depends on v_0 and then show that the VSL controller works for all values of q_1. Furthermore, in this case, the desired equilibrium density is $\rho_i^e = \frac{C_d}{v_f}$ for $i = 1,2,\dots,N$ and the equilibrium flow speed and flow rate are $v_i^e = v_f$ and $q_{i+1}^e = v_i^e\rho_i^e = C_d$ respectively for $i = 1,2,\dots,N-1$. Therefore, we initially assume that $q_i = v_{i-1}\rho_{i-1}$ for $i = 2,3,\dots,N$, which we relax in Theorem 1.6 below.

Let $x = [x_1, x_2, \dots, x_N]^T$, where $x_i = \rho_i - \frac{C_d}{v_f}, i = 1,2,\dots,N$. If $x_N(0) \le 0$, i.e., $\rho_N(0) \le \frac{C_d}{v_f}$, we choose $v = [v_0, v_1, \dots, v_{N-1}]^T$, such that

$$q_i = q_{i+1} - \lambda_{i-1}x_i, i = 1,2,..,N, \tag{52}$$

where $\lambda_i > 0, i = 0,1,\dots,N-1$ are design constants. Thus we have

$$\dot{x}_i = \dot{\rho}_i = q_i - q_{i+1} = -\lambda_{i-1}x_i, i = 1,2,\dots,N,$$

which implies that $x_i(t)$ converges to 0 exponentially fast and $x_N(0) \leq 0, \forall t \geq 0$. Since we assume that $q_1 = \frac{v_0 w \rho^j}{v_0 + w}$ and $q_i = v_{i-1}\rho_{i-1}$ for $i = 2, \ldots, N$, solving (52) for v gives

$$\begin{aligned} v_0 &= \frac{(q_2 - \lambda_0 x_1)w}{w\rho^j - q_2 + \lambda_0 x_1}, \\ v_i &= \frac{q_{i+2} - \lambda_i x_{i+1}}{\rho_i}, i = 1,2,\ldots,N-1. \end{aligned} \tag{53}$$

If $x_N(0) > 0$, i.e., $\rho_N(0) > \frac{C_d}{v_f}$, we choose v such that

$$\begin{aligned} q_i &= q_{i+1} - \lambda_{i-1} x_i, i = 1,2,..,N-1, \\ q_N &= q_{N+1} - \lambda_{N-1}(x_N + \delta_1), \end{aligned} \tag{54}$$

where $\delta_1 > 0$ is a design constant. Then we have

$$\begin{aligned} \dot{x}_i &= \dot{\rho}_i = q_i - q_{i+1} = -\lambda_{i-1} x_i, i = 1,2,\ldots,N-1, \\ \dot{x}_N &= \dot{\rho}_N = q_N - q_{N+1} = -\lambda_{N-1}(x_N + \delta_1), \end{aligned}$$

which implies that $\forall x_N(0) > 0$, $x_N(t)$ will decrease exponentially toward $-\delta_1 < 0$. Therefore there exists $t_0 > 0$, such that $x_N(t_0) = -\delta_2$, i.e., $\rho_N(t_0) = \frac{C_d}{v_f} - \delta_2$, where $0 < \delta_2 < \min\{\delta_1, \frac{C_d}{v_f}\}$, which is in the region of (52)-(53). At $t = t_0$, we have $x_N(0) < 0$ and the controller (53) is switched on, in which case $x(t)$ converges to 0 exponentially fast as shown above. Solving (54) for v, we have

$$\begin{aligned} v_0 &= \frac{(q_2 - \lambda_0 x_1)w}{w\rho^j - q_2 + \lambda_0 x_1}, \\ v_i &= \frac{q_{i+2} - \lambda_i x_{i+1}}{\rho_i}, i = 1,2,\ldots,N-2, \\ v_{N-1} &= \frac{q_{N+1} - \lambda_{N-1}(x_N + \delta_1)}{\rho_{N-1}}. \end{aligned} \tag{55}$$

Using the above VSL controller and assuming that the speed is not allowed to go below zero or exceed the speed limit v_f, the following equations summarize the VSL controller for the N-section road system under the assumption that $q_1 = \frac{v_0 w \rho^j}{v_0 + w}$ and $q_i = v_{i-1}\rho_{i-1}, i = 2,3,\ldots,N$,

which we will relax subsequently. For all $I \in \bigcup_{i=1}^{4} \Omega_i$, the VSL commands are generated as follows:

$$
\begin{aligned}
\bar{v}_0 &= \frac{(q_2 - \lambda_0 x_1)w}{w\rho^j - q_2 + \lambda_0 x_1}, \\
\bar{v}_i &= \begin{cases} \dfrac{q_{i+2} - \lambda_i x_{i+1}}{\rho_i} & \rho_i > 0 \\ v_f & \rho_i = 0 \end{cases}, i = 1,2,\dots,N-2, \\
\bar{v}_{N-1,1} &= \frac{q_{N+1} - \lambda_{N-1}(x_N + \delta_1)}{\rho_{N-1}}, \\
\bar{v}_{N-1,2} &= \begin{cases} \dfrac{q_{N+1} - \lambda_{N-1} x_N}{\rho_{N-1}} & \rho_{N-1} > 0 \\ v_f & \rho_{N-1} = 0 \end{cases}, \\
\bar{v}_{N-1} &= \begin{cases} \bar{v}_{N-1,1} & \text{if } \rho_N(0) > \dfrac{C_d}{v_f} \text{ and } \rho_N(t) > \dfrac{C_d}{v_f} - \delta_2 \\ \bar{v}_{N-1,2} & \text{if } \rho_N(0) > \dfrac{C_d}{v_f} \text{ and } \rho_N(t) = \dfrac{C_d}{v_f} - \delta_2 \\ \bar{v}_{N-1,2} & \text{if } \rho_N(0) \le \dfrac{C_d}{v_f} \text{ and } \rho_N(t) \le \dfrac{C_d}{v_f} \end{cases}, \\
v_i &= \text{med}\{0, \bar{v}_i, v_f\}, i = 0,1,\dots,N-1,
\end{aligned}
\tag{56}
$$

where $\delta_1 > 0, 0 < \delta_2 < \min\{\delta_1, \frac{C_d}{v_f}\}$, $\lambda_i > v_f$, for $i = 1,2\dots,N-1$, $0 < \lambda_0 < \frac{v_f w \rho^j}{C_d}$. In controller (56), $\lambda_0 < \frac{v_f w \rho^j}{C_d}$ guarantees that the denominator of v_0 is always greater than 0. $\lambda_i > v_f$, for $i = 1,2\dots,N-1$ guarantees the exponential convergence of the density states, which we will show in the proof of Theorem 1.6 below. The switching logic of v_{N-1} is similar to that of v shown in Figure 87 for the single section case. For $I \in \Omega_5$, the VSL command is

$$v_i = v_f, i = 0,1,\dots,N-1. \tag{57}$$

Similar to the single section case, we can show that v_0 is well-defined as its denominator is always greater than 0. For $i = 1,2,\dots,N-1$, v_i is also well-defined by setting $v_i = v_f$ when its denominator is equal to 0.

Theorem 1.6. *We consider the traffic flow model described by (51) with the VSL controller (56):*

a) *Let $I \in \bigcup_{i=1}^{4} \Omega_i$, i.e., $C_d < C$. The closed-loop system has a unique equilibrium state $\rho^e = \frac{\min\{d, C_d\}}{v_f} \times \mathbf{1}$. In addition, $\forall \rho(0) \in \{\rho | 0 \le \rho_N \le \frac{C_d}{v_f}\}$, the density vector $\rho(t)$ converges to ρ^e exponentially fast and $\forall \rho(0) \in \{\rho | \frac{C_d}{v_f}\} < \rho_N \le \rho^j$, $\rho_N(t)$ decreases to $\frac{C_d}{v_f} - \delta_2$ exponentially fast, which brings it to the region where the density vector $\rho(t)$ converges to ρ^e exponentially fast. Furthermore, the flow rates $q_i, i = 1,2,\dots,N+1$ and flow*

speeds $v_i, i = 0,1,\dots,N-1$ converge to $\min\{d, C_d\}$ and v_f respectively which is the state which corresponds to the maximum possible flow.

b) *Let $I \in \Omega_5$, i.e., $C_d \geq C$. The closed-loop system has a unique equilibrium state $\rho^e = \frac{\min\{d,C\}}{v_f} \times \mathbf{1}$. In addition, $\forall \rho(0) \in S$, the density vector ρ converges exponentially fast to ρ^e. Furthermore, the flow rates and flow speeds converge exponentially fast to $\min\{d, C\}$ and v_f respectively, achieving the maximum possible flow at steady state.*

The proof of Theorem 1.6 is presented in Appendix F. Theorem 1.6 shows that the VSL controller guarantees that for all cases $I \in \bigcup_{i=1}^{5} \Omega_i$, the steady state densities, flow rates and speeds of flow are stabilized at the desired values which correspond to the maximum flow rate through the road segment while achieving homogeneous density distribution.

We should note that in Theorem 1.6, the design of the VSL controller and the stability analysis of the closed-loop system are performed under the assumption that we have perfect knowledge of system parameters and accurate measurement of the density vector ρ. However when $I \in \Omega_4$, since the desired equilibrium point of the closed-loop system (51), i.e., $\rho^e = \frac{C_d}{v_f} \times \mathbf{1}$ lies exactly on the discontinuity plane of the fundamental diagram, which is $\{\rho | \rho_N = \frac{C_d}{v_f}\}$, when $\rho(t) = \frac{C_d}{v_f} \times \mathbf{1}$ at steady state, any disturbance in model parameters or measurement noise may push the density in section N to $\rho_N > \frac{C_d}{v_f}$, which may lead to temporary capacity drop which the controller tries to correct leading to a possible oscillation around the desired equilibrium point. Even though such oscillations may not have any significant impact in an actual traffic situation, the proposed controller can be easily modified to avoid such oscillatory response. This is achieved by setting the desired equilibrium point to be $\rho^e = \left(\frac{C_d}{v_f} - \sigma\right) \times \mathbf{1}$, where $\sigma > 0$, in order to provide a margin between ρ^e and the discontinuity at $\rho_N = \frac{C_d}{v_f}$. Thus in (56), $x_i = \rho_i - \left(\frac{C_d}{v_f} - \sigma\right)$. With sufficiently large feedback gains $\lambda_0, \dots, \lambda_N$, the controller is able to stabilize the density state ρ at a point that is arbitrarily close to $\rho^e = \left(\frac{C_d}{v_f} - \sigma\right) \times \mathbf{1}$, therefore avoid the capacity drop. We will demonstrate this with numerical simulations in Numerical Experiments. Thus, although the controller (56) is designed for accurate system model, it can be robust with respect to system disturbance with simple modification. How to modify the controller of this paper to be robust with respect to a wide range of uncertainties is currently under investigation and it is outside the scope of this paper. However the ideal properties of the controller of this paper form the basis for comparison of any other controller under less ideal situations and for this reason it has its own merit.

Numerical Experiments

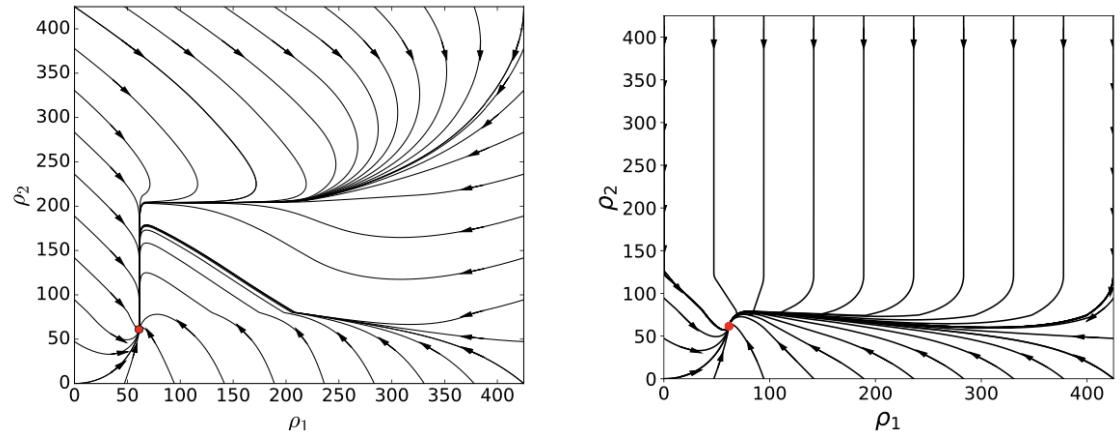
In this section, we use numerical simulations to demonstrate the analytical results of the previous sections, for both open-loop and closed-loop systems. The simulations are performed on a two-section road network, whose parameters are: $C = 6500$ veh/h, $w = 20$ mi/h, $\rho^j =$

425 veh/mi, $v_f = 65$ mi/h, $\tilde{w} = 10$ mi/h, $\tilde{\rho}^j = 750$ veh/mi, $\rho_c = 100$ veh/mi. When I belongs to Ω_1 to Ω_4, we set the downstream capacity $C_d = 5200$ veh/h, which is less than C, and $\epsilon_0 = 0.15$. When I belongs to Ω_5, we set $C_d = 7000$ veh/h, which is greater than C. The upstream demand d is set to be 4000 veh/h, 4420 veh/h, 5000 veh/h, 6000 veh/h and 6000 veh/h for the cases of I in Ω_1 to Ω_5 respectively. We apply controller (56),(57) to the two-section system with the following design constants: $\lambda_1 = \lambda_2 = 70$ mi/h, $\delta_1 = 20$ veh/mi, $\delta_2 = 5$ veh/mi. Among the abbreviated units we used above, "veh" stands for number of vehicles, "mi" stands for miles and "h" stands for hours.

Figure 90-Figure 94 show the phase portraits of the two-section open-loop and closed-loop systems when I belongs to Ω_1 to Ω_5. When $I \in \Omega_1$, all the density state trajectories of the open-loop system, shown in Figure 90 (a) converge to the unique equilibrium state $\rho^e = \left(\frac{d}{v_f}, \frac{d}{v_f}\right) = (61.5,61.5)$, indicated by the red dot. In Figure 90 (b), all density state trajectories of the closed-loop system converge to the same equilibrium state as in the open-loop case as expected from the analysis. When $I \in \Omega_2$, all density state trajectories of the open-loop system shown in Figure 91 (a) converge to the isolated equilibrium state $\rho^e = \left(\frac{d}{v_f}, \frac{d}{v_f}\right) = (68,68)$, indicated by the red dot, or to the equilibrium set

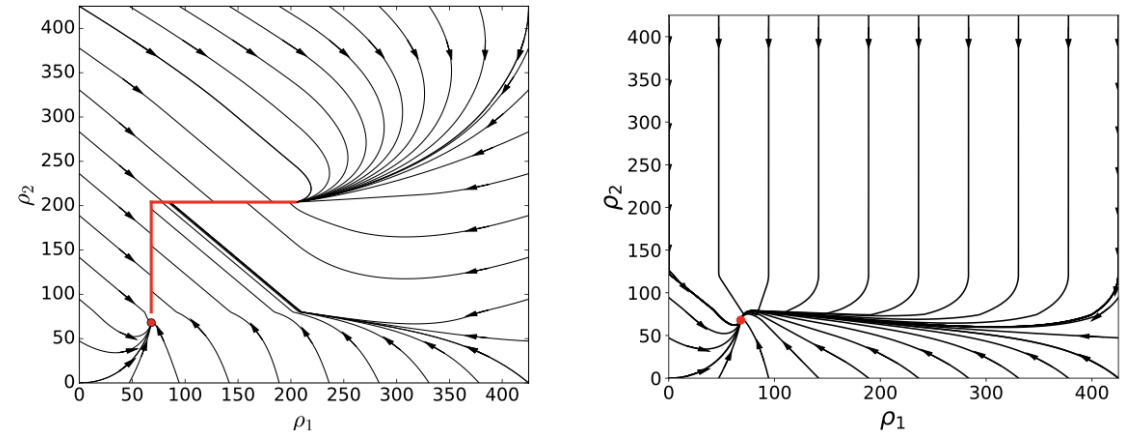
$$\begin{aligned} S^e = & \{\rho \left| \rho_1 = \frac{d}{v_f}, \frac{C_d}{v_f} < \rho_2 \le \rho^j - \frac{d}{w} \right.\} \cup \{\rho \left| \frac{d}{v_f} \le \rho_1 \le \rho^j - \frac{d}{w}, \rho_2 = \rho^j - \frac{d}{w} \right.\} \\ = & \{\rho | \rho_1 = 68{,}80 < \rho_2 \le 204\} \cup \{\rho | 68 \le \rho_1 \le 204, \rho_2 = 204\}, \end{aligned}$$

indicated by the red line. When the VSL control is applied, all the density state trajectories of the closed-loop system converge to the unique equilibrium state $\rho^e = \left(\frac{d}{v_f}, \frac{d}{v_f}\right) = (68,68)$, as shown in Figure 91 (b). When $I \in \Omega_3$, each density state trajectory of the open-loop system shown in Figure 92 (a) converges to one of the two isolated equilibrium states, $\rho_1^e = \left(d/v_f, d/v_f\right) = (77,77)$ and $\rho_2^e = \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}, \rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) = (204,204)$, indicated by the red dot and red star respectively. All closed-loop state trajectories shown in Figure 92 (b) converge to the unique equilibrium state $\rho_1^e = \left(d/v_f, d/v_f\right) = (77,77)$. Figure 93 (a) shows that when $I \in \Omega_4$, all the density state trajectories converge to the unique equilibrium state $\rho^e = \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}, \rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) = (204,204)$, indicated by the red dot. The phase portrait of the corresponding closed-loop system is plotted in Figure 93 (b). As shown in Theorem 1.6, all density state trajectories converge to the desired equilibrium state $\rho^e = \left(C_d/v_f, C_d/v_f\right) = (80,80)$, indicated by the red dot in Figure 93 (b). Furthermore, ρ_2 converges to $\rho_2 = 80$ when the initial condition satisfies $\rho_2(0) \le 80$. If $\rho_2(0) > 80$, $\rho_2(t)$ decreases to $\rho_2 = 75$ first, then increases and converges to 80, which guarantees the steady-state flow rate $C_d = 5200$ veh/h. When $I \in \Omega_5$, capacity drop will not occur since the downstream capacity is higher than the capacity of the road sections. All state trajectories in Figure 94 converge to the unique equilibrium state $\rho^e = \left(\min\{d, C\}/v_f, \min\{d, C\}/v_f\right) = (92.3,92.3)$. The open-loop and closed-loop behavior when $I \in \Omega_5$ are the same as expected.



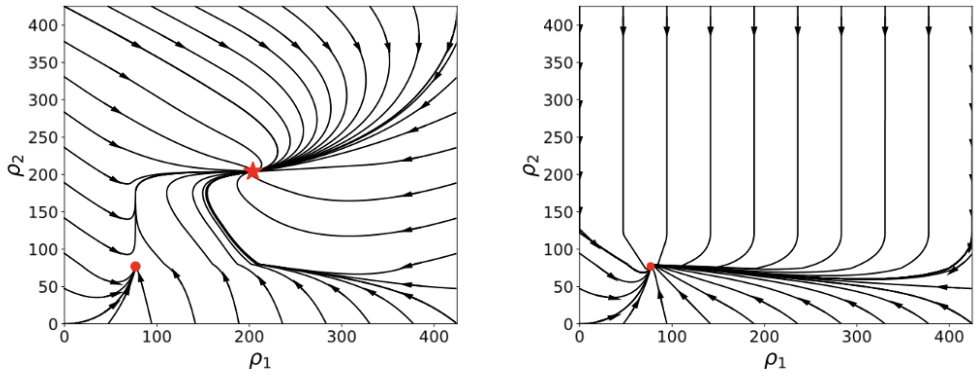
(a) Open-loop (b) Closed-loop

Figure 90. Phase portrait when $I \in \Omega_1(C_d < C, d < (1-\epsilon_0)C_d)$. Both the open-loop and closed-loop densities converge to the same low density equilibrium state. ● Single low density equilibrium state.



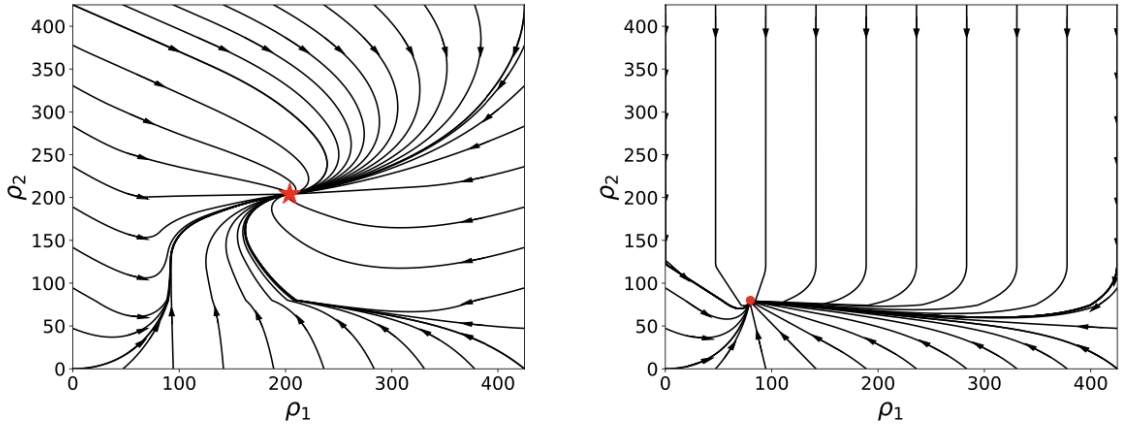
(a) Open-loop (b) Closed-loop

Figure 91. Phase portrait when $I \in \Omega_2(C_d < C, d = (1-\epsilon_0)C_d)$. The open-loop system has an infinite number of equilibrium density states which do not correspond to the maximum possible flow speed. Closed-loop system has a single low density equilibrium state. ● Equilibrium state; ___ Equilibrium manifold.



(a) Open-loop **(b) Closed-loop**

Figure 92. Phase portrait when $I \in \Omega_3(C_d < C, (1-\epsilon_0)C_d < d \leq C_d)$. The open-loop system has two equilibrium density states one in the low density and the other in the high density region. The closed-loop system has a unique equilibrium state at low density. ● Low density equilibrium state; ★ High density equilibrium state.



(a) Open-loop (b) Closed-loop

Figure 93. Phase portrait when $I \in \Omega_4(C_d < C, d > C_d)$. The open-loop system has a uniqueequilibrium state in the high density region. The closed-loop system has a unique equilibrium state at low density. ● Low density equilibrium state; ★ High density equilibrium state.

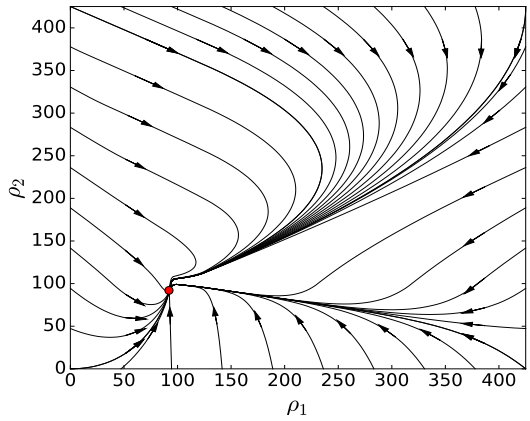
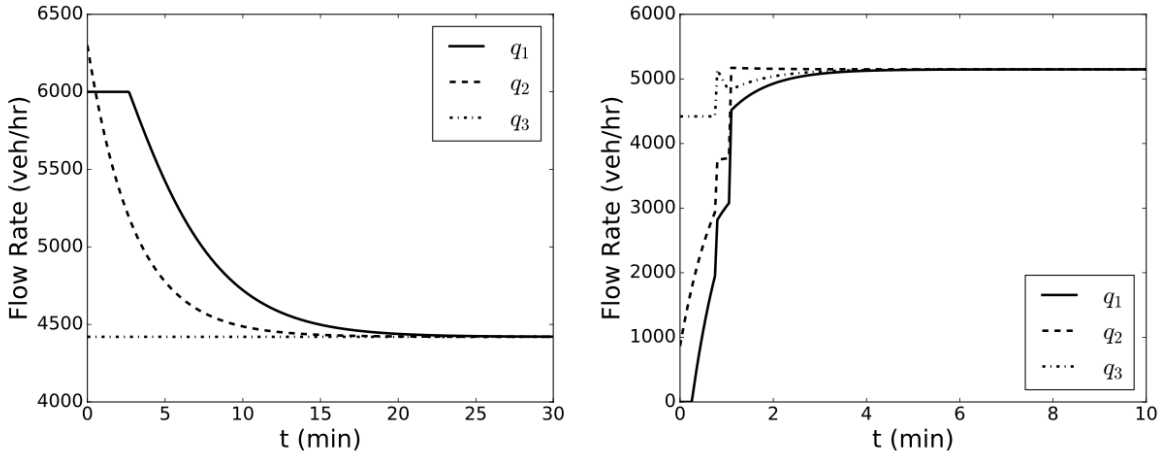


Figure 94. Phase portrait when $I \in \Omega_5 (C_d \geq C)$. Same open-loop and closed-loop response. ● Single low density equilibrium state.

Figure 95 shows the flow rate time responses of the open-loop and closed-loop systems when $I \in \Omega_4$ with initial condition $\rho = (110,110)$. From Figure 95 (a), we can see that at $t = 0$, $q_1(0) = d = 6000$ veh/h, $q_2(0) = v_f \rho_1(0) = 6500$ veh/h $< q_1(0)$ and decrease to the steady state value of 4420 veh/h. On the other hand, $q_3 = (1 - \epsilon_0) C_d = 4420$ veh/h remains constant during the entire simulation time (30 min). In Figure 95 (b), $q_3 = (1 - \epsilon_0) C_d = 4420$ veh/h at the beginning of the simulation, then jumps to 5200 veh/h, then oscillates a little and converges to $C_d = 5200$ veh/h. The jump in the value of q_3 is due to the fact that ρ_2 decreases and crosses the value $\frac{C_d}{v_f}$, at which q_3 jumps from $(1 - \epsilon_0) C_d$ to C_d. The values of q_1 and q_2 also have a jump between $t = 1$ min and $t = 2$ min. This jump is caused by the switching of the VSL control which at this time does not affect q_3 since q_3 is only a function of ρ_2, and does not jump when the VSL switches.



(a) Open-loop (b) Closed-loop

Figure 95. Flow rate when I ∈ Ω4

Figure 96 shows the performance of the closed-loop system in the same scenario as in Figure 95, however with perturbed v_f. In this case, the actual free flow speed $v_f = 0.9v_{fn}$, where v_{fn} is the nominal value of v_f, based on which the controller (56) is designed. That is, the VSL controller is over-estimating the flow rate at the bottleneck, therefore sends more flow to section 2 than it can handle, which leads to temporary capacity drop, which the controller corrects creating an oscillation around an average that corresponds to the desired flow as shown in Figure 96 (a). Our controller however can be easily modified to take care of the uncertainty without changing the fundamentals of the design and analysis. As shown in Figure 96 (b), we modified the controller (56) as stated earlier by setting $x_i = \rho_i - \left(\frac{C_d}{v_{fn}} - 5\right) = \rho_i - 75, i = 1,2$ and increasing the feedback gains to be $\lambda_i = 100$ mi/h, $i = 0,1$. The modified controller tries to stabilize the density vector at $\rho^e = (75,75)$, which gives a margin between ρ^e and the boundary of capacity drop. The increased feedback gains are able to suppress the steady state error to make sure that the steady state value of ρ is close to ρ^e thus capacity drop does not occur. With the modified VSL controller, the steady state density is $\rho = (74.2,78.44)$, and the steady state traffic flow is $q_1 = q_2 = q_3 = 4590$ veh/h. This a simple case how an uncertainty can be dealt with by the proposed controller. The robustness of the proposed controller with respect to a wide range of uncertainties is currently under investigation and it is outside the scope of this paper that focuses on the control design and analysis under ideal conditions. The results form the basis for comparison as uncertainties are included in the model.

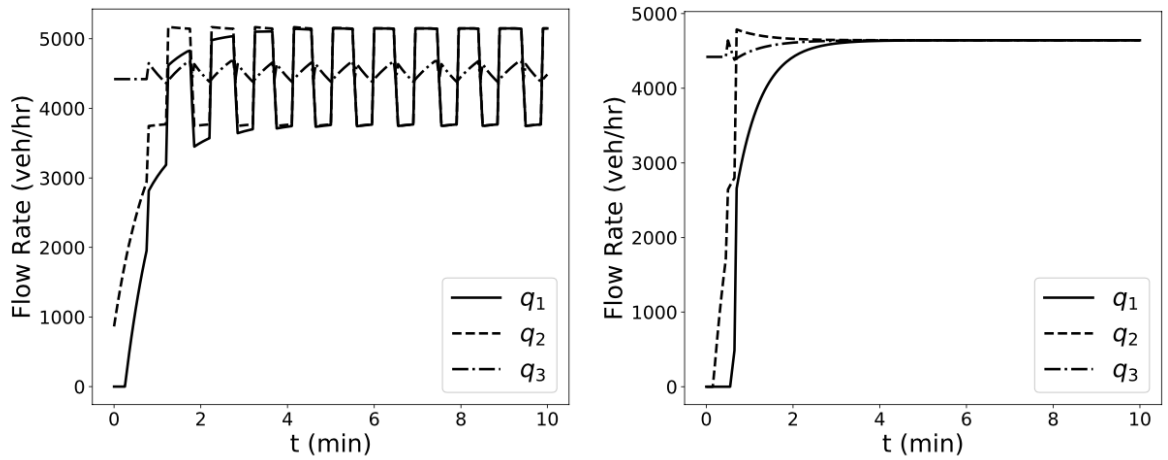


Figure 96. Flow rate with Perturbed vf when I ∈ Ω4

Robust VSL Control of Cell Transmission Model with Disturbance

Section based on the publication:

F. Alasiri, Y. Zhang, and P.A. Ioannou, "Robust variable speed limit control with respect to uncertainties," European Journal of Control 59 (2021): 216-226.

Introduction

VSL Control: Single Section shows the possibility of achieving the maximum possible flow rate at a bottleneck and avoiding capacity drop, under the assumption that we have perfect knowledge of model parameters of the CTM and accurate measurement of the vehicle densities. We have shown as well in VSL Control: Single Section that due to the discontinuous nature of the desired equilibrium point, any disturbance or measurement noise may lead to an oscillatory behavior of the closed-loop system. We also demonstrate with numerical simulations that with simple modification, the VSL controller can help the system avoid the oscillation and stabilize the density at an equilibrium point close to the desired one. In this section, we modify the VSL controller by adding the integral action in order to reject the constant disturbance, which may be introduced by the parametric modeling and measurement errors.

The Cell Transmission Model with Disturbance

Consider a single road section of unit length under the assumption that vehicle density ρ is uniform along the section. The road section is expected to meet a demand of flow d. Let $\tilde{q}_1$ and $\tilde{q}_2$ represent the true value of the inflow and outflow, respectively. Then, according to the conservation law of traffic flow, we have $\dot{\rho} = \tilde{q}_1 - \tilde{q}_2$. In practice, however, we can only measure the corrupted values of $\tilde{q}_1$ and $\tilde{q}_2$ due to the inevitable parametric modeling errors and imperfect measurements of the true flows. Let q_1 and q_2 denote the measured inflow and outflow, respectively, and let μ account for the uncertainties in the model related to modeling and measurement errors. Then, the evolution of the traffic density ρ with respect to time of the considered road section, shown in Figure 85, is given by the following differential equation:

$$\dot{\rho} \quad = q_1 - q_2 + \mu, 0 \le \rho(0) \le \rho^j \tag{58}$$

Where

$$\begin{aligned} q_1 \quad &= \min\{d, C, w(\rho^j - \rho)\}, \\ q_2 \quad &= \begin{cases} \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), (1-\epsilon(\rho))C_d\}, & \text{if } C_d < C \\ \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), C_d\}, & \text{otherwise} \end{cases}, \\ v_f\rho_c \quad &= w(\rho^j - \rho_c) = \tilde{w}(\tilde{\rho}^j - \rho_c) = C, \\ 0 \quad &< \rho_c < \rho^j, 0 < \tilde{w} < w, v_f > 0, \\ \epsilon(\rho) \quad &= \begin{cases} 0 & \text{if } 0 \le \rho \le \dfrac{C_d}{v_f} \\ \epsilon_0 & \text{otherwise} \end{cases}, \end{aligned} \tag{59}$$

and the parameters in equation (59) follow the same definition as in VSL Control: Single Section. In model (58)- (59) the parameter μ is an unknown disturbance that accounts for all uncertainties in the model, and q_1, q_2 are the measured flows. We assume that μ is bounded by a constant μ_m and satisfies $|\mu| \leq \mu_m \ll C_d$. In other words, compared to the bottleneck capacity, the magnitude of the disturbance μ is small, which also guarantees that $0 \leq \rho(t) \leq \rho^j, \forall t \geq 0$. We also assume that μ is approximately constant for large intervals of time.
In the following subsection, the equilibrium points of the open-loop system described by (58)-(59) are identified and their stability properties are analyzed.

The Stability Analysis of the Open-Loop CTM with Disturbance

In order to analyze the stability properties of the dynamical model presented in (58)- (59), all possible operating scenarios are investigated. These scenarios are defined by the capacity of the road section C, the capacity of the downstream section C_d, the capacity drop factor ϵ_0 (which may depend on microscopic results (Kontorinaki et al. 2016; Srivastava and Geroliminis 2013; Y. Zhang and Ioannou 2017b)), and the level of both demand d and disturbance μ. The following theorem presents the results of the analysis.

Theorem 1.7. *Consider the open loop system (58)- (59), where the disturbance term μ is assumed to be constant but otherwise unknown. We have the following results:*

1. *If $(d+\mu) < (1-\epsilon_0)C_d$ and $C > C_d$, then $\rho(t)$ converges exponentially fast to $\frac{d+\mu}{v_f}$, $\forall \rho(0) \in [0, \rho^j]$.*
2. *If $(d+\mu) = (1-\epsilon_0)C_d$ and $C > C_d$, then*
   - *$\rho(t)$ converges exponentially fast to $\frac{d+\mu}{v_f} = \frac{(1-\epsilon_0)C_d}{v_f}$, $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$.*
   - *$\rho(t) = \rho(0)$, $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j - \frac{d+\mu}{w}]$.*
   - *$\rho(t)$ converges exponentially fast to $\rho^j - \frac{d+\mu}{w} = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, $\forall \rho(0) \in (\rho^j - \frac{d+\mu}{w}, \rho^j]$.*
3. *If $(1-\epsilon_0)C_d < (d+\mu) \leq C_d$ and $C > C_d$, then*
   - *$\rho(t)$ converges exponentially fast to $\frac{d+\mu}{v_f}$, $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$.*
   - *$\rho(t)$ converges exponentially fast to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$, $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$.*
4. *If $(d+\mu) > C_d$ and $C > C_d$, then $\rho(t)$ converges exponentially fast to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$, $\forall \rho(0) \in [0, \rho^j]$.*
5. *If $(d+\mu) < C$ and $C \leq C_d$, then $\rho(t)$ converges exponentially fast to $\frac{d+\mu}{v_f}$, $\forall \rho(0) \in [0, \rho_j]$.*

6. *If* $(d+\mu) \geq C$ *and* $C \leq C_d$*, then* $\rho(t)$ *converges exponentially fast to* $\frac{C}{v_f}$*,* $\forall \rho(0) \in [0, \rho_j]$.

Proof. Under the assumption that $(d+\mu) \geq 0$, the proof follows from that of Theorem 1.3 in VSL Control: Single Section for all six cases by replacing d with $(d+\mu)$.

Theorem 1.7 shows that the equilibrium density states of the open-loop system are directly affected by the external disturbance. In fact, the disturbance term has a direct effect on the actual demand and could shift the equilibrium of the flow from the uncongested region to the congested one or trigger a capacity drop much earlier. Depending on the initial condition of the density, the system would reach a state of equilibrium when $(q_1+\mu) = q_2$. While some cases have more than one isolated equilibrium point associated with high density, in case 2, when $(d+\mu) = (1-\epsilon_0)C_d$ *and* $C > C_d$, there is an infinite number of equilibrium points. The aim is to design a controller such that the traffic flow of the road section operates within the free flow region in the fundamental diagram despite the presence of the disturbance term μ. The control input is variable speed limit commands to vehicles upstream in order to protect the section under consideration and maximize the throughput under different demands, initial density conditions, and constant disturbances.

Robust VSL Control of the CTM with Disturbance

The purpose of this section is to design a variable speed limit (VSL) controller that rejects the constant disturbance μ, guarantees convergence to a desired density located in the free-flow region in the fundamental diagram, and improves the throughput of an active bottleneck. The basic idea of the VSL control is to reduce the incoming flow by informing the upstream vehicles to follow a speed limit so that the density and flow rate of the road section converge to the desired possible values, which correspond to the maximum possible throughput at the bottleneck. The control problem would have been trivial if one could directly control the inflow q_1 via traffic light control. Since such an approach is not feasible in most traffic situations, controlling the inflow via the upstream speed is the only feasible choice. The nonlinear relationship between the inflow and upstream speed makes the design and analysis of VSL control more challenging. As shown in Figure 86, the VSL action v is applied to the upstream road section called the "VSL zone". All vehicles in the VSL zone are asked to follow the speed limit v and then follow the free flow speed v_f within the road section under consideration.

If the VSL command v is less than the free flow speed v_f, then the fundamental diagram of the VSL zone is distorted, as shown in Figure 86, assuming that the VSL zone has similar characteristics as the road section (Csikós and Kulcsár 2017; Hadiuzzaman and Qiu 2013; H.-Y. Jin and Jin 2015). From the geometry of the fundamental diagram, it follows that the parameters ρ^j, w, and $\widetilde{w}$ remain unchanged, while the maximum possible flow rate the VSL zone could send to the considered road section is given by the term $\frac{vw\rho^j}{v+w}$. The use of VSL control will affect the inflow q_1, which in addition to upstream demand d will also depend on how

much flow is allowed by the VSL. Therefore, the model (58)- (59) with VSL control inputs is given by:

$$\dot{\rho} = q_1 - q_2 + \mu, 0 \le \rho(0) \le \rho^j, \tag{60}$$

Where

$$\begin{aligned} q_1 &= \min\{d, \frac{vw\rho^j}{v+w}, C, w(\rho^j - \rho)\}, \\ q_2 &= \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), (1-\epsilon(\rho))C_d\}, \\ v_f\rho_c &= w(\rho^j - \rho_c) = \tilde{w}(\tilde{\rho}^j - \rho_c) = C, \\ 0 &< \rho_c < \rho^j, 0 < \tilde{w} < w, v_f > 0, \\ \epsilon(\rho) &= \begin{cases} 0 & \text{if } 0 \le \rho \le \frac{C_d}{v_f}, \\ \epsilon_0 & \text{otherwise} \end{cases} \end{aligned} \tag{61}$$

In equation (61), the term that could be influenced by the VSL is $\frac{vw\rho^j}{v+w}$ since it is the only term that depends on the control action, namely the upstream speed of flow v. Let $q_{1v} = \frac{vw\rho^j}{v+w}$, and, without loss of generality, assume $d < C$. Then, system (58)- (59) can be rewritten as follows:

$$\dot{\rho} = q_1 - q_2 + \mu, 0 \le \rho(0) \le \rho^j, \tag{62}$$

Where

$$\begin{aligned} q_1 &= \min\{d, q_{1v}, w(\rho^j - \rho)\}, \\ q_2 &= \min\{v_f\rho, \tilde{w}(\tilde{\rho}^j - \rho), (1-\epsilon(\rho))C_d\}, \\ q_{1v} &= \text{median}\{0, \bar{q}_{1v}, C\}, \end{aligned} \tag{63}$$

and the parameters in equation (63) follow the same definition as in (61). The median function is used to guarantee that when mapping the controlled flow rate q_{1v} into the VSL command, i.e., $v = \frac{wq_{1v}}{w\rho^j - q_{1v}}$, the speed does not become less than zero or exceed the free flow speed limit. $\bar{q}_{1v}$ is the unconstrained control variable to be designed.

The following constants $0 < \rho^L \le \rho^\star < \rho^U < \frac{C_d}{v_f}$, shown in Figure 97, are defined to help design the controller, where $\rho^\star$ is the desired value to which we want the traffic density of the road section to converge. ρ^L and ρ^U denote the lower and upper bounds, respectively, of $\rho^\star$.

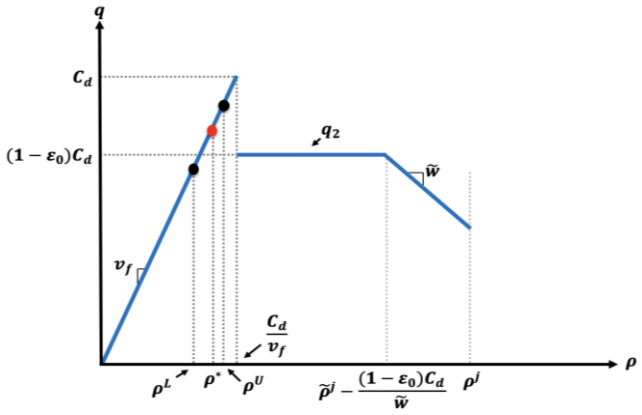


Figure 97. Design Constants

Selecting the value of $\rho^\star$ is critical for designing the controller. $\rho^\star$ needs to be chosen so that the bottleneck congestion is reduced or prevented and the discharging flow rate is improved as compared to the case without applying a control action. Intuitively, $\rho^\star$ should be in the uncongested region of the fundamental diagram to take advantage of the free-flow speed in that region. A trivial choice is to make $\rho^\star = \frac{C_d}{v_f}$. This choice corresponds to the highest discharging flow in this case. However, small disturbances or faulty measurements may push the density towards the capacity-drop region, resulting in an oscillatory behavior of the closed-loop system, as shown in (Y. Zhang and Ioannou 2018). The values of $\rho^\star$ and ρ^U should be chosen so that we do not lose excessive potential capacity. Therefore, $\rho^\star$ and ρ^U can be arbitrarily close to the value $\frac{C_d}{v_f}$ as long as the inequality $0 < \rho^L \leq \rho^\star < \rho^U < \frac{C_d}{v_f}$ holds, as will be shown later in Theorem 1.8. ρ^L and ρ^U are introduced to prevent unwanted rapid switching when applying the controller.

When the traffic of the road section is congested due to the activation of the bottleneck, we choose v so that we decrease the inflow q_1 in order to bring ρ to the uncongested region in the fundamental diagram by letting

$$\bar{q}_{1v} = q_s \tag{64}$$

where q_s is a small constant flow such that $q_s < \min\{v_f\rho^L, (1-\epsilon_0)C_d, \tilde{w}(\tilde{\rho}^j - \rho^j)\}$, which guarantees that $\forall \rho \geq \rho^L$, $q_s < q_2$, implying that $\dot{\rho} < 0$. A trivial selection is $q_s = 0$. Theoretically, this selection is valid. However, the implication of such a choice leads to sacrificing the potential road capacity. Therefore, the constant flow q_s should be selected in a way that we do not significantly lose some of the potential capacity. Thus, q_s can be arbitrarily close to the value $\min\{v_f\rho^L, (1-\epsilon_0)C_d, \tilde{w}(\tilde{\rho}^j - \rho^j)\}$ as long as the inequality $q_s < \min\{v_f\rho^L, (1-\epsilon_0)C_d, \tilde{w}(\tilde{\rho}^j - \rho^j)\}$ is satisfied. Since the outflow is greater than the inflow in

this case, the traffic density of the road section is decreasing. Thus, there exists a finite time instant $t_0 > 0$ at which $\rho(t_0) = \rho^L$, and we choose v such that

$$\bar{q}_{1v} = q_2 - \lambda_1(\rho - \rho^\star) - \lambda_2\left(\int_{t_0}^{t}(\rho - \rho^\star)\,d\tau + c\right) \tag{65}$$

is the integral gain used to determine how fast the steady-state error is eliminated, $\lambda_1 > \max\left\{2\sqrt{\lambda_2}, v_f + \frac{\lambda_2}{v_f}\right\} > 0$ is the proportional gain used to determine how fast the control system responds, and c is a design constant to be selected in order to guarantee that ρ asymptotically converges to $\rho^\star$, as we will show later. Equation (65) is a proportional–integral (PI) controller, which rejects the disturbance μ and stabilizes the density at $\rho = \rho^\star$. The integral action enables the PI controller to eliminate the offset with respect to $\rho^\star$ introduced by the disturbance μ. If the integral part is removed from equation (65), the steady-state error with respect to $\rho^\star$ may result in capacity drop, as shown in (Y. Zhang and Ioannou 2018). We can increase the value of λ_1 to suppress the steady-state error in this case. However, the controlled flow rate q_{1v} is constrained by the saturation of the VSL as well as the road capacity. Therefore, the proportional plus integral control is essential for stability. It rejects the effect of the disturbances and drives the density to the desired point.

If ρ decreases to the uncongested region $\rho \leq \frac{C_d}{v_f}$, we do not want the disturbance to push back the density to the capacity drop region. Therefore, if ρ increases and reaches ρ^U, we switch back to (64). In order to avoid undesirable stability phenomena and oscillations as a result of frequent switching, we introduce a hysteresis to make the switching continuous. Therefore, the unconstrained control flow $\bar{q}_{1v}$ with hysteresis characteristics is described as follows

$$\begin{aligned}
\bar{q}_{1v} &= k(t), \\
k(0) &= \begin{cases} k_1(0) & \text{if } \rho(0) > \rho^L \\ k_2(0) & \text{otherwise,} \end{cases} \\
k(t) &= \begin{cases} k_1(t) & \text{if } k(t^-) = k_2 \text{ and } \rho(t) = \rho^U \\ k_2(t) & \text{if } k(t^-) = k_1 \text{ and } \rho(t) = \rho^L \\ k(t^-) & \text{otherwise,} \end{cases} \forall t > 0
\end{aligned} \tag{66}$$

where

$$\begin{aligned}
k_1(t) &= q_s, \\
k_2(t) &= q_2 - \lambda_1(\rho - \rho^\star) - \lambda_2\left(\int_{t_0}^{t}(\rho - \rho^\star)\,d\tau - c\right), t \geq t_0.
\end{aligned}$$

Mapping the flow rate into the VSL command v, we have

$$v = \frac{wq_{1v}}{w\rho^j - q_{1v}}. \tag{67}$$

where $q_{1v} = \text{median}\{0, \bar{q}_{1v}, C\}$, and the constants in equations k_1 and k_2 are defined as follows:

$$q_s < \min\{v_f\rho^L, (1-\epsilon_0)C_d, \tilde{w}(\tilde{\rho}^j - \rho^j)\}, \quad \lambda_2 > 0, \quad \lambda_1 > \max\left\{2\sqrt{\lambda_2}, v_f + \frac{\lambda_2}{v_f}\right\} > 0,$$

$$c = \frac{\lambda_1(\rho(t_0) - \rho^\star) - \mu_m}{\lambda_2}.$$

In equation (66), $k_1(t) = q_s$, where q_s is the constant flow rate that satisfies $q_s < \min\{v_f\rho^L, (1-\epsilon_0)C_d, \tilde{w}(\tilde{\rho}^j - \rho^j)\}$ in order to decrease the density and keep it operating within the free flow region. Since the goal is to improve the throughput, the value of q_s should be arbitrarily close to the value $\min\{v_f\rho^L, (1-\epsilon_0)C_d, \tilde{w}(\tilde{\rho}^j - \rho^j)\}$ as long as the inequality holds in order to take advantage of some of the road capacity potential. $k_2(t)$ is a proportional-integral (PI) controller, in which the term $\lambda_1(\rho - \rho^\star)$ is the proportional term to drive ρ towards $\rho^\star$, the term $\lambda_2\left(\int_{t_0}^{t}(\rho - \rho^\star)\,d\tau - c\right)$ is the integral term to reject the disturbance μ, and c is a constant term that guarantees the asymptotic convergence of ρ towards $\rho^\star$. The selection of c and the lower bound of λ_1 is explained in detail in the proof of Lemma 1.2 and Theorem 1.8.

Depending on the value of the flow density at the current time $\rho(t)$ and on whether at the previous time $k_1(t^-)$ or $k_1(t^-)$ is activated, the controller $k(t)$ switches according to the 'if conditions' in equation (66). Note that while k_2 is active, $\rho(t) < \rho^L$. To prevent the disturbances from pushing the density towards the capacity drop region, the controller switches from k_2 to k_1 at $\rho(t) = \rho^L$ in order to reduce the value of the traffic density and maintain it within the uncongested region of the fundamental diagram. Once $\rho(t) = \rho^U$, the controller switches to the PI control k_2 to drive the traffic density of the road section to the predetermined density ρ^*. In the $k_2(t)$ equation, if $k(0) = k_2(0)$ at $t = 0$, then $t_0 = 0$. If $k(t)$ switches from $k_1(t)$ to $k_2(t)$ at $t \neq 0$, then $t_0 = t$ is the switching time instant. After obtaining the value of the unconstrained control flow $\bar{q}_{1v}$, we map the constraint controlled flow rate $q_{1v} = \text{median}\{0, \bar{q}_{1v}, C\}$ into the VSL control action via equation (67).

Note that when $k_2(t)$ is activated in equation (66), the value of the section density is strictly less than ρ^U, i.e., $\rho < \rho^U < \frac{C_d}{v_f}$. Therefore, $q_2 = v_f\rho$ and $q_1 = \text{median}\{0, d, \bar{q}_{1v}\}$, due to $d < C$. Let

$$\Phi(t) = \int_{t_0}^{t}(\rho(\tau) - \rho^\star)\,d\tau - \frac{\lambda_1(\rho(t_0) - \rho^\star) - \mu_m}{\lambda_2}, t \geq t_0. \tag{68}$$

Then, we have $\dot{\Phi}(t) = \rho(t) - \rho^\star$. Since $\Phi(t) = \Phi(t_0) + \int_{t_0}^{t} \dot{\Phi}(\tau)d\tau,\ t \geq t_0$, we have that $\Phi(t_0) = -\frac{\lambda_1(\rho(t_0)-\rho^\star)-\mu_m}{\lambda_2}$. Therefore, with $k(t) = k_2(t)$, system (62)-(63) can be written as:

$$\begin{aligned} &\dot{\rho}(t) = \text{median}\{0, d, \bar{q}_{1v}\} - v_f\rho + \mu \\ &\dot{\Phi}(t) = \rho(t) - \rho^\star,\ \forall t \geq t_0 \\ &\rho(t_0) \leq \rho^L, \Phi(t_0) = -\frac{\lambda_1(\rho(t_0) - \rho^\star) - \mu_m}{\lambda_2} \end{aligned} \tag{69}$$

where

$$\bar{q}_{1v} = k_2(t) = v_f\rho - \lambda_1(\rho - \rho^\star) - \lambda_2\Phi,$$

and $\lambda_2 > 0, \lambda_1 > \max\{2\sqrt{\lambda_2}, v_f + \frac{\lambda_2}{v_f}\} > 0$. Let us first assume that capacity drop is ignored, i.e., (69) holds for all $\rho(t) \in \Re$. The following lemma describes the stability properties of (69). Lemma 1.2 is subsequently used to analyze the closed-loop system (62)-(63), with the controller (67), taking into account the capacity drop effect.

Lemma 1.2. *Consider system (69), if $d + \mu \geq v_f\rho^\star$, we have the following results:*

1. *System (69) has a unique equilibrium point $[\rho(t)^e, \Phi(t)^e]^T = \left[\rho^\star, \frac{\mu}{\lambda_2}\right]^T$.*
2. *$\forall\big(\rho(t_0), \Phi(t_0)\big) \in \Re^2$, $[\rho(t), \Phi(t)]^T$ asymptotically converges to $\left[\rho^\star, \frac{\mu}{\lambda_2}\right]^T$.*
3. *$\forall\big(\rho(t_0), \Phi(t_0)\big) \in S = \{(\rho, \Phi)\,\big|\, v_f\rho^\star - \mu - d \leq \big(\lambda_1 - v_f\big)(\rho - \rho^\star) + \lambda_2\left(\Phi - \frac{\mu}{\lambda_2}\right) \leq v_f\rho^\star - \mu\} \cap \{(\rho, \Phi)\big| -\frac{v_f\rho^\star-\mu}{v_f} < (\rho - \rho^\star) < -\frac{v_f\rho^\star-\mu-d}{v_f}\}$, then $(\rho(t) - \rho^\star) \in S, \forall t \geq t_0$.*

Proof. The proof of Lemma Lemma 1.2 is in Appendix G.

Based on the value of the inflow q_1, Lemma 1.2 shows that the trajectories of system (69) asymptotically converge to the unique equilibrium point $[\rho(t)^e, \Phi(t)^e]^T = \left[\rho^\star, \frac{\mu}{\lambda_2}\right]^T$. It follows from the proof that if system (69) holds for $\forall\rho \in \Re$, then the unique equilibrium point $[\rho(t)^e, \Phi(t)^e]^T = \left[\rho^\star, \frac{\mu}{\lambda_2}\right]^T$ is globally asymptotically stable. However, the proof also indicates that if the initial conditions $\big(\rho(t_0), \Phi(t_0)\big)$ are in the region $S_3 = \{(\rho, \Phi)|\big(\lambda_1 - v_f\big)(\rho - \rho^\star) + \lambda_2\left(\Phi - \frac{\mu}{\lambda_2}\right)\}$, it is possible that the trajectory of $\rho(t)$ approaches $-\frac{v_f\rho^\star-\mu-d}{v_f}$, which is already in the capacity drop region, where (69) does not hold. Fortunately, the initial conditions in system (69) are set to be in a certain region. Therefore, we only need to show that for some specific $\rho(t_0)$, system (69) holds for all $t \geq t_0$, and $\rho(t)$ converges to the predetermined density $\rho^\star$. The results of Lemma 1.2 are used as a stepping stone to help analyze the stability properties of the closed-loop system (62)-(63) with (67) in the following theorems.

Theorem 1.8. *Consider the system in (62)-(63) with the controller (67), if* $(d+\mu) \geq v_f\rho^\star$, *then* $\rho(t)$ *converges to* $\rho^\star$ *asymptotically with time,* $\forall\rho(0) \in [0, \rho^j]$.

Proof. $\forall\rho(0) < \rho^L$, $k(0) = k_2(0)$. If $\forall t \geq 0,\ \rho(0) < \rho^U$, then (69) holds for all $t \geq 0$, thus $\rho(t)$ converges to $\rho^\star$ asymptotically according to Lemma 1.2. If $\exists t > 0$, such that $\rho(t) = \rho^U$, then $k(t)$ switches to $k_1(t)$ and for all $\rho(t) > \rho^L$, $\dot{\rho}(t) < 0$, thus $\exists t_0 > 0$, such that $\rho(t_0) = \rho^L$ and $k(t_0) = k_2(t_0)$. Similarly, $\forall\rho(0) > \rho^L$, $k(0) = k_1(0)$, then $\exists t_0 > 0$, such that $\rho(t_0) = \rho^L$ and $k(t_0) = k_2(t_0)$. Therefore, we only need to consider the case when $\rho(t_0) = \rho^L$ and $k(t_0) = k_2(t_0)$. To shift the equilibrium point of the system to the origin, define $x_1(t) = \rho(t) - \rho^\star$ and $x_2(t) = \Phi(t) - \frac{\mu}{\lambda_2}$. As long as $k(t) = k_2(t)$, system (62)-(63) can be written as follows:

$$\dot{x}_1 = \text{median}\{0, d, \bar{q}_{1v}\} - \left(v_f\rho^\star + v_f x_1\right) + \mu$$
$$\dot{x}_2 = x_1,\ \forall t \geq t_0$$
$$x_1(t_0) = \rho^L - \rho^\star, x_2(t_0) = -\frac{\lambda_1 x_1(t_0) - \mu_m + \mu}{\lambda_2}$$

Since $x_2(t_0) = -\frac{\lambda_1 x_1(t_0) - \mu_m + \mu}{\lambda_2}$, we have that $\lambda_1 x_1(t_0) + \lambda_2 x_2(t_0) = \mu_m - \mu > 0$, and because of $\mu_m \ll v_f\rho^\star$, we also have $\lambda_1 x_1(t_0) + \lambda_2 x_2(t_0) < v_f\rho^\star - \mu$. Furthermore, $x_1(t_0) = \rho^L - \rho^\star > -\frac{v_f\rho^\star - \mu_m}{v_f} > -\frac{v_f\rho^\star - \mu}{v_f}$. Similarly, $x_1(t_0) < -\frac{v_f\rho^\star - \mu - d}{v_f}$. As a result, $x(t_0) \in S$. According to Lemma 1.2, as long as $x_1(t) < \rho^U - \rho^\star$ we have

$$\dot{x} = Ax$$
$$x_1(t_0) = \rho^L - \rho^\star, x_2(t_0) = -\frac{\lambda_1 x_1(t_0)}{\lambda_2}$$

where

$$A = \begin{bmatrix} -\lambda_1 & -\lambda_2 \\ 1 & 0 \end{bmatrix}.$$

Since $\lambda_1 > 2\sqrt{\lambda_2}$, *A* has two real negative roots, i.e.,

$$p_1 = \frac{-\lambda_1 + \sqrt{\lambda_1^2 - 4\lambda_2}}{2},\ p_1 = \frac{-\lambda_1 - \sqrt{\lambda_1^2 - 4\lambda_2}}{2}$$

where $0 > p_1 > p_2$. We can calculate that

$$e^{At} = \mathcal{L}^{-1}[(sI - A)^{-1}] = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix}$$

where

$$a_{11}(t) = \frac{1}{2}(e^{p_1 t} + e^{p_2 t}) - \frac{\lambda_1}{p_1 - p_2}(e^{p_1 t} - 2e^{p_2 t})$$

and

$$a_{12}(t) = -\frac{\lambda_2}{p_1 - p_2}(e^{p_1 t} - e^{p_2 t}).$$

Therefore,

$$\begin{aligned} x(t) &= x_1(t_0)a_{11}(t - t_0) + x_2(t_0)a_{12}(t - t_0) \\ &= \frac{1}{2}x_1(t_0)\left(e^{p_1(t-t_0)} + e^{p_2(t-t_0)}\right) - \frac{\lambda_1 x_1(t_0) + \lambda_2 x_2(t_0)}{p_1 - p_2}\left(e^{p_1(t-t_0)} - e^{p_2(t-t_0)}\right) \end{aligned}$$

Since $p_1 > p_2, x_1(t_0) \leq 0$ and $\lambda_1 x_1(t_0) + \lambda_2 x_2(t_0) > 0$, we have that $x(t) < 0,\ \forall t \geq t_0$. Consequently, $\dot{x}_1 = -v_f x_1 - v_f \rho^\star + \mu + d$ and $\dot{x}_2 = x_1$, for all $t > t_0$. According to Lemma 1.2, $x_1(t)$ converges to 0 asymptotically, which implies that $\rho(t)$ asymptotically converges to $\rho^\star$, $\forall \rho(0) \in [0, \rho^j]$.

Theorem 1.8 shows that if $(d + \mu) \geq v_f \rho^\star$, then controller (67) drives $\rho(t)$ to $\rho^\star$ asymptotically. In the case when $(d + \mu) < v_f \rho^\star$, the dynamics and the stability properties of the closed-loop system are given by the following theorem.

Theorem 1.9. *Consider the system in (62)-(63) with the controller (67), if* $(d + \mu) < v_f \rho^\star$, *then* $\rho(t)$ *converges to* $\frac{(d+\mu)}{v_f}$ *asymptotically with time,* $\forall \rho(0) \in [0, \rho^j]$.

Proof. Since $d + \mu < v_f \rho^\star$, then $\exists \eta > 0$, such that $d + \mu \leq v_f \rho^\star - \eta$. Similar to the analysis in Theorem 1.8, we only need to consider the case where $\rho(t_0) \leq \rho^L$ and $k(t_0) = k_2(t_0)$. Let $x_1(t) = \rho(t) - \rho^\star$ and $x_2(t) = \Phi(t) - \frac{\mu}{\lambda_2}$ in order to shift the equilibrium point of the system to $(0,0)$. Then, according to (62)-(63), we have

$$\dot{x}_1 \leq d - q_2 + \mu = d + \mu - \left(v_f \rho^\star + v_f x_1\right)$$

Thus, $\forall x_1(t_0) \leq \rho^L - \rho^\star$, $x_1(t) \leq \frac{d+\mu}{v_f} - \rho^\star, \forall t \geq t_0$. Therefore,

$$x_2(t) = x_2(t_0) + \int_{t_0}^{t} x_1(\tau)d\tau - \frac{\mu}{\lambda_2} \leq x_2(t_0) - \frac{\mu}{\lambda_2} - \eta(t - t_0), t \geq t_0$$

which decreases to negative infinity as t increases. As a result, q_1 saturates at the value of d, and thus $\dot{x}_1 = d + \mu - \left(v_f \rho^\star + v_f x_1\right)$. Therefore, x_1 converges to $\frac{(d+\mu)}{v_f} - \rho^\star$, which implies that $\rho(t)$ asymptotically converges to $\frac{(d+\mu)}{v_f}$, $\forall \rho(0) \in [0, \rho^j]$.

In summary, with the proposed controller, if the sum of the upstream demand d and the disturbance μ is greater than or equal to the predetermined equilibrium flow, the density in the section will converge to the equilibrium point $\rho^\star$. If the sum of the upstream demand d and the disturbance μ is less than the predetermined equilibrium flow, the density converges to $\frac{d+\mu}{v_f}$, at

which the steady state flow $d + \mu$ is the maximum possible value. Note that the selection of ρ^L affects the distance from the switching point to the desired equilibrium point. According to the proof of Theorem 1.8, we can select $\rho^L = \rho^\star$, which still guarantees convergence. In addition, since $\rho(t)$ always converges to $\rho^\star$ from the left side, $\rho^\star$ can be arbitrarily chosen to be close to $\frac{C_d}{v_f}$, as long as the inequality $0 < \rho^L \leq \rho^\star < \rho^U < \frac{C_d}{v_f}$ holds.

Numerical Simulations

In this section, the commercial microscopic simulator VISSIM is used to evaluate the performance of the proposed robust VSL control. The results are compared with those generated by the macroscopic CTM used in the design and analysis.

Simulation Network and Fundamental Diagram

The traffic flow on the southbound segment of the I-710 freeway in Long Beach, California, United States, is simulated using the microscopic simulator VISSIM, without considering the on-ramps and off-ramps. The considered section of the freeway to be controlled has 3 lanes and is divided into two segments of length 0.71 $mile$ and 0.36 $mile$ upstream of a bottleneck location, as shown in Figure 98.

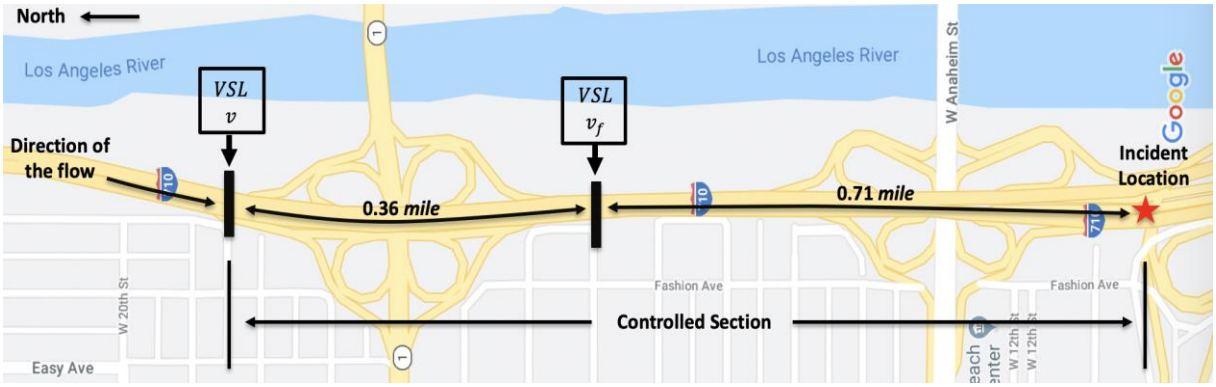


Figure 98. Simulation Network of the I-710 Freeway

Under different levels of traffic demand and by taking into account the high volume of heavy trucks (since the freeway is close to the port of Long Beach), the data of the density and flow rate of the 0.71-mile segment are collected in order to obtain the fundamental diagram for the case of no bottleneck. Then, a bottleneck is created by introducing an incident that blocks the middle lane, and the data of the density and flow rate are gathered to demonstrate the fundamental diagram for this case. As shown in Figure 99, the blue line describes the fundamental diagram of the 0.71 mile-long segment without incident. The maximum capacity of the road segment is 8460 veh/hr, which corresponds to a critical density $\rho_c = 140$ $veh/mile$. The free flow speed is $v_f = 60$ $mile/hr$. If the bottleneck is activated, then the resulting fundamental diagram is shown by the red line in the same figure, where the maximum capacity in this case is 5200 veh/hr.

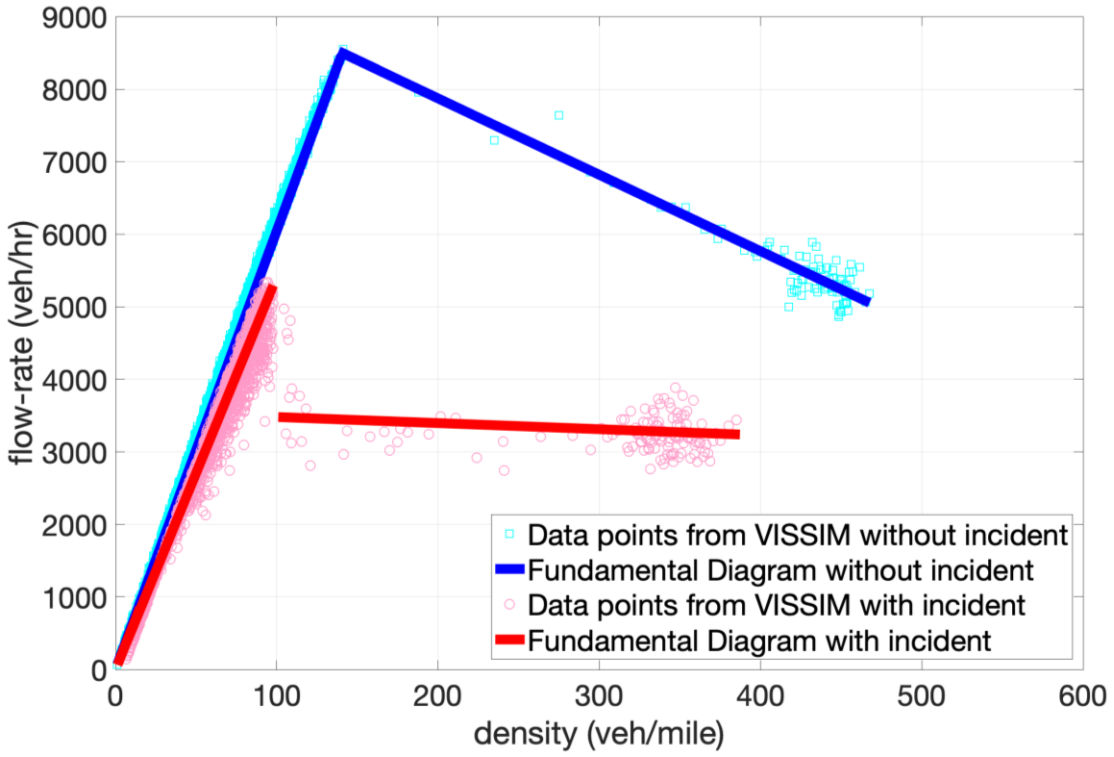


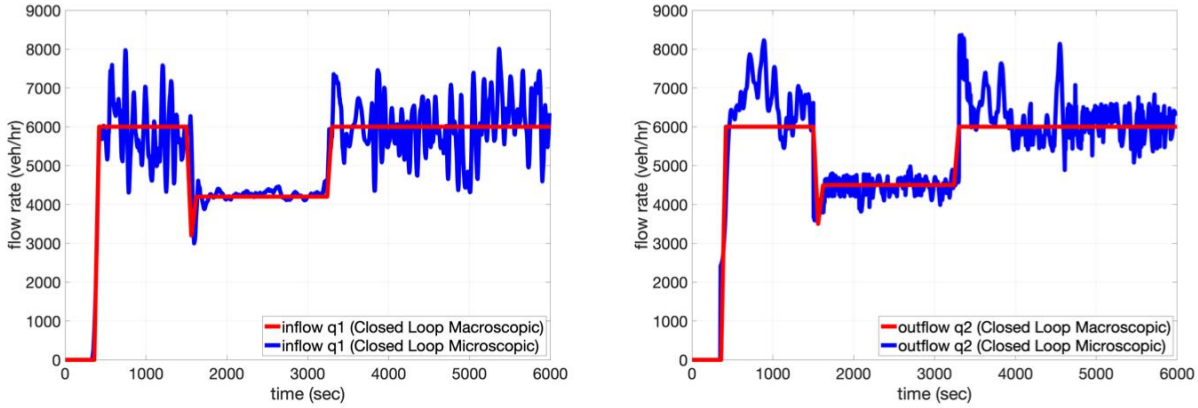
Figure 99. Fundamental Diagrams of the Road Section with VISSIM Data Points

Simulation Results

Macroscopic Simulation

The resulting data from the microscopic simulator are used to validate the macroscopic model. The macroscopic simulations are performed on a single-section road network with the following parameters: $C = 8460\ veh/hr$, $v_f = 60\ mile/hr$, $\rho^j = 460\ veh/mile$, $\tilde{\rho}^j = 980\ veh/mile$, $w = 21\ mile/hr$, $\tilde{w} = 10\ mile/hr$, $C_d = 5200\ veh/hr$, and the demand $d = 6000\ veh/hr$. The controller is applied to the single-section system with the following design constants: $\lambda_1 = 200$, $\lambda_2 = 900$, $\mu_m = 350\ veh/hr$, $q_s = 3200\ veh/hr$, $\rho^* = \rho^L = 75\ veh/mile$, and $\rho^U = 84\ veh/mile$. It is assumed that the incident takes place 25 minutes after the simulation starts and remains for 30 minutes. Figure 100, Figure 101, and Figure 102 show the behavior of the flow rates, density and the VSL commands, respectively, with both the macroscopic and microscopic simulations shown together for the sake of comparison. The solid red lines illustrate the macroscopic results. As seen from the Figure 101, when the incident happens at 1500 sec, the density is 100 $veh/mile$, which is greater than ρ^L. It results in reducing the inflow q_1, represented by the red line in Figure 100 (a), since $q_{1v} = \bar{q}_{1v} = q_s = 3200\ veh/hr$. Thus, the VSL command is $v = 10\ mile/hr$. The outflow q_2, represented by the red line Figure 100 (b), decreases as the value of ρ keeps decreasing until $\rho \leq \rho_L$. At that moment, the PI controller takes over and eventually forces the outflow to converge to 4500 veh/hr, which corresponds to the predetermined density value $\rho^* = 75\ veh/mile$, where the VSL commands are $v = 15\ mile/hr$. During the incident time (from 1500 to 3300 sec), there is a difference of 300 veh/hr between the values of q_1 and q_2 since $q_1 - q_2 + \mu = 0$. After the incident is removed, the VSL

displays the value of 60 $mile/hr$, and both inflow and outflow converge to the constant demand $d = 6000\ veh/hr$.



(a) Inflow q1 (b) Outflow q2

Figure 100. Macroscopic/Microscopic Behavior of q1 & q2 of the Closed-loop System

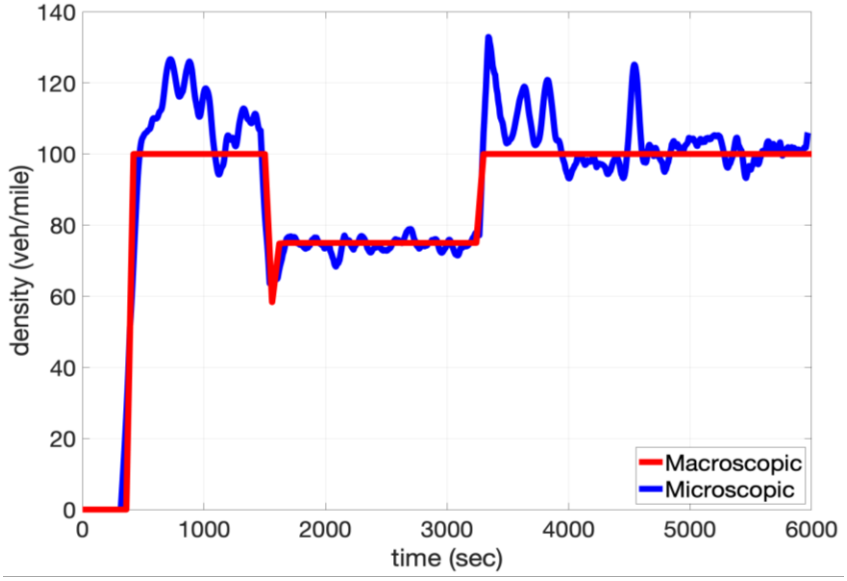


Figure 101. Density of Discharging Section

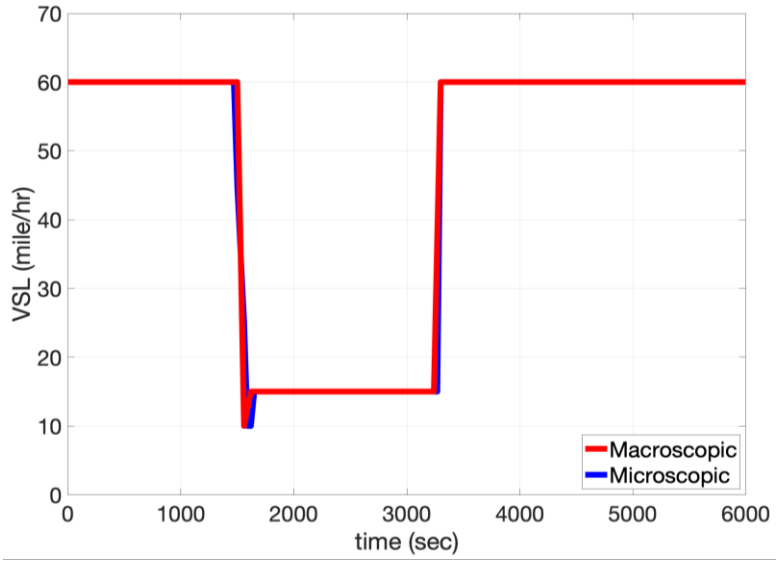


Figure 102. VSL Commands

Microscopic Simulation

The proposed controller is evaluated using the commercial software VISSIM by applying it to the same simulated segments of the I-710 freeway in Long Beach. The freeway is injected with demand $d = 6000\ veh/hr$, with the consideration that the traffic flow in that area has a high volume of trucks since it is near Long Beach Port. The middle lane is blocked with an accident in order to create a bottleneck. The accident takes place 25 minutes after the simulation starts and lasts for 30 minutes (1500 to 3300 sec). The VSL commands are given to vehicles at the beginning of each section, as illustrated in Figure 98.

The blue curves in Figure 100, Figure 101, and Figure 102 show the microscopic behavior of the inflow, the outflow, the density of the discharging section, and the VSL commands. When the accident happens at 1500 sec, the VSL control is activated by gradually reducing the speed in steps of 5 $mile/hr$ to ensure safety for the drivers, until it reaches 10 $mile/hr$. Then it increases to 15 $mile/hr$ where it stays the same until the incident is cleared, as represented by the blue curve in Figure 102. The outflow q_2 reduces dramatically from around 6500 veh/hr to approximately 3600 veh/hr. Then, it slightly increases and fluctuates around 4500 veh/hr. However, once the incident is removed, the flow jumps immediately to around 8400 veh/hr and then gradually decreases with time, when it fluctuates around 6000 veh/hr. As illustrated by the blue curve in Figure 101, the VSL stabilizes the density around $\rho^* = 75\ veh/mile$ and stays less than $\rho^U = 84\ veh/mile$. The resulting VISSIM curves of the inflow, the outflow, the density of the discharging section, and the VSL commands are similar to the ones carried out by the macroscopic simulations.

The microscopic VISSIM simulator is also used to compare the open loop to the closed-loop system for the outflow, density, and the queue length. Figure 103 represents the discharging flow of the bottleneck with and without control.

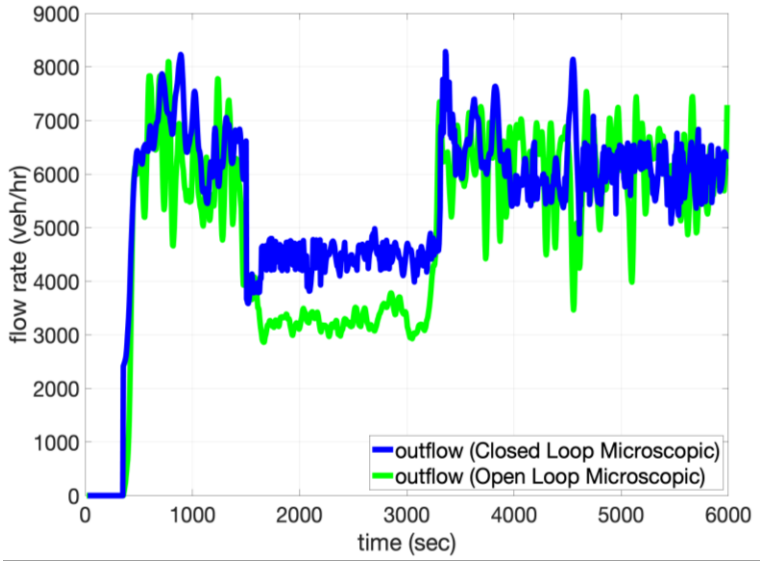


Figure 103. Discharging Flow Rate

In the case of no control, illustrated by the green curve, the flow rate at the bottleneck drops to around 3380 veh/hr due to the lane closure and capacity drop. When the incident is cleared, the flow rate increases directly to 7500 veh/hr. Then, it concentrates around 6000 veh/hr. When the VSL controller is applied, illustrated by the blue curve, the flow rate decreases to around 4500 veh/hr, which is higher by 33% compared to that in the no control case during the time of the incident. It then increases to approximately 6000 veh/hr after the incident is over.

As seen in Figure 104, the density of the discharging section (green curve) starts increasing due to the incident from around 100 to 350 $veh/mile$ since no control action is applied.

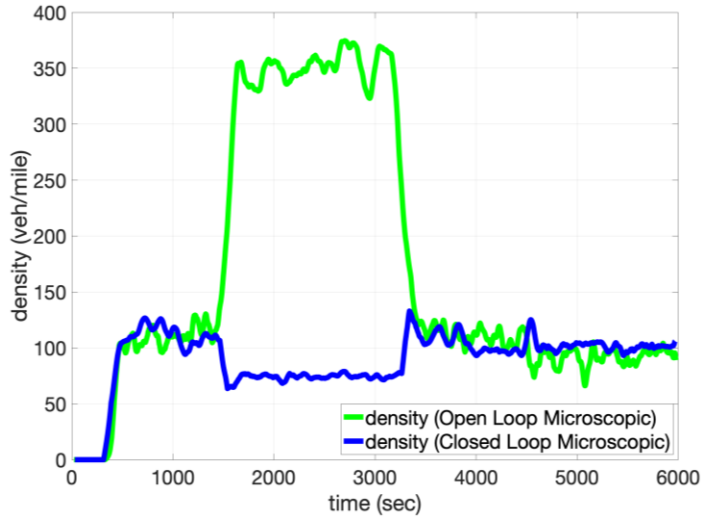


Figure 104. Density of Discharging Section

Then the density stays around 350 $veh/mile$ until the incident is removed. When the VSL control is applied, the density (blue curve) is stabilized around the preset density value $\rho^* = 75$ $veh/mile$, which is much lower than the scenario without control.

The definition that the authors used in (Y. Zhang et al. 2018) to track the number of vehicles lined up at the entrance of the controlled sections is used here to measure the queue length. Let Q represent the number of vehicles in the queue. Using the flow conservation equation, we have $\dot{Q} = d - q_1$. Note that Q only tracks the length of the upstream queue. Therefore, the stability of the closed-loop system is not affected.

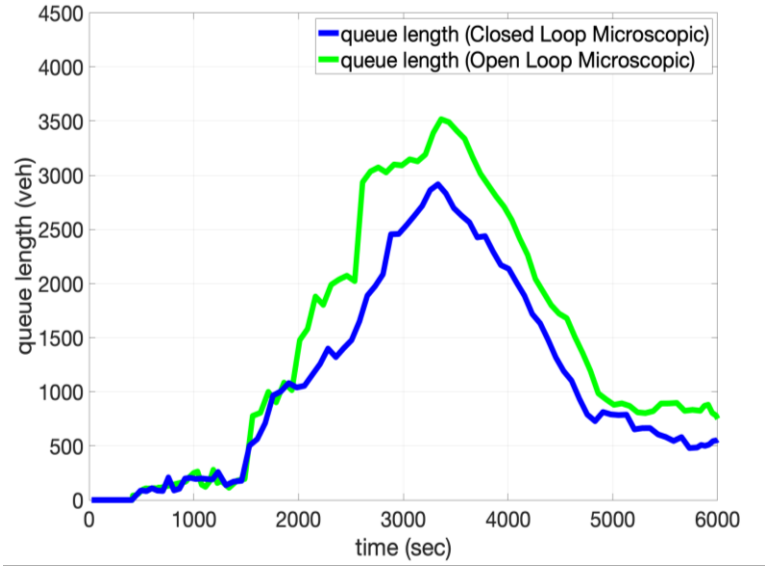


Figure 105. Growth and Discharge of the Queue

Figure 105 shows the time evolution of the queue length for both open-loop and closed-loop systems, represented by the green and blue curves, respectively. In the case of closed-loop system, the queue length increases less rapidly, reaches a lower maximum and decreases fast when the incident is removed compared to the no control case.

Conclusion

Based on the first-order macroscopic Cell Transmission Model (CTM) of traffic flow, we design, analyze, and evaluate the performance of several integrated highway traffic flow control strategies in both macroscopic and microscopic simulations. We discover that the unmanaged forced lane-changing behavior in the vicinity of highway bottlenecks is a major reason for the capacity drop phenomenon. Motivated by that, we propose a Lane Change (LC) controller that provides lane change recommendations to upstream vehicles in order to avoid the capacity drop. Two types of Variable Speed Limit (VSL) controllers, together with the LC controller, are designed and demonstrated improvements in traffic mobility, safety, and environmental impact at highway bottlenecks. The combined LC and feedback linearization VSL controller show analytically that the trajectories of the closed-loop system converge to the desired equilibrium point, which corresponds to the maximum possible throughput.

In addition, the combined LC and VSL controller is extended to coordinate with ramp metering controllers. As demonstrated by the Monte-Carlo microscopic simulations, the coordinated VSL, RM, and LC controller has significantly improved the system performance, maintained the queue length of on-ramps, and fulfilled the fairness between the mainline traffic flow and on-ramp flows. Then, the proposed controller was compared to the widely used MPC controller. The performance and robustness of both integrated control schemes are compared with respect to perturbations on traffic demand, model parameters, and measurement noise. The simulation results confirm that both controllers are able to improve the total time spent (TTS). Though the MPC follows an optimization-based routine, the FL-based VSL guarantees exponential stability with negligible computational effort and similar robustness.

Furthermore, the most updated version of the cell transmission traffic flow model (CTM), which considers the capacity drop phenomenon and the bounded acceleration effects, is modified to include a constant disturbance term, which accounts for the uncertainties in the model due to the inevitable parametric modeling and measurement errors. The open-loop stability properties of the modified CTM are investigated under all possible traffic flow scenarios. The open-loop stability analysis shows that the disturbance directly influences the location of the equilibrium points on the fundamental diagram, resulting in triggering the capacity drop phenomena much earlier. To reject the disturbance, therefore, while improving the throughput, a robust VSL controller is introduced, and the stability properties of the closed-loop system are analyzed. Our findings have shown that the proposed controller guarantees that the traffic density of a considered highway road section converges asymptotically with time to a predetermined desired density located in the free-flow region of the fundamental diagram. Both macroscopic and microscopic simulation results demonstrate that the robust control scheme is able to stabilize the density of the road section at the desired density, and, as a result, it improves the discharging flow rate by 33%, compared to the case of no control action.

Intelligent Parking Assist System

Filipe Vital, Department of Electrical and Computer Engineering, University of Southern California
Petros Ioannou, Department of Electrical and Computer Engineering, University of Southern California

Introduction

It was estimated that, in the year of 2013, trucks were responsible for carrying around 70% (in weight) of USA's total freight shipments, without considering multimodal shipments that use trucks at some point (U.S. Department Of Transportation 2015). It is expected that this value will still be as high as 66% by the year of 2040, despite substantial increases in multimodal and rail shipments (U.S. Department Of Transportation 2015). This shows just how important trucks are to the USA economy. However, the increasing demand for trucks comes with a need for supporting infrastructure and legislation, in particular, appropriate parking infrastructure. In 2015, a survey by the Federal Highway Administration identified truck parking shortages in 36 US states, with more pronounced issues along major trade corridors (U.S. Department of Transportation 2015). The lack of truck parking can have significant impact on road safety, industry costs, and the environment (Rodier et al. 2010; Sochor and Mbiydzenyuy 2013), and ranked among the trucking industry's top concerns in recent surveys by the American Transportation Research Institute (ATRI) (American Transportation Research Institute 2019). As expanding the infrastructure would require significant capital investment, this issue points to the need for better utilization of the existing truck parking capacity.

One of the requirements for efficient utilization of the truck parking capacity is having enough information about the truck parking system. Recently, the topic of smart parking systems has become increasingly popular, with research on parking occupancy sensing and information distribution systems for both passenger vehicles and trucks (T. Lin, Rivano, and Le Mouel 2017; de Almeida Araujo Vital, Ioannou, and Gupta 2020). Nevertheless, drivers' direct usage of occupancy information can only help with last minute adjustments to the schedule and might not be as helpful in high-demand areas where all rest areas might be full during peak hours. Parking availability information could be used more effectively if included in the earlier stages of planning, when the route and schedule are decided. Ideally, an integrated planning system with access to data on the location and predicted availability of all rest areas would be able to choose an optimal route and schedule, such that: working hours regulations are satisfied; and off-duty time is scheduled only at locations that are expected to have available parking at the time of arrival. This type of system is the subject of this research.

An important aspect of truck parking availability is its time dependence. Although drivers often report difficulties finding parking and truck stops report operating overcapacity, they usually refer to the period between 7 PM and 5 AM when drivers are looking for overnight parking (U.S. Department of Transportation 2015; NCDOT 2017; Boris and Brewster 2018; Martin and Shaheen 2013). This suggests that it may be possible to mitigate the truck parking shortage by encouraging drivers to plan their stops on off-peak periods, thus balancing the demand.

Balancing the demand with respect to space may also be an option. Several factors influence drivers' choice of parking location; however, lack of information can lead drivers to park illegally even when there are facilities with available parking nearby (Rodier et al. 2010). In (U.S. Department of Transportation 2015), less than 50% of truck stops reported working overcapacity. Many of the facilities operating under capacity may be in regions with low parking demand, yet, some may be in high-demand areas and are underutilized because drivers do not know they are viable options. With this in mind, it is worth considering to include parking availability earlier in the planning process. Instead of only providing drivers information on facilities along their path to allow for on-trip decisions, if parking information is considered when choosing the path itself, the existing parking capacity could be better utilized. Depending on the situation, it may be advantageous for the driver to take a longer route if that can guarantee available parking. This suggestion of addressing parking issues at the supply chain level has been brought up before in (Phelan et al. 2016); however, literature on the topic is still scarce. Previous studies focused on developing methods to estimate truck parking demand, predict parking occupancy, and measure parking occupancy for parking management systems. Some pilot projects distribute real-time availability information through websites or variable-message signs on nearby highways. However, judging how to best use the information, if at all, is up to the drivers. These systems can only assist drivers with small on-trip schedule adjustments as using all this information to generate regulation-compliant itineraries is not trivial, and most systems do not include an occupancy prediction feature either.

Another interesting topic to be considered is how this issue affects battery electric trucks (BET). Among the concerns regarding heavy-duty vehicles electrification are the reduced range, longer recharge time and the reduction in maximum payload due to added battery weight. However, it is often overlooked that commercial drivers are required to stop and rest regularly regardless of the vehicle's range. Even if long-haul trucks drive on average 600 miles per day (Smith et al. 2019), they are required to stop for at least 30 minutes every 8 hours and for at least 10 hours after 11 hours of driving time. Adequate infrastructure and efficient trip planning can mitigate the range issue by recharging the vehicle during mandatory stops. Nevertheless, recharging does take longer than refueling and charging stations are not abundant, making charging station availability an issue for BETs.

Truck parking is already considered a critical issue for the trucking industry, and the similar issue of recharge station availability and electric vehicles' range are obstacles to the adoption of heavy-duty BETs. Therefore, this project focuses on the following topics:

1. How to integrate parking availability information into the planning of truck drivers' itineraries?
2. How are trip cost and duration affected by stricter parking restrictions? Is it financially interesting to the industry?
3. What is the effect of working hour regulations and parking restrictions on the economic viability of battery electric trucks?

To answer the above questions, this study addresses the 'Truck Driver Scheduling Problem under Parking Availability Constraints' (TDSP-PA) and its variants. The TDSP-PA, presented in the section Truck Driver Scheduling Problem, aims to calculate optimal schedules such that they are regulation-compliant and guarantee appropriate parking at required stops. Section Shortest Path and Truck Driver Scheduling Problem extends the TDSP-PA to include path planning, studies how the optimal route is affected by parking conditions, and the potential costs of disregarding parking shortages. Section Long Haul Battery Electric Truck Planning extends the problem to electric trucks and studies the effect of parking constraints on electric trucks' performance. Additional details on TDSP extensions are also provided in Appendix H.

Truck Driver Scheduling Problem

Section based on the publication:

F. Vital and P. Ioannou, "Long-Haul Truck Scheduling with Driving Hours and Parking Availability Constraints," in 2019 IEEE Intelligent Vehicles Symposium (IV), vol. June, jun 2019, pp. 620–625.

Introduction

According to the U.S. Department of Transportation (USDOT), 36 states are experiencing shortages in rest areas (U.S. Department of Transportation 2015). Truck parking shortages can lead to illegal parking, drowsy driving and prolonged time looking for parking, affecting negatively drivers' safety, operating costs and the environment. As expanding the infrastructure would require significant capital investment, this issue points to the need for better utilization of the existing truck parking capacity.

One of the requirements for efficient utilization of the truck parking capacity is having enough information about the truck parking system. Recently, the topic of smart parking systems has become increasingly popular, with research on parking occupancy sensing and information distribution systems for both passenger vehicles and trucks (Bayraktar et al. 2015; Morris et al. 2018). Nevertheless, the direct usage of the occupancy information by the drivers can only help with last minute adjustments to the schedule and might not be as helpful in high-demand areas where all rest areas might be full during peak hours. Parking availability information could be used more effectively if included in the earlier stages of planning, when the route and schedule are decided. Ideally, an integrated planning system with access to data on the location and predicted availability of all rest areas would be able to choose an optimal route and schedule, such that the schedule satisfies the driving hours regulations and only schedules stops at locations that are expected to have available parking at the time of arrival.

The truck driver scheduling problem (TDSP) under HOS regulations has been studied both as a part of a vehicle routing problem (Kok, Hans, et al. 2010; Asvin Goel and Irnich 2017) and by itself (A. Goel 2010; Asvin Goel and Kok 2012; Asvin Goel 2012; Koç et al. 2016; Kok, Hans, and Schutten 2011; Archetti and Savelsbergh 2009). Different versions of the TDSP were considered, the main differences being: parking restrictions, the HOS regulation considered, and if the solution is optimal or only feasible. In (Archetti and Savelsbergh 2009; A. Goel 2010; Asvin Goel and Kok 2012; Asvin Goel and Irnich 2017), the authors do not consider specific parking locations, treating the problem as if trucks could park at any point along the route, which is not true in practice. In some cases, the authors restrict parking to client locations (Kok, Hans, and Schutten 2011; Asvin Goel 2012; Kok, Hans, et al. 2010), and propose to model rest areas as customer locations with zero service time and an unbounded time-window. In (Asvin Goel 2012), a mixed integer programming (MIP) model to the TDSP was presented. This model restricts parking to client locations and considers the common types of restrictions present in HOS regulations, allowing it to model or approximate different regulations. Similar MIP models were used in (Kok, Hans, and Schutten 2011; Koç et al. 2016). In (Koç et al. 2016), Koç includes an environmental impact factor dependent on the types of idling used in each stop. Parking is restricted to rest areas, but they are assumed to be always available. In (Kok, Hans, and

Schutten 2011), Kok studied the problem of optimizing the departure time and schedule with time-dependent travel times. Although many methods already restrict parking to suitable areas such as customer sites or rest areas, the possibility of the rest areas being full at certain times is not considered. Furthermore, the existing work usually limits the planning horizon and/or the total on-duty time to avoid the rules that regulate longer trips.

In this paper we focus on the issues of parking availability and long trips HOS regulations. Our first contribution is introducing the TDSP with parking availability, which is a variant of the TDSP that assumes that parking locations are subject to availability constraints. The second contribution is to include the USA HOS rule for long trips, *i.e.,* trips with more than 60 hours of on-duty time, in the MIP model for the TDSP. This paper is organized as follows: section USA's Hours of Service Regulations introduces the HOS regulation that was considered; Problem Description presents the problem description; section Model describes the MIP model; section Experiments describes the experiments and results; and section Conclusion presents the conclusion and future work.

USA's Hours of Service Regulations

The current USA HOS regulation (Code of Federal Regulations, n.d.) differentiates between driving time, on-duty time and off-duty time. In summary, driving time is the time spent operating the truck, on-duty time is the time from when the driver is required to be ready for work until he/she is relieved from work, and off-duty is the time when the driver is not on-duty. The time restrictions set by the regulation can be reset by an off-duty periods with minimum durations specified in the regulation. We refer to off-duty periods lasting at least 0.5, 10, and 34 consecutive hours, as *breaks*, *daily rests* and *weekly rests*, respectively. Note that the longer off-duty periods can be used to reset the restrictions related to the shorter ones. The USA HOS regulation can be summarized as follows:

- Daily Driving Time Limit: A driver may drive at most 11 hours between 2 consecutive *daily rests*.
- 14-Hour Elapsed Time Limit: A driver cannot drive after 14 hours have elapsed since the last *daily rest* ended.
- Rest Breaks: A driver cannot drive after 8 hours have elapsed since the last *break* ended.
- 60-Hour Limit: A driver cannot drive after having been on duty for 60 hours in any period of 7 consecutive days. The 7 days period can be reset by taking a *weekly rest*.

Problem Description

We consider the problem of scheduling the rest stops for a long-haul truck trip with a known route and a single client while taking into account the USA HOS regulations and estimated parking availability windows for all rest areas along the route. It is assumed that the rest areas are located on the route and require no detours to be accessed. The parking availability time-windows are assumed known. The route has $n + 1$ nodes, 2 of which are the origin, node 0, and destination of the truck, node n. The other $n - 1$ are rest areas located along the route. For each node $i \in \{0,1, \dots, n\}$ the variable $x_i = (x_{i,a}, x_{i,d})$ represents the arrival and departure

times of the truck at that node. Each rest area i has T_i parking availability time-windows $\left[t_{i,\tau}^{min}, t_{i,\tau}^{max}\right]$, where $\tau \in \{1,2,\ldots,T_i\}$ indicates the time-window's index. The time-windows restrict the arrival time at that node and are only in effect when the truck has to stop at that specific node, driving by it is not constrained by the time windows. For each location and time-window, a binary variable $y_{i,\tau}$ represents if that specific time window is being used (yes:1, no:0). Driving by without stopping is represented by the variable $y_{i,0}$ (drive by:1, stop:0). The travel time $d_{i,i+1}$ in between nodes is considered known and independent of time. The planning horizon is denoted by t_{hor}. The driver must reach its destination before the specified planning horizon. Figure 106 shows an example of a route with origin v_0, 3 rest areas v_1, v_2 and v_3 with 3 time-windows each, and a destination v_4 also with 3 time-windows.

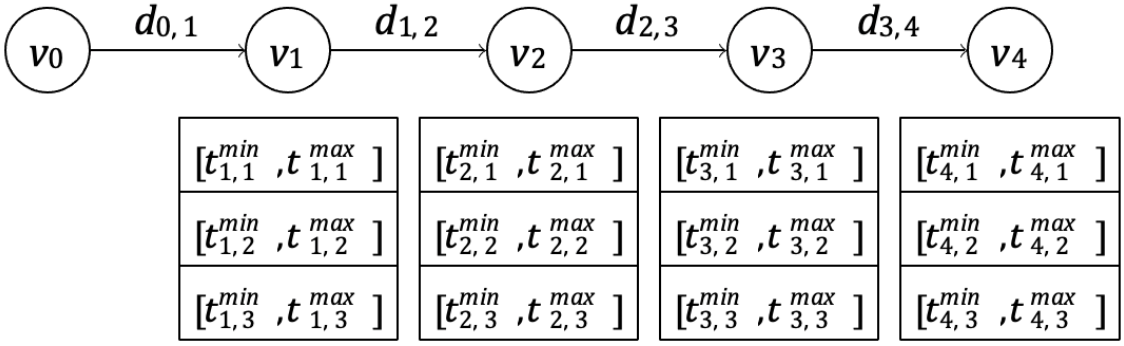


Figure 106. Simple route with 5 locations (origin, 3 rest areas and destination) with 3 time-windows each.

The schedule must comply with the HOS regulations described in the section USA's Hours of Service Regulations. R is defined as the set of different types of rest period described in the regulation. For each $r \in R$, t_r defines the minimum duration of that type of rest period. C is the set of constraints imposed by the regulation. $C_1 \subseteq C$ is the set of constraints controlling the maximum elapsed time between off-duty periods. $C_2 \subseteq C$ is the set of constraints controlling the maximum accumulated driving time between off-duty periods. $C_3 \subseteq C$ is the set of constraints controlling the maximum accumulated on-duty time during a rolling time-window; the width of the time-window for a constraint $c \in C_3$ is represented by δ_c. In the USA regulation δ_c is 7 days, so these rolling time-window constraints will be referred to as weekly constraints. For each constraint $c \in C$, t_c is the time limit imposed by the regulation and $R_c \subseteq R$ is the set of rest types that can reset this counter. The binary variable $z_{i,r}$ indicates whether a rest of type r is taken at location i (yes:1, no:0). The driver cannot take more than 1 type of rest at the same location. If no type of rest is schedule for a rest area, the driver cannot stop there. The departure time from the origin must be within the interval $\left[t_0, t_{dep}\right]$. It is assumed that the driver has been off-duty for long enough before the departure time, so that all constraints' counters are reset before departure. Table 21 lists all the variables and parameters used in the model, some of which are defined in the following section.

Table 21. Variables & Parameters

Variables	
Symbol	**Description**
$x_{i,a}, x_{i,d}$	Arrival/Departure times from location i
$y_{i,\tau}$	Used time-window τ at location i?
$y_{i,0}$	Drove by location i?
$z_{i,r}$	Rest of type r was taken at location i?
$\lambda_{i,j,c}$	Accumulated driving time from a trip departing location i at time $x_{j,a}$, relative to constraint $c \in C_3$
$\psi_{i,j,c}$	Accumulated driving time from trips departing locations 0 to i at time $x_{j,a}$, relative to constraint $c \in C_3$
α_*, β_*	Auxiliary variables for ramp constraints
Parameters	
Symbol	**Description**
Ti	Number of time-windows at location i
$t_{i,\tau}^{min}, t_{i,\tau}^{max}$	Lower/Upper limit of τ-th time-window at location i
R	Set of rest types defined in the regulation
C	Set of constraints defined in the regulation
t_c	Time limit related to constraint $c \in C$
δ_c	Rolling time-window's width for constraint $c \in C_3$
R_c	Set of rest types that can reset constraint $c \in C$
t_r	Minimum duration for rest of type $r \in R$
$d_{i,i+1}$	Travel time between location i and $i + 1$
t_{hor}	Planning time horizon
t_{dep}	Maximum departure time from the origin

Model

A MIP model for the TDSP under HOS regulations has been proposed by Goel in (Asvin Goel 2012). This model considers that the prediction horizon is limited to 1 week, that drivers may rest at client locations, and that parking locations are always available. Our model considers similar HOS regulations, but restricts parking to rest areas with available parking spaces, and includes the USA regulations for longer trips. Section Parking Availability Constraints presents the model used for short trips which includes the parking availability. The weekly constraints are introduced on section Weekly Constraints.

Parking Availability Constraints

Other formulations model all nodes in the network as required stops with client locations having one or more time-windows constraints, and parking locations having a single unbounded time-window. In this model we aim to address the issue of parking availability, so these past

approaches are not directly applicable. We model parking availability as time-windows constraints for each parking location. However, unlike client locations, the rest areas are not required stops, and if the driver is not stopping at a certain location, there is no need to restrict the schedule with its parking availability. Therefore, the time-windows must be conditioned to the scheduling of off-duty periods at the rest areas. The formulation is as follows:

$$\text{Minimize Total travel time} = x_{n,a} - x_{0,d} \tag{70}$$

Subject to:

$$x_{i,d} + d_{i,i+1} = x_{i+1,a}, \qquad \forall 0 \le i \le n-1 \tag{71}$$

$$x_{i,a} + \sum_{r \in R} t_r \, z_{i,r} \le x_{i,d}, \qquad \forall 1 \le i \le n \tag{72}$$

$$x_{i,d} \le x_{i,a} + \left(1 - y_{0,\tau}\right) t_{hor}, \qquad \forall 1 \le i \le n \tag{73}$$

$$y_{i,0} + \sum_{\tau=1}^{T_i} y_{i,\tau} = 1, \qquad \forall 1 \le i \le n \tag{74}$$

$$\sum_{\tau=1}^{T_i} y_{i,\tau} = \sum_{r \in R} z_{i,r}\,, \qquad \forall 1 \le i \le n-1 \tag{75}$$

$$\sum_{\tau=1}^{T_i} y_{i,\tau} \, t_{i,\tau}^{min} \le x_{i,a}, \qquad \forall 1 \le i \le n \tag{76}$$

$$x_{i,a} \le t_{hor} - \sum_{\tau=1}^{T_i} \left[y_{i,\tau} \left(t_{hor} - t_{i,\tau}^{max} \right) \right], \qquad \forall 1 \le i \le n \tag{77}$$

$$x_{k,a} - x_{i,d} \le t_c + t_{hor} \sum_{j=i+1}^{k-1} \sum_{r \in R_c} z_{j,r}\,, \qquad \forall 0 \le i < k \le n, c \in C_1 \tag{78}$$

$$\sum_{j=i}^{k-1} d_{j,j+1} \le t_c + t_{hor} \sum_{j=i+1}^{k-1} \sum_{r \in R_c} z_{j,r}\,, \qquad \forall 0 \le i \le k \le n, c \in C_2 \tag{79}$$

$$x_i \in [0, t_{hor}]^2, y_i \in \{0,1\}^{T_i+1}, z_i \in \{0,1\}^{|R|}, \qquad \forall 1 \le i \le n \tag{80}$$

$$x_{0,d} \in \left[0, t_{dep}\right], \qquad y_{n,0} = 0 \tag{81}$$

The objective function (70) is set to minimize the total trip duration. Constraint (71) guarantees that the arrival time equals the departure time of the previous location plus the driving time. Constraint (72) states that the vehicle must not depart before the arrival time plus the minimum rest time decided for that location. Constraint (73) controls what happens when the driver does not stop at a certain location. If the vehicle does not stop at location i, the arrival time equals the departure time. This constraint works with constraints (72), (74) and (75) to

assure this. Equality will hold when $y_{i,0} = 1$. If $y_{i,0} = 0$, then constraint (73) is always true as t_{hor} is large. Constraint (74) states that at any location, either exactly 1 time-window is used or the vehicle does not stop. Constraint (75) states that the driver only stops if an off-duty period is scheduled. Constraints (76) and (77) check the time-windows. Arrival must happen after the beginning and before the end of the chosen time window. Constraint (78) checks that the time elapsed since the last rest in $R_c, c \in C_1$ is less than t_c. Constraint (79) checks if the accumulated driving time between rest periods in $R_c, c \in C_2$ is less than t_c. Constraint (80) sets the variables' domains, and (81) guarantees that the departure time from the origin is within the required period and that the vehicle will stop at the destination.

Weekly Constraints

Some authors limit the planning horizon to less than a week and model the weekly constraint as an accumulated driving time constraint over the whole trip (Koç et al. 2016) or between 2 consecutive weekly rests (Asvin Goel 2012). The formulation in (Koç et al. 2016) is unable to deal with trips requiring on-duty time above the weekly limit. Without the planning horizon assumption, the model in (Asvin Goel 2012) loses its guarantee of optimality due to not considering the different structure of the USA regulation, but still guarantees a valid schedule. In this model, the weekly constraint c_w can be defined as a constraint of the set C_2 as follows: $c_w \in C_2$, $t_{cw} = 60$, $R_{cw} = \{\textit{weekly rest}\}$. We will refer to this formulation as *Simplified* model.

In the USA, this weekly constraint is defined as a 7-day rolling time-window in which the driver cannot drive after working for 60 hours. This restriction can be modeled as:

$$\lambda_{i,c}(t) = R\big(t - x_{i,d}\big) - R\big(t - x_{i+1,a}\big) - R\big(t - x_{i,d} - \delta_c\big) + R\big(t - x_{i+1,a} - \delta_c\big), \quad \forall 0 \le i \le n-1, c \in C_3 \tag{82}$$

$$\lambda_c(t) = \sum_{i=0}^{n-1} \lambda_{i,c}(t), \quad \forall c \in C_3 \tag{83}$$

$$\lambda_c(t) \le t_c \quad \forall t \in \{x_{1,a}, x_{2,a}, \dots, x_{n,a}\},\ c \in C_3 \tag{84}$$

where $R(t)$ is the unit ramp function. $\lambda_{i,c}(t)$ represents the accumulated driving time generated by the displacement between locations i and $i+1$ at time t, and $\lambda_c(t)$ represents the accumulated driving time over the last δ_c hours at time t, both relative to constraint $c \in C_3$. It is sufficient to check these constraints at the arrival times x_i, a. If the constraints are broken anywhere they will also be broken at the arrival time that follows.

MIP formulation

The accumulated driving time over the last δ_c hours, $\lambda_c(t)$, needs to be evaluated at all arrival times $x_{j,a}$, so each of its component functions $\lambda_{i,c}(t)$ must also be evaluated at these times. Constraints were defined for each evaluated time using the method for writing piecewise linear functions in MIP models described in (Schoomer 1964). The domains of the functions $\lambda_{i,c}(t)$ are

divided in sections according to when the slope of the function changes and auxiliary variables are used to write t according to where it is located relative to the sections' boundaries. For each function and evaluation time $\lambda_i(x_{j,a})$, the sets of variables $\{\alpha_{i,j,p}\}$, $\{\beta_{i,j,q}\}$, and $\{\lambda_{i,j}\}$ are defined as follows:

$$\alpha_{i,j,p,c} \in 0,1,\ \beta_{i,j,q,c} \in [0,1] \qquad \forall 0 \le i < j \le n,\ 0 \le p \le 4,\ 1 \le q \le 5,\ c \in C_3 \tag{85}$$

$$1 \ge \alpha_{i,j,0,c} \ge \beta_{i,j,1,c} \ge \alpha_{i,j,1,c} \ge \cdots \ge \alpha_{i,j,4,c} \ge \beta_{i,j,5,c}, \quad \forall 0 \le i < j \le n, c \in C_3 \tag{86}$$

$$\alpha_{i,j,p,c} < \beta_{i,j,p+1,c} + 1, \qquad \forall 0 \le i < j \le n, c \in C_3 \tag{87}$$

$$\begin{aligned} x_{j,a} = x_{i,d}\beta_{i,j,1,c} + d_{i,i+1}\beta_{i,j,2,c} + (\delta_c - d_{i,i+1})\beta_{i,j,3,c} + d_{i,i+1}\beta_{i,j,4,c} \\ + t_{hor}\beta_{i,j,5,c}, \qquad \forall 0 \le i < j \le n,\ c \in C_3 \end{aligned} \tag{88}$$

$$\lambda_{i,j} = d_{i,i+1}\beta_{i,j,2} - d_{i,i+1}\beta_{i,j,4}, \qquad \forall 0 \le i < j \le n \tag{89}$$

$$\sum_{i=0}^{j-1} \lambda_{i,j} \le t_c, \qquad \forall 1 \le j \le n,\ c \in C_3 \tag{90}$$

where $\lambda_{i,j} = \lambda_i(x_{j,a})$. The α's and β's are auxiliary variables used to model the piecewise definition of $\lambda_{i,c}(t)$. The α's determine in which section of the function domain t is, and the β's define its exact position within the section, the indexes p and q represent the sections. Constraints (85), (86) and (87) imply that, for a section q, whenever $0 < \beta_{*,q} < 1$, then $\beta_{*,q^-} = 1, \alpha_{*,q^-} = 1, \quad \forall q^- < q$, and $\beta_{*,q^+} = 0, \alpha_{*,p} = 0, \quad \forall q^+ > q, p \ge q$. Constraint (88) writes the time instant to be evaluated, $x_{j,a}$, as a function of the α's and β's. Constraint (89) uses the β's to calculate $\lambda_i(x_{j,a})$, and constraint (90) calculates, and limits, the accumulated driving time over the moving time-window relative to regulation $c \in C_3$. This set of constraints substitutes constraints (82), (83) and (84), and guarantees that the accumulated driving time in any period of δ_c consecutive hours is kept below t_c. Due to (88) this problem would be a quadratically constrained problem. However, as (88) only considers $j > i$, the variables $\alpha_{i,j,p,c}$ for $p < 2$ and $\beta_{i,j,q,c}$ for $q < 3$ will be always 1 and can be defined as constants. This model still does not include the possibility of using *weekly rests* to reset the constraint, so it will be referred to as *No Reset* model.

Reset for weekly constraint

According to USA's regulation, a driver may restart the 168 consecutive hours (7 days) period, by taking an *weekly rest*. When this *weekly rest* is taken the system should be able to set the weekly accumulated driving time at the end of that rest to zero and start counting again from there. This was implemented using indicator constraints controlled by the variables $z_{i,r}$. A set of

variables $\{\psi_{i,j}\}$ was created to represent the accumulated driving time generated by all trips starting at locations $\{0, \dots, i\}$ measured at time x_j, a. The formulation is as follows:

$$\psi_{i,j} = \begin{cases} \psi_{i-1,j} + \lambda_{i,j} & \text{if } \sum_{r \in R_c} z_{i,r} = 0 \\ \lambda_{i,j} & \text{if } \sum_{r \in R_c} z_{i,r} = 1 \end{cases} \qquad \forall 1 \le i < j \le n,\ c \in C_3 \qquad (91)$$

$$\psi_{0,j} = \lambda_{0,j}, \qquad \forall 1 \le i < j \le n \qquad (92)$$

$$\psi_{j-1,j} \le t_c, \qquad \forall 1 \le j \le n,\ c \in C_3 \qquad (93)$$

where constraint (91) defines $\psi_{i,j}$ and sets to zero all contributions from nodes before location i when an appropriate rest is taken at location i. This model assumes that all constraints' counters are reset before departure, so (92) sets the initial accumulated driving time to zero. These 3 constraints replace constraint (90). In the regulation considered, only *weekly rests* can be used to reset this constraint, so $|R_c| = 1$ for $c \in C_3$. In this case the conditions turn into $z_{i,r} = 1$ and $z_{i,r} = 0$. This model will be referred to as *Reset* model.

Experiments

Parking Availability Impact

This section describes the experiment used to test the impact of considering availability windows for every parking lot along a truck route. A model without the parking availability constraints was used as baseline for comparison. A route, approximately 1960Km long, going from San Diego to Seattle using the I-5 highway was chosen. Data from the FHWA (U.S. Department Of Transportation 2013) was used to find rest areas and truck stops located close to the route and position them along the route; 94 truck stops and rest areas were considered. Figure 107 shows the parking lots along the route (gray circles), as well as the chosen parking locations for the base case (triangles and squares) and for one of the tested scenarios (crosses).

This trip requires less than 60 hours of on-duty time, so the rolling time-window constraints are not needed. In order to simulate parking availability, time windows with start and end times normally distributed were considered for each rest area/truck stop. The distribution used for the start times had mean 5 hours (5am) and standard deviation of 0.5 hours, and the one for the end time had mean 20 hours (8pm) and standard deviation of 1 hour. For this experiment, a 100 scenarios with different parking availability time-windows were generated. It was considered that the final destination has daily time-windows from 8am to 6pm. The day was divided in 8 3-hours long intervals, and these intervals were used as departure constraints. Both models were solved for each pair of departure constraints and scenario using the solver CPLEX.

This experiment compared the feasibility and the average trip duration of solutions generated by the two models. As the scenarios do not affect the baseline model, only one solution was generated for each departure interval, the solution's feasibility was then evaluated in each of the 100 scenarios. Our model generates different solutions for each scenario, so the average

cost was used for comparison. Our model includes additional constraints in the problem, and does not relax the existing ones. Consequently, it is impossible for our model to reduce the optimal cost (trip duration). The advantage of our model lies in the practical feasibility of its solutions. As can be seen on Figure 108, the schedules generated without considering parking are often infeasible, meaning that they scheduled rest stops at locations without available parking. Even if rules of thumb were used to choose departure constraints that improve feasibility, such measures can have a significant impact on the total duration/cost of the trip, as seen on Figure 109. When the departure was restricted to earlier times the feasibility improved, however the cost deteriorated. Figure 109 also shows that the average cost of our model's solution is not significantly higher than the baseline model. Therefore, our model is able to guarantee the feasibility of its schedules with only a small impact to the cost. In fact, in this experiment, if we do not restrict the departure time, the average cost of our solutions is practically equal to the optimum cost of the baseline model.

It is also important to note that these costs for the baseline model were generated as if the driver did not incur any penalty for not finding parking. In practice the driver would have to either keep driving looking for parking or park illegally somewhere nearby. In the first case, the search time itself would already increase the trip duration, and it might cause other changes in the rest of the schedule, further increasing the cost. In the second case, the driver is subject to the possibility of being fined and to higher safety risks. The estimation of the extra costs incurred by infeasible schedules, in particular the second case, is non-trivial and was not treated in this paper.

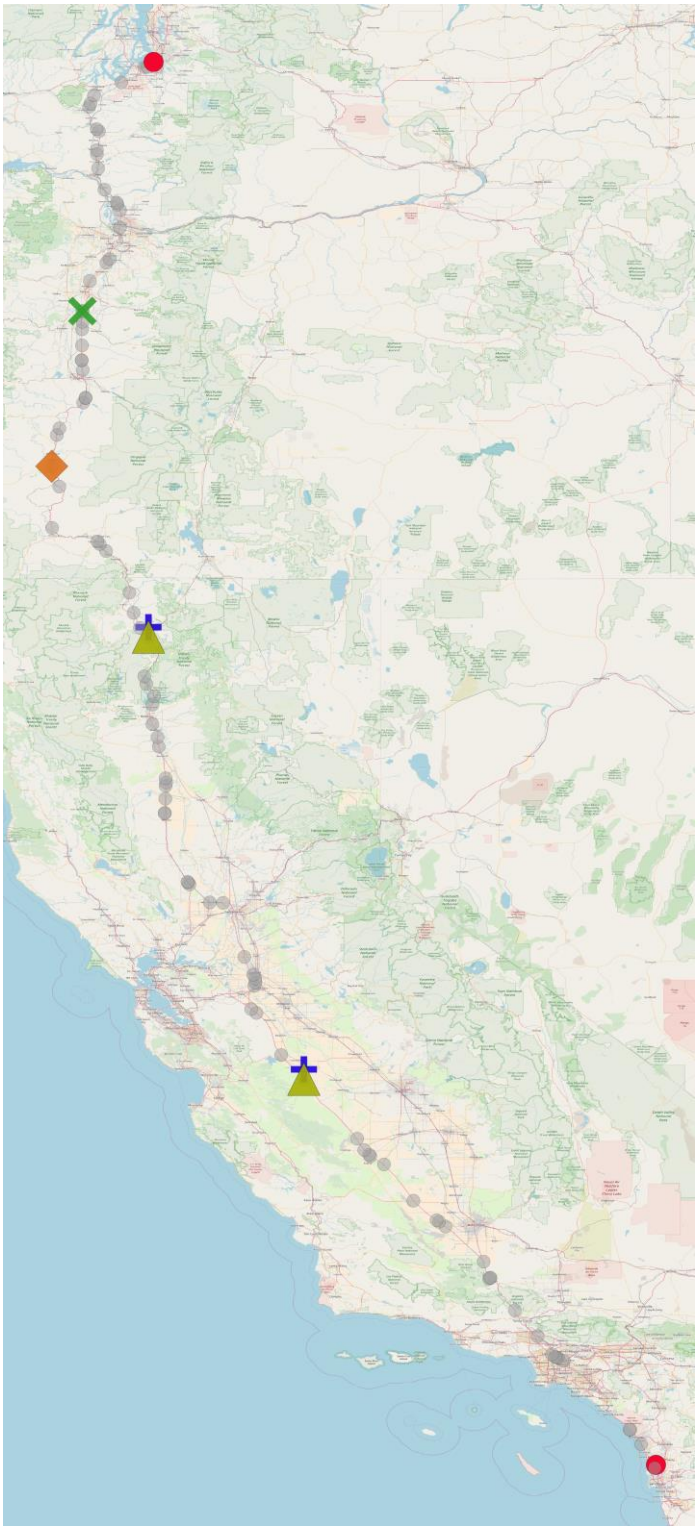


Figure 107. Route used on short trip experiment. San Diego to Seattle through the I-5 freeway. The triangles (base model) and +s (new model) represent truck stops chosen for daily rests, and the square (base model) and × (new model) represent the ones chosen for short breaks. The gray circles represent the truck stops near the chosen route.

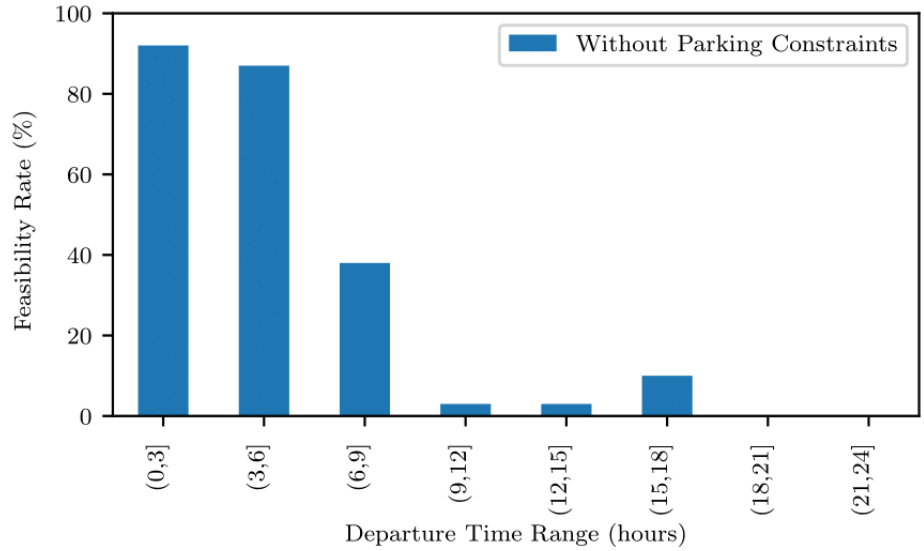


Figure 108. Feasibility rate of the schedules generated without considering the parking constraints. The feasibility rate of schedules that consider parking constraints is always 100%, so it was omitted.

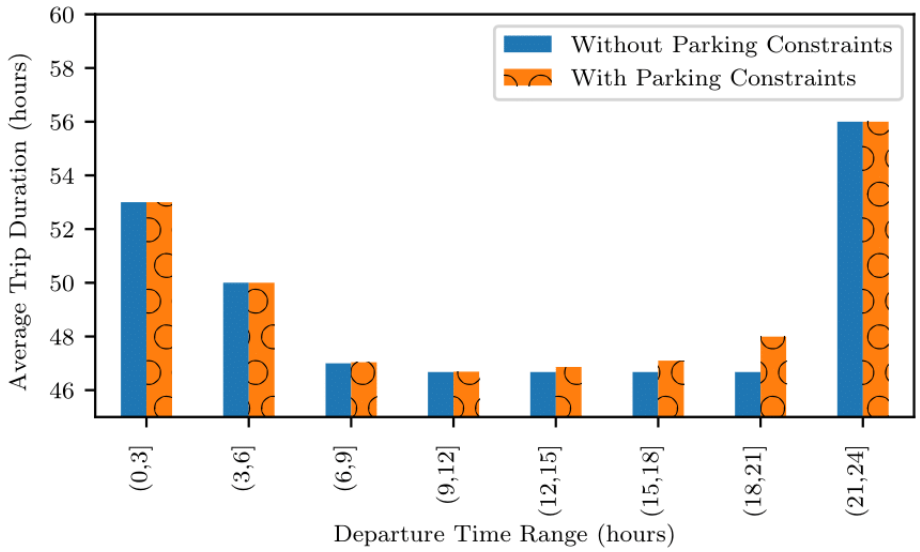


Figure 109. Average trip duration of schedules generated with and without parking constraints.

Long Trips

In order to test the long trip models, a route was generated with equally spaced truck stops, the travel time between two adjacent truck stops was set to 1 hour. Like in the previous experiment, normal distributions were used to generate time-windows for each truck stop. The distribution used for the start times had mean 4 hours (4am) and standard deviation of 1 hour, and the one for the end time had mean 21 hours (9pm) and standard deviation of 2 hours. It

was considered that the truck must depart from the origin during the first 24 hours and that the final destination has daily time-windows from 8am to 6pm.

This experiment tested the performance of the 3 models, *Reset*, *No Reset*, and *Simplified*, which were presented on section Model.

Performance It can be seen on Figure 110 that when the reset option is not implemented the total trip duration automatically increases to more than 1 week when the needed driving time is larger than the weekly limit. When the reset is implemented the trip duration only increases by the duration of the weekly rest needed to reset the counter. This is the reason why the *Simplified* model is likely to find an optimal schedule for the USA regulations. In general, it is more efficient for the drivers to take the 34 hours rest and reset the counter than to reduce their average daily driving hours to match the rolling time-window. However, this is not necessarily true for every scenario. In a scenario with more restrictive time constraints, the need to extend offduty periods beyond the minimum required in order to meet said constraints would lower the average daily driving hours, making *weekly rests* not as advantageous. Figure 65 also shows, as 'TW Bound', the lower bound for solutions that need the rolling time-window to be found, i.e., solutions that assign at least 1 period longer than 168 hours without a *weekly rest* and with more than 60 hours of driving time. This lower bound was obtained by solving this problem without restricting the parking locations, but forcing the solution to have at least 1 interval with more than 60 hours of driving without a *weekly rest*. As this bound is dependent only on the HOS regulation and the total driving time, it can be calculated off-line to be used for comparison when needed. The solution generated by the *Simplified* model has the minimum cost among the subset of solutions that do not use the rolling time-window constraint. Therefore, if its cost is smaller than the lower bound of the cost of solutions that use the rolling time-window constraint, we can guarantee that this solution is optimal. For the tested scenarios, the results for the *Simplified* model were optimal and significantly lower than the bound. For schedules that exceed this bound we can only show what is the maximum possible improvement to the solution if the *Reset* model were used, and use this information to decide whether to accept the current solution or try to improve it by using the *Reset* model.

Complexity Figure 111 shows that the average solve times for all 3 models are almost the same when the number of locations used is smaller than 61. Due to fixed spacing between locations used in the experiment, at 61 locations the total driving time reaches the weekly driving limit (60h) and the weekly constraints start being needed. The solve times for the *Reset* model rises sharply after that threshold. The solve times for the other models also increase, but at a slower rate. Although the *Reset* and *No Reset* models have a similar number of constraints and variables, the indicator constraints make the *Reset* model's solve time increase significantly faster. Unexpectedly, even though the *Simplified* model has a notably smaller number of variables and constraints, the variation of its solve time was very inconsistent and did not show a significant improvement compared to the *No Reset* model. Nevertheless, its solve time is still shorter than the *Reset* model while finding solutions of same or similar costs, which are significantly better than the *No Reset* solutions. The oscillations seen in the solve time plot of Figure 111 show that the problem is significantly harder to solve in certain scenarios. This is

most likely caused by how the rest areas spacing and total driving time match with the regulations. We believe that the solve time is very sensitive to the configuration of the rest areas, and that those oscillations will change for other configurations and regulations.

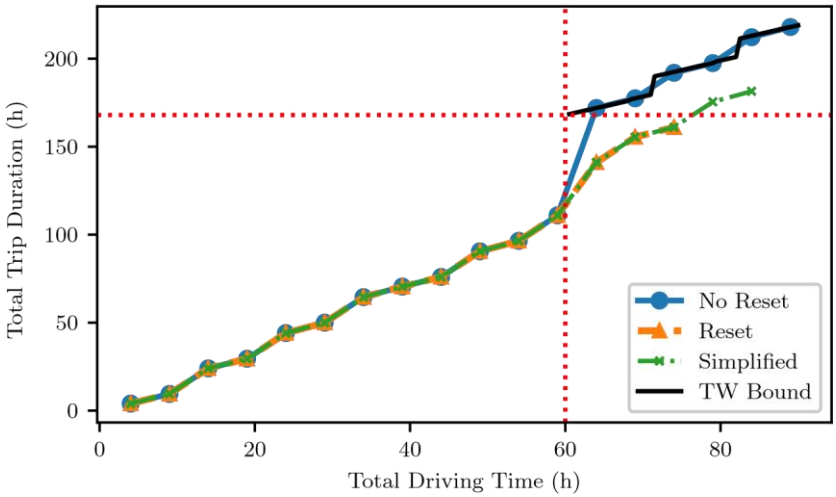


Figure 110. Average trip duration of schedules calculated by the 3 methods with varying total driving time, and the lower bound of solutions that use the rolling time-window. The vertical dotted line marks the on-duty time weekly limit (60h) and the horizontal dotted line represents a trip duration of 1 week (168h).

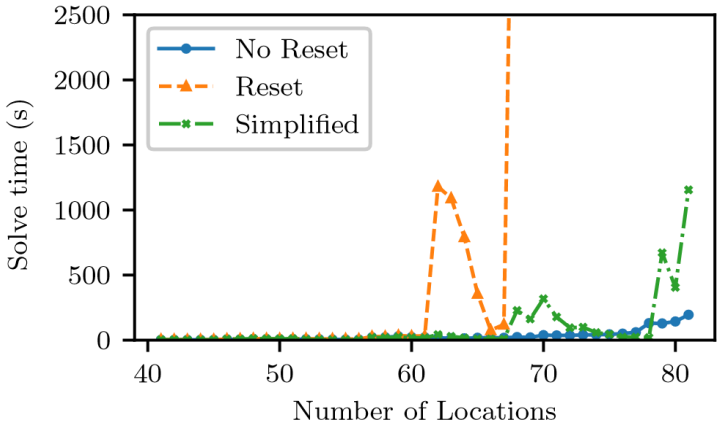


Figure 111. Solve time of the 3 presented methods, with varying number of locations and total travel distance.

Conclusion

In this study, a MIP model for the TDSP with parking availability was presented and extended to include the weekly constraints of the USA regulation. Moreover, the effects of the inclusion of parking availability and weekly working hours constraints to the performance and complexity of the model were studied.

The experiments showed that by including parking availability constraints, our model guarantees the practical feasibility of the schedules with minimal impact on the cost for the driver/company, whereas the practical feasibility of other models' solutions is compromised by the risk of not finding parking. However, using only the simplified weekly constraint was found to be more advantageous compared to using the complete constraint due to scalability issues, except when the driver is expected to consistently drive less than the allowed daily limit.

To the extent of our knowledge, this was the first work to consider the time-dependent parkingavailability of the rest areas and the rolling time window nature of the USA HOS regulation for long trips in the TDSP. As long as data is available, this model could be used by drivers to better plan their schedules, both off-line and during the trip whenever new data came in. In future work, it would be interesting to consider the impact of each truck's schedule on the parking availability and optimize schedules for large fleets. Another topic which was not treated and can also be pursued as future work is accounting for the uncertainty in parking availability.

Shortest Path and Truck Driver Scheduling Problem

Section based on the publication:

F. Vital and P. Ioannou, "Scheduling and Shortest Path for Trucks with Working Hours and Parking Availability Constraints," Transportation Research Part B: Methodological 148 (2021): 1-37, doi:10.1016/j.trb.2021.04.002

Introduction

In 2015, a survey by the Federal Highway Administration identified truck parking shortages in 36 US states, with more pronounced issues along major trade corridors (U.S. Department of Transportation 2015). The lack of truck parking can have significant impact on road safety, industry costs, and the environment (Rodier et al. 2010; Sochor and Mbiydzenyuy 2013), and ranked among the trucking industry's top concerns in recent surveys by the American Transportation Research Institute (ATRI) (American Transportation Research Institute 2019).

Due to the Hours-of-Service (HOS) regulations, truck drivers are required to take regular breaks. These rules aim to ensure that drivers take adequate rest and avoid fatigue-related accidents. However, when appropriate parking is scarce, drivers may find themselves having to choose between driving beyond the legal limits or parking in unauthorized and often unsafe locations, such as highway shoulders and freeway ramps. In recent surveys, most drivers reported using unauthorized parking locations at least once a week (Boris and Brewster 2018; U.S. Department of Transportation 2015; Rodier et al. 2010; Martin and Shaheen 2013)]. As truck crashes can be very costly (Zaloshnja and Miller 2007; Hagemann et al. 2013), such practices may lead to significant losses for the trucking industry due to potential accidents. According to a study by the Virginia Department of Transportation, 25% of all truck-related crashes along Virginia's major corridors occurred on entrance and exit ramps (Kimley Horn 2015). Although the data is not specific to parked trucks, it shows that parking on ramps poses a significant safety risk. Truck insurance premiums have increased in recent years and represented 5% of the average marginal operational cost of trucking (not including workers compensation costs/insurance, physical damage, jury awards, and out-of-court settlements) in 2019 (Murray and Glidewell 2019). This is in part due to recent increases in workers compensation claims, settlements, and jury awards, which at times can surpass $10 million (Lysiak 2019). These growing financial risks push trucking companies to reevaluate safety and how much risk they are willing to take.

Truck parking shortage costs go beyond safety-related ones (Sochor and Mbiydzenyuy 2013; Boris and Brewster 2018). Surveys report that drivers often spend more than 30 minutes looking for parking (American Transportation Research Institute 2018; NCDOT 2017). Although a survey by ATRI (Boris and Brewster 2018) reported that most surveyed drivers spent less than 15 minutes looking for parking, the drivers had, on average, parked one hour earlier than required by the regulation, which also contributes to decreasing their daily revenue-earning miles. Stopping for rest early or spending a long time looking for parking is inefficient use of driver's time. As most truck drivers are not paid by the hour, this can have a significant impact on their compensation (Boris and Brewster 2018). Spending long times looking for parking also means higher fuel consumption, increasing both costs and emissions. Besides, as drivers often

idle the truck's engine to power their appliances, when trucks end up parking near residential areas, the emissions generated can significantly affect the region's air quality (Palaniappan, Wu, and Kohleriter 2003).

An important aspect of truck parking availability is its time dependence. Although drivers often report difficulties finding parking and truck stops report operating overcapacity, they usually refer to the period between 7 PM and 5 AM when drivers are looking for overnight parking (U.S. Department of Transportation 2015; NCDOT 2017; Boris and Brewster 2018; Martin and Shaheen 2013). This suggests that it may be possible to mitigate the truck parking shortage by encouraging drivers to plan their stops on off-peak periods, thus balancing the demand. Balancing the demand with respect to space may also be an option. Several factors influence drivers' choice of parking location; however, lack of information can lead drivers to park illegally even when there are facilities with available parking nearby (Rodier et al. 2010). In (U.S. Department of Transportation 2015), less than 50% of truck stops reported working overcapacity. Many of the facilities operating under capacity may be in regions with low parking demand, yet, some may be in high-demand areas and are underutilized because drivers do not know they are viable options. With this in mind, it is worth considering to include parking availability earlier in the planning process. Instead of only providing drivers information on facilities along their path to allow for on-trip decisions, if parking information is considered when choosing the path itself, the existing parking capacity could be better utilized. Depending on the situation, it may be advantageous for the driver to take a longer route if that can guarantee available parking. This suggestion of addressing parking issues at the supply chain level has been brought up before in (Boris and Brewster 2018; Phelan et al. 2016; de Almeida Araujo Vital, Ioannou, and Gupta 2020); however, literature on the topic is still scarce.

Although planning methods that account for HOS rules, the truck driver scheduling problem (TDSP), have been extensively studied in the truck scheduling literature (Archetti and Savelsbergh 2009; A. Goel 2010; Drexl and Prescott-Gagnon 2010; Asvin Goel and Kok 2012; Asvin Goel 2012; Koç et al. 2016; Vital and Ioannou 2019), the issue of truck parking availability being time-dependent was only considered by Vital and Ioannou in (Vital and Ioannou 2019). The usual assumptions are that any valid parking location is available 24/7, and that drivers can arrive at client locations as early as needed and wait until their delivery time-window. Given the truck parking shortage in many States, we believe that such assumptions can lead drivers to locations that are unable to accommodate them at their arrival time. Therefore, as in (Vital and Ioannou 2019), we consider that parking is restricted in both time (only within certain time-windows) and space (only at rest areas), and that early arrivals at client locations are not allowed. However, (Vital and Ioannou 2019) studied a scheduling problem where the path to be taken is given, whereas, in this project, we address a more general version of the problem where the path taken must also be optimized. While HOS rules have been studied in the context of shortest path problems in (Asvin Goel and Irnich 2017; Drexl and Prescott-Gagnon 2010; Mayerle et al. 2020), these works do not account for parking availability information. The methods presented by Goel and Irnich (Asvin Goel and Irnich 2017), and by Drexl and Prescott-Gagnon (Drexl and Prescott-Gagnon 2010), consider that drivers may rest anywhere along the

path, whereas a recent study by Mayerle et al. (Mayerle et al. 2020) restricts parking to appropriate facilities but does not account for the parking shortage.

HOS-compliant schedules, efficient path choices and safe parking conditions (time and location) are all important aspects of planning for long-haul trucking. Although these topics have been studied separately, work combining two of them is very limited (i.e., HOS rules and parking availability in (Vital and Ioannou 2019), and HOS rules and path planning in (Mayerle et al. 2020), and the intersection of all three topics has yet to be investigated. Our purpose is to study the interaction between HOS-compliant scheduling, path planning and time-dependent parking availability, and the importance of considering all three when facing truck parking shortages.

Scientific contributions and structure

In this study, we focus on the issue of integrating time-dependent parking availability information into long-haul truck planning (path and schedule). The main contributions of the study are the following: First, we introduce the shortest path and truck driver scheduling problem with parking availability constraints (SPTDSP-PA). The SPTDSP-PA extends the Truck Driver Scheduling Problem (TDSP) by adding a parking availability and a path optimization component. Each parking facility in the road network has a set of parking availability time-windows that restrict when drivers can park. Multiple paths may exist between two consecutive clients, each one with its own set of parking facilities. Second, we propose a resource-constrained shortest path formulation for the SPTDSP-PA along with a tailored label-correcting algorithm used to find an optimal solution. Third, we analyze the impact that parking constraints have on trip duration and compare it to estimated potential costs of disregarding parking during planning.

The truck parking shortage has been recognized as a safety concern by the USA (U.S. Department of Transportation 2015) and the European Union (Weenen et al. 2019). In addition, estimates point to a substantial impact in the economy (Weenen et al. 2019; Sochor and Mbiydzenyuy 2013; Hernández and Anderson 2017). Nevertheless, research on freight planning accounting for parking availability remains very limited. The same is true for ways to estimate how parking availability relates to trip duration, costs, and illegal parking. Our study provides a tool for policymakers to estimate how the current truck parking infrastructure affects truck drivers' ability to safely comply with the HOS regulations. The ability to estimate the average cost of following a safe schedule at different parking availability levels can be used to support decisions on parking infrastructure spending. From the industry's standpoint, our model allows drivers and companies to improve their operations' safety standards and estimate their costs more accurately.

This study is organized as follows: Section Literature Review reviews the relevant literature. Section USA's Hours of Service Regulations describes the HOS regulations. Section Problem Description presents the problem addressed. Section Model describes the proposed formulation. Section Label-Correcting Method describes the label-correcting algorithm used. Section Case Study presents a case-study used to evaluate the impact of the proposed method.

Section Randomized Networks Experiments presents the experiments used to measure the algorithms performance. Section Conclusion presents the conclusion.

Literature Review

The inclusion of HOS rules in scheduling algorithms, the truck driver scheduling problem (TDSP), was approached in many studies in recent years (Archetti and Savelsbergh 2009; A. Goel 2010; Asvin Goel and Kok 2012; Asvin Goel 2012; Koç et al. 2016; Vital and Ioannou 2019). Multiple regulations have been considered, including ones from the United States (Asvin Goel and Kok 2012), Europe (A. Goel 2010) and Canada (Asvin Goel and Rousseau 2011). Furthermore, it is often studied as part of a vehicle routing and truck driver scheduling problem (VRTDSP) (Kok, Hans, et al. 2010; Kok, Hans, and Schutten 2011; Rancourt, Cordeau, and Laporte 2013; Asvin Goel and Vidal 2014; Gaddy, Hernandez, and Nurre 2018; Koç, Jabali, and Laporte 2018), which is a variant of the vehicle routing problem (VRP) that accounts for HOS rules, and, less commonly, it is studied in the context of shortest path problems (SPP) (Drexl and Prescott-Gagnon 2010; Asvin Goel and Irnich 2017; Mayerle et al. 2020). Besides the particular methods used, the differences between problems treated in the literature usually relate to the following aspects: regulation considered, optimality of the solutions, parking restrictions, cost function, and main problem (TDSP, VRP or SPP). We are most interested in how they approached parking restrictions and path planning.

Parking restrictions

Although truck parking is currently a critical issue, it is often overlooked in the literature, with many methods not even restricting parking to appropriate facilities. In (Archetti and Savelsbergh 2009), Archetti et al. considered the problem of determining whether a sequence of n full truckload transportation requests is feasible given a set of HOS regulations and pick-up time-windows. The proposed method allows drivers to park anywhere and finds a feasible schedule in $O(n^3)$ time. In (A. Goel 2010), Goel considered a similar problem using the European regulations, and in (Asvin Goel and Kok 2012), presented an algorithm to find feasible schedules to visit n locations using the US regulations in $O(n^2)$ time. However, these methods assumed that drivers could park anywhere, which is not valid in practice. This assumption is also present in (Kok, Meyer, et al. 2010; Drexl and Prescott-Gagnon 2010; Rancourt, Cordeau, and Laporte 2013; Asvin Goel and Vidal 2014; Asvin Goel and Irnich 2017). In (Asvin Goel 2012), Goel presented a mixed integer programming (MIP) formulation and a dynamic programming algorithm for the TDSP that restricts parking to client locations and calculates a schedule with minimum trip duration. Rest areas were modeled as clients with zero service time and unbounded time-windows. Similar MIP models were used in (Kok, Hans, and Schutten 2011; Koç et al. 2016; Vital and Ioannou 2019), focusing on different aspects of the problem but keeping parking restricted to appropriate facilities. In (Kok, Hans, and Schutten 2011), Kok et al. addressed the issue of traffic congestion by considering time-dependent travel times and proposed a heuristic approach to integrate the TDSP model into a VRP method. In (Koç et al. 2016), Koç et al. approached the environmental impact caused by truck idling and how it is affected by the truck's equipment and rest areas' infrastructure. The drivers can only park at rest areas, which have different types of infrastructure available. Early arrival is allowed at

client locations, but it does not count as off-duty time. The cost function accounts for the type of idling used in each stop given the equipment installed on the truck and the infrastructure available at each rest area. This method was later used as a base for a VRTDSP algorithm with the same focus (Koç, Jabali, and Laporte 2018). In (Vital and Ioannou 2019), Vital and Ioannou approached the issue of truck parking availability and US HOS rules for long trips. Their model considered a single client trip, which hinders drivers' ability to plan consecutive trips. Parking was restricted to rest areas, and parking availability was modeled as time-window constraints for each rest area. Each rest area's availability time-windows take effect only if a stop is scheduled for that particular location. Due to the focus on parking availability issues, the model assumed that parking is unavailable outside of the delivery time-window and did not allow early arrival at the client or rest areas. As short-term staging due to warehouse or terminal hours is a source of truck parking demand (U.S. Department of Transportation 2015; Cambridge Systematics 2019), we see the restriction on early arrivals (also included in our model) as an important distinction when considering truck parking shortages. This study is the only one that considered time-dependent parking availability in the TDSP. Nevertheless, as (Vital and Ioannou 2019) addresses only the scheduling problem, it does not account for alternative paths or parking locations that require a detour to be reached. This limitation motivates the other aspect of our work: path planning.

Path planning

The inclusion of parking constraints and HOS regulations when determining the shortest path between locations is relevant not only to individual drivers that need to plan their itineraries, but also to carriers and other stakeholders that need to estimate operational costs and allocate resources. Hence why we are interested in the shortest path problem with resource constraints (SPPRC) that lies between the TDSP and the VRTDSP. VRTDSP methods assume that the shortest path between any two clients is known (and independent of the current status of the HOS constraints), and use TSDP algorithms to calculate the cost of each route generated. The rest areas considered in these problems are located along these known shortest paths. If the driver is allowed to rest anywhere or only at client locations, this assumption does not affect the route cost. However, when parking is restricted and rest areas are considered, the minimum cost path between two clients will depend on the location of every reachable rest area and the HOS constraints' status at the departure time from the client. The inclusion of parking availability constraints makes it even more important to consider alternative paths and rest areas. When parking is scarce at the usual routes, it may be cost-effective to take a slightly longer path if it has better parking conditions. Failing to consider how parking availability and HOS constraints affect the shortest path between clients may cause planners to underestimate the trip's duration and cost. This inaccuracy can upset operations planning as well as fair driver remuneration (depending on how wages are determined). The issue is aggravated when drivers lack the flexibility to adjust their route, as some of the drivers surveyed in (Sun et al. 2013). In this case, the driver is limited to taking a sub-optimal route, further increasing the difference between estimated and actual trip cost and duration.

The shortest path problem with resource constraints (SPPRC) often appears in column generation solutions to the VRP (Costa, Contardo, and Desaulniers 2019) and several

approaches have been proposed for its variants (Asvin Goel and Irnich 2017; Pugliese and Guerriero 2013; Lozano, Duque, and Medaglia 2016; Horváth and Kis 2016; Irnich and Desaulniers 2005). The SPPRC is often solved through dynamic programming-based labeling algorithms, applying tailored dominance rules and bound estimates to identify and discard inferior paths. SPPRC formulations and algorithms are tailored to their own problem variants and may not be directly applicable to other problems. Hence the need to develop tailored methods for the SPPRC in the context of HOS regulations and parking availability constraints. However, the number of studies using SPPRC formulations in the context of HOS-compliant planning is very limited. In (Drexl and Prescott-Gagnon 2010), Drexl and Prescott-Gagnon present a SPPRC formulation to the problem of finding HOS-compliant routes and schedules, and propose exact and heuristic labeling algorithms. In (Asvin Goel and Irnich 2017), Goel and Irnich propose an exact method for the VRTDSP using a branch and price algorithm where a SPPRC is used to generate HOS-compliant routes and their costs. An auxiliary network is used to model drivers' possible activities, but parking locations are not considered. Even though they consider HOS regulations, both (Asvin Goel and Irnich 2017) and (Drexl and Prescott-Gagnon 2010) assume that drivers may stop and rest anywhere on a route. This limitation is partially addressed in (Mayerle et al. 2020), where Mayerle et al. study the impact of Brazilian regulations in the planning of long-haul full truckload shipments. Differently from (Drexl and Prescott-Gagnon 2010; Asvin Goel and Irnich 2017), this study is not aimed at deciding which clients to visit and in what order for a VRTDSP, but at how changes to HOS rules affect the best path to reach a client. They use a labeling algorithm and pruning heuristics to optimize the path a truck takes to reach a single client, while scheduling stops at allowed locations to satisfy regulations. Their model includes some time-restrictions to all rest stops by restricting departure times at the beginning of each work day, as well as the start time of lunch breaks. However, they also overlook the question of whether those parking locations will be available at the desired times. In addition, it shares the same single client limitation as (Vital and Ioannou 2019).

USA's Hours of Service Regulations

The current USA HOS regulation (Code of Federal Regulations, n.d.) differentiates between driving time, on-duty time and off-duty time. In summary, driving time is the time spent operating the truck, on-duty time is the time from when the driver is required to be ready for work until he/she is relieved from work, and off-duty is the time when the driver is not on-duty. The time restrictions set by the regulation can be reset by off-duty periods with minimum durations specified in the regulation. We refer to off-duty periods lasting at least 0.5, 10, and 34 consecutive hours, as *breaks*, *daily rests* and *weekly rests*, respectively. Note that the longer off-duty periods can be used to reset the restrictions related to the shorter ones. The USA HOS regulation can be summarized as follows:

- Daily Driving Time Limit: A driver may drive at most 11 hours between 2 consecutive *daily rests*.
- 14-Hour Elapsed Time Limit: A driver cannot drive after 14 hours have elapsed since the last *daily rest* ended.
- Rest Breaks: A driver cannot drive after 8 hours have elapsed since the last *break* ended.

- 60-Hour Limit: A driver cannot drive after having been on duty for 60 hours in any period of 7 consecutive days. The 7 days period can be reset by taking a *weekly rest*.

Unless the driver expects to be on-duty for less than 8.6 hours per day (60/7) on average, taking *weekly rests* is more efficient than not. Therefore, a simplified version of the 60-Hour Limit was considered in this paper. Instead of restricting the on-duty time over any period of 7 consecutive days, the on-duty time between two consecutive *weekly rests* was restricted to 60 hours. This greatly simplifies the implementation of the 60-Hour limit, while still guaranteeing regulation compliance.

Problem Description

The problem consists of planning a single truck's minimum cost path and schedule from the origin to an ordered set of client locations, while complying with the USA HOS regulations, client service time-windows and only scheduling off-duty time at truck parking locations (TPL) which are expected to have available parking at the time of arrival. The road network's travel times are fixed and known. Parking availability is modeled as time-windows for each TPL, within which the driver is guaranteed to find parking. The parking availability time-windows are assumed known. The collection of data and statistical analysis required to generate reasonable time-windows is beyond the scope of this study and is not addressed. Some recent work on parking prediction can be seen in (Rajabioun and Ioannou 2015; Bayraktar et al. 2015; Sadek, Martin, and Shaheen 2018; Monteiro and Ioannou 2018; Morris et al. 2018; Tavafoghi, Poolla, and Varaiya 2019).

An important difference between this study and some of the scheduling papers mentioned earlier, such as (Asvin Goel 2012; Koç et al. 2016), is that we do not allow early arrival at nodes with time-window constraints. In (Asvin Goel 2012), the driver was allowed to arrive early at a client location and wait until the start of the service time-window. This waiting time was treated as an off-duty period and, when long enough, could be used to satisfy HOS rest requirements. In (Koç et al. 2016), the driver was also allowed to arrive early at a client, however, the waiting time is treated as on-duty time and could not be used to satisfy the HOS rest requirements. Both studies set unbounded time-windows for TPLs, so early arrivals at these facilities were not a concern. In this study, we assume that TPLs and clients can only accommodate new vehicle arrivals within the time-windows. Therefore, allowing early arrival would be equivalent to telling the driver to stop outside the parking or customer facility and wait until it becomes available, which goes against the objective of avoiding stops at inappropriate locations. According to (U.S. Department of Transportation 2015; Cambridge Systematics 2019), short-term staging due to warehouse or terminal hours is a source of truck parking demand. Therefore, we believe that the ability to accommodate drivers whenever they arrive should not be taken for granted. If a client allows drivers to arrive early, this should be explicitly modeled as a TPL right before the customer location. This TPL should have its own time-windows to limit how early the driver may arrive, and possibly include restrictions on the duration of stay.

The problem is solved over a road network that includes only the main routes the truck can take for that specific trip, and the TPLs around them. Trucks have road restrictions due to size and weight, so using the full road network would include a lot of superfluous information. In addition, the order in which clients must be visited can also be used to narrow down the set of relevant paths. The truck route network is generated from the full network by calculating a set of paths between every two consecutive clients, then connecting each path to the set of nearby TPLs to be considered. The number of paths considered, as well as the relevant TPLs for each path, is at the user's discretion. Appendix M presents a simple heuristic that could be used to choose the number of paths and its impact on solution quality and running time. The network is defined as an acyclic directed graph $G = (V, A)$, where V is the set of nodes of the graph and A is the set of arcs. The nodes represent locations of interest in the road network, as TPLs, client locations, intersections, and road branching spots. The arcs represent road sections, and each $(i, j) \in A$ is assigned a fixed travel time d_{ij}, and length l_{ij}. Figure 112 shows an example network, the nodes with a number index are road nodes and the ones with a letter index are TPLs. The edges connected to TPLs are represented as dashed arrows and the main paths as continuous arrows.

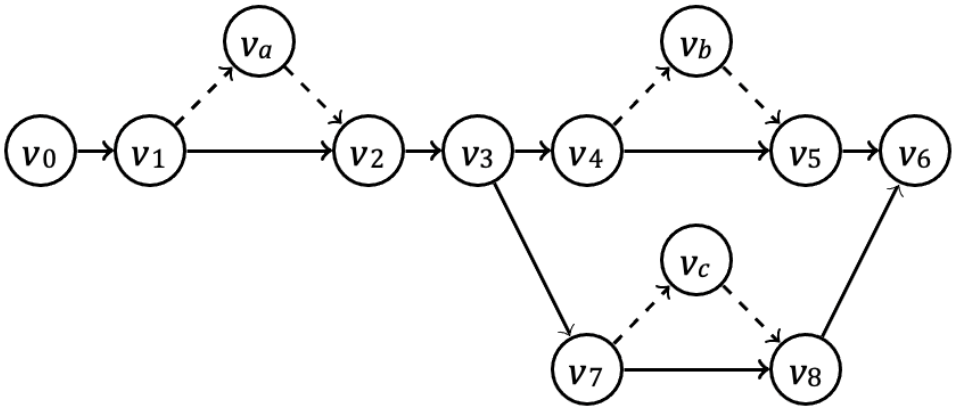


Figure 112. Example of simplified road network. The nodes with a number index are road nodes (intersections, branching or merging spots) and the ones with a letter index are TPLs.

Model

The SPTDSP-PA is modeled as a resource-constrained shortest path problem (Irnich and Desaulniers 2005), where the time, cost, and counters for the different HOS regulations are treated as resources. This section presents the extended network and system dynamics used to model the problem presented in the section Problem Description.

Extended Network

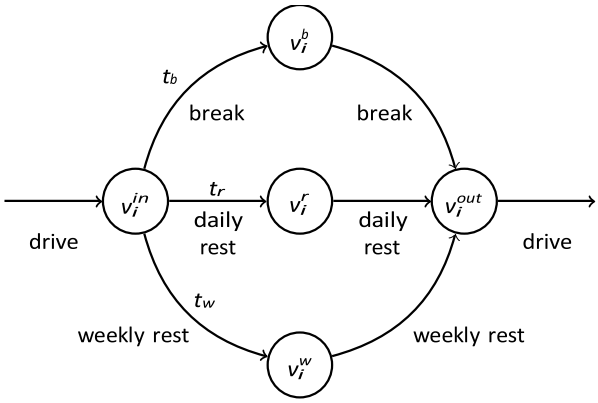
The network introduced in the section Problem Description describes the road sections and locations considered in the problem, but it does not portray the activities that take place at each location. In order to represent the different activities and decisions involved in the problem, we define an extended network $G' = (V', A')$. Nodes representing locations where non-driving activities take place are expanded to explicitly include these activities as edges in

the graph. Figure 113 shows the subnetworks that will replace the expanded nodes. Each edge has its activity indicated below the arrow. Non-drive edges that have a fixed duration have their duration indicated above the arrow. The incoming/outgoing edges of the subnetwork are the incoming/outgoing edges of the expanded node.

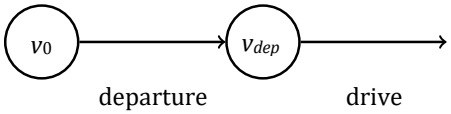
Rest Areas: Off-duty periods of different durations satisfy different HOS constraints, so rest area nodes are expanded to include a different path for each type of off-duty period. In this case, only 3 types are considered. They are *break*, *daily rest* and *weekly rest*, which have minimum durations of t_b, t_r and t_w, respectively. Figure 113a shows the subnetwork representing the node after expansion. The duration of the second half of each path is a non-negative decision variable used to model rest time beyond the minimum required. We only restrict arrival time, so time-windows that were assigned to the original rest area node are assigned only to the entrance node v_i^{in}. However, if a parking facility closes at certain times and requires vehicles to leave before then, this restriction can be modeled by a time-window at the exit node v_i^{out}. Rest areas with restrictions on the duration of stay can be modeled by restricting the values that the second half of each path can take and by removing paths with minimum duration exceeding the limit (e.g., if a rest area does not allow stops longer than 4 hours, the paths for *daily* and *weekly rests* can be removed, and the second half of the *break* path restricted to at most 3.5 hours). In the remainder of the chapter, the central nodes of each path, v_i^b, v_i^r and v_i^w will be referred to as *break*, *daily rest* and *weekly rest* nodes, respectively. Let V_b, V_r and V_w represent the set of all *break*, *daily rest* and *weekly rest* nodes, respectively.

Origin: The choice of departure time from the Origin node does not have the same effect as the choice of departure time from TPLs. We assume it does not affect HOS resources, but that it might affect cost. Therefore, it must be treated differently. The Origin node is replaced by the subnetwork in Figure 113b to model the vehicle's departure time. Departure time constraints can be modeled either by the set of allowed values for the duration of that edge, or as a time-window constraint on the departure node v_{dep}.

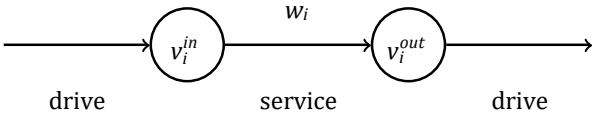
Clients: Client nodes are expanded as shown in Figure 113c. The service edge in the subnetwork is used to model the service time at each client. As the regulation differentiates between driving and on-duty time, service time must be treated differently from the driving time. The time-windows that were assigned to the original client node will be assigned only to the entrance node v_i^{in}.



(a) Sub-network used to expand rest area nodes.



(b) Sub-network used to expand the origin node.



(c) Sub-network used to expand client nodes. w_i is the service time.

Figure 113. Sub-networks used to model non-driving activities.

System Equations

As the driver's schedule must comply with the HOS regulations, each regulation constraint is modeled by a different resource that must be kept below the limits described in the regulation throughout the whole path. The resources considered are:

- current time (η^0)
- cost (c)
- elapsed time since last *break* (η^b)
- elapsed time since last *daily rest* (η^r)
- accumulated driving time since last *daily rest* (ψ^r)
- accumulated on-duty time since last *weekly rest* (ψ^w)

Each state of the system can be described by the tuple $x = (i, \theta)$, where $i \in V'$ is the current node and $\theta = (\eta^0, c, \eta^b, \eta^r, \psi^r, \psi^w)$ lists the current resource values. Each arc $(i, j) \in A'$ is assigned a set of allowed durations Δ_{ij} and a length l_{ij}. When (i, j) represents a road section, then $\Delta_{ij} = \{d_{ij}\}$ and l_{ij} is the length of the associated arc from A. For the other types of arcs (*break*, *daily rest*, *weekly rest*, *departure*, *service*), present only in A', l_{ij} is zero. If the arc has a fixed duration, Δ_{ij} is a single element set containing that duration. If the arc has a variable duration, like the departure arc and the arcs used to define off-duty period extensions, $\Delta_{ij} = [0, \infty)$. This interval can be further restricted to avoid unwanted decisions being checked, e.g.,

breaks should not be extended to the point where they are longer than a *daily rest*'s minimum duration.

Let $x_k = (v_k, \theta_k)$, where $\theta_k = \left(\eta_k^0, c_k, \eta_k^b, \eta_k^r, \psi_k^r, \psi_k^w\right)$, represent the system's state after k decisions. Let $U(x_k)$ represent the set of feasible decisions at state x_k. Each decision is described by a tuple $u_k = (v_{k+1}, \delta_k)$, where v_{k+1} is the next node to be visited, and δ_k is the time required to reach it. The evolution of the system is described by:

$$x_{k+1} = f(x_k, u_k), \quad \forall x_k \in X, u_k \in U(x_k), \tag{94}$$

$$\begin{aligned} U(x_k) \quad &= \{u_k = (v_{k+1}, \delta_k) | f(x_k, u_k) \in X, \\ &\quad x_k = (v_k, \theta_k), (v_k, v_{k+1}) \in A', \delta_k \in \Delta_{v_k v_{k+1}}\}, \end{aligned} \tag{95}$$

where X is the set of feasible states. As the resources are affected differently by the activities involved, the function $f(\cdot)$ is defined separately for each activity. Furthermore, as the definition of $f(\cdot)$ is trivial for the update of the next node, i.e., it is always equal to the first element of the decision tuple, this part will be omitted. The resource update rules are modeled by a set (one for each activity) of resource extension functions (REF), which are described in Table 22. These functions take the resource vector θ_k, the duration δ_k and length $\mu_k = l_{v_k v_{k+1}}$ of the chosen arc as arguments, and return the new resource vector θ_{k+1}. The functions f^d, f^s, f^b, f^r, f^w and f^0 are used for the activities *drive*, *service*, *break*, *daily rest*, *weekly rest* and *departure*, respectively. For example, if edge (v_k, v_{k+1})'s activity is *drive*, then $\eta_{k+1}^b = \eta_k^b + \delta_k$, but if (v_k, v_{k+1})'s activity is *break*, then $\eta_{k+1}^b = 0$. The assignment of activities to the arcs of the extended network is described in the section Extended Network. The cost is modeled as a linear combination of the distance traveled and the time spent in each activity throughout the trip. α^d, α^s, α^b, α^r, α^w and α^0 are the hourly costs applied to activities *drive*, *service*, *break*, *daily rest*, *weekly rest* and *departure*, respectively. We assume that $\alpha^0 < \alpha^b = \alpha^r = \alpha^w$. β_d is the cost per kilometer traveled, it is applicable only to the activity *drive*. Related works usually use total trip duration, on-duty time, travel distance or a combination of those factors as the cost, so our formulation is flexible enough to model most cost functions found in the literature.

Table 22. Resource Extension Functions

	f^d	f^s	f^b	f^r	f^w	f^0
$\eta^0_{k+1} =$	$\eta^0_k + \delta_k$					
$c_{k+1} - c_k =$	$\alpha_d\delta_k + \beta_d\mu_k$	$\alpha_s\delta_k$	$\alpha_b\delta_k$	$\alpha_r\delta_k$	$\alpha_w\delta_k$	$\alpha_0\delta_k$
$\eta^b_{k+1} =$	$\eta^b_k + \delta_k$		0			η^b_k
$\eta^r_{k+1} =$	$\eta^r_k + \delta_k$			0		η^r_k
$\psi^r_{k+1} =$	$\psi^r_k + \delta_k$	ψ^r_k		0		ψ^r_k
$\psi^w_{k+1} =$	$\psi^w_k + \delta_k$			ψ^w_k	0	ψ^w_k

Constraints

There are two types of constraints, time-window constraints, used to model arrival time restrictions at client locations and parking availability at TPLs, and HOS constraints to model the compliance to HOS regulations. A state $x_k = (v_k, \theta_k)$ is only feasible if it satisfies all constraints. Below we describe each constraint:

Time-Window Constraints: Each node i representing a client location or TPL has a set of T_i disjoint time-windows. Each time-window is defined by a tuple $\left(t^{min}_{i,\tau}, t^{max}_{i,\tau}\right)$ representing the minimum and maximum arrival times allowed by that time-window, where $\tau \in \{1, \cdots, T_i\}$ is the index of the window. A state x_k with $v_k = i$ satisfies the time-windows constraints if and only if $\eta^0_k \in \cup_{\tau=1}^{T_i}\left[t^{min}_{i\tau}, t^{max}_{i\tau}\right]$.

HOS Constraints: Each HOS related resource has a maximum allowed value defined by the regulation introduced in the section USA's Hours of Service Regulations. Let t_{eb} be the limit for elapsed time between *breaks*, t_{er} the limit for elapsed time between *daily rests*, t_{ar} the limit for accumulated driving time between *daily rests*, and t_{aw} the limit for accumulated on-duty time between *weekly rests*. Then a feasible state x_k must satisfy:

$$\eta^b_k \le t^{eb}, \quad \eta^r_k \le t^{er}, \quad \psi^r_k \le t^{ar}, \quad \psi^w_k \le t^{aw} \tag{96}$$

Note that the regulation restricts driving, but not other working activities. So, if a client has a parking facility that does not require driving to be reached, that parking facility can be modeled as a TPL right after the client exit node and the HOS restrictions could be relaxed for those particular client exit and TPL entrance nodes. The time limits may be exceeded during service, but the driver would be able to rest before driving, so the schedule would still satisfy the HOS regulations.

SPTDSP-PA Formulation

The objective is to find a minimum cost route and schedule that satisfies the above constraints. The cost is modeled by the resource c. Let X_d represent the set of terminal states, i.e., the feasible states at the destination node. We refer to the node with no outgoing edges in G' as the destination node. It is assumed that all terminal states are absorbing, i.e., $f(x,u) = x, \forall x \in X_d$. Let $\overline{n}$ be the maximum number of decisions required to reach the destination from the Origin node v_0. As G' is an acyclic directed graph, $\overline{n}$ is finite and bounded by the number of nodes $|V'|$. Given an initial state x_0, the SPTDSP-PA is formulated as:

$$\min_{u_0,\cdots,u_{\overline{n}-1}} c_{\overline{n}} \tag{97}$$

$$\text{s.t.} \quad x_{k+1} = f(x_k, u_k), \quad k = 0,1,\cdots,\overline{n}-1 \tag{98}$$

$$x_k \in X, \quad u_k = (v_k, \delta_k) \in U(x_k) \tag{99}$$

where X is the set of feasible states. The objective function (97 minimizes the total trip cost $c_{\overline{n}}$, defined as the cost resource at the last state, $x_{\overline{n}}$. Constraint (98 controls the evolution of the system. Constraint (99) defines the domains of the variables used. $U(x_k)$ is defined in (95), and feasibility is defined in the section Constraints. In the following section we present a method for solving the SPTDSP-PA problem (97)-(99).

Label-Correcting Method

This section describes the dynamic programming-based label-correcting method used to solve the SPTDSP-PA. An overview of resource constrained shortest path problems, including labeling algorithm approaches, can be found in (Irnich and Desaulniers 2005; Pugliese and Guerriero 2013). The general idea behind label-correcting methods is to progressively calculate the shortest path from the origin node to all other nodes until the shortest path to the destination is found. Due to the large number of paths generated, it is necessary to identify and discard inefficient paths as soon as possible.

Overview

This subsection explains the general working of the algorithm. Figure 114 shows the work-flow of the algorithm used. To every partial solution $(u_0, u_1, \cdots, u_{k-1})$ going from the origin node v_0 to a node $v_k \in V'$ we assign a label consisting of the resource vector θ_k at node v_k, the slacks that will be described in the section Label Improvement, and information necessary to reconstruct the partial solution. For simplicity, we refer to a label by its associated resource vector θ_k.

Label Choice: Let *OPEN* be the list of untreated labels. *OPEN* is initialized with the initial state's label and we use the small labels first method (D. P. Bertsekas 1993) to choose which label to treat first. New labels are inserted at *OPEN*'s start if smaller than the first label, and at the end otherwise. At each iteration *OPEN*'s first label is picked and expanded to generate new labels. Labels are compared lexicographically by their resources, following the order $(\eta^0, c, \psi^w, \eta^r, \psi^r, \eta^b)$. We give priority to resources that are reset less often. Giving priority to

the arrival time (η^0) also helps with the dominance check as a label can only be dominated by labels with lesser or equal arrival time. The method used to manage *OPEN* is represented in Figure 114 by the 'Search Method' block.

Expansion: Each label describes a certain state of the system (and also contains information on the partial solution that generated it). The 'expansion' step consists of heuristically choosing a set of feasible decisions and applying them to that label, thus generating new labels/partial solutions. This step is represented in Figure 114 by the 'Expand Partial Solution' block, and, as shown in the diagram, depends on several factors. The set of feasible decisions at any given state depends on the network topology and on the regulations, which include both the HOS and time-window/parking constraints, being considered. Ideally, all feasible decisions should be tested during expansion. However, our model also has a number of continuous decision variables, making it impossible to test all of them. The 'Expansion Criteria' block refers to the heuristics defining how the set of feasible decisions is sampled. In addition, the A^* algorithm and dominance rules are also used to speed up computation by discarding inefficient partial solutions before their labels are inserted into *OPEN*.

Label Improvement: Due to the usage of heuristics during node expansion, the partial solutions generated may contain inefficiencies. Our algorithm keeps track of these inefficiencies and, at nodes where the driver can choose to extend an off-duty period, it checks whether future labels can be improved by updating upstream decisions. This approach is based on how Goel tracked unnecessary off-duty time to adjust infeasible partial solutions in (Asvin Goel 2012). Under certain conditions, by changing the departure time from the origin or the duration of some upstream rest stops, it is possible to reach a particular node at the same time but spending fewer resources or at a lower cost. This step is represented by the 'Update Upstream Labels' block.

Dominance Check: We say that a label is dominated when it cannot generate solutions that are better than the ones generated by another label. A label can only dominate or be dominated by a label of the same node, so the algorithm also keeps a list of untreated labels separated by node to speed up the dominance check. When new labels for rest nodes are created, the algorithm performs dominance checks between the new label and the existing ones. If the new label is dominated by an existing one, it is not inserted into *OPEN*. Any labels dominated by the new one are discarded. In order to reduce the computation load, dominance is checked only on rest nodes, other nodes' labels are always accepted.

Termination Conditions: The algorithm stops if it finds a solution that is close enough to the estimated lower-bound, or if *OPEN* is empty.

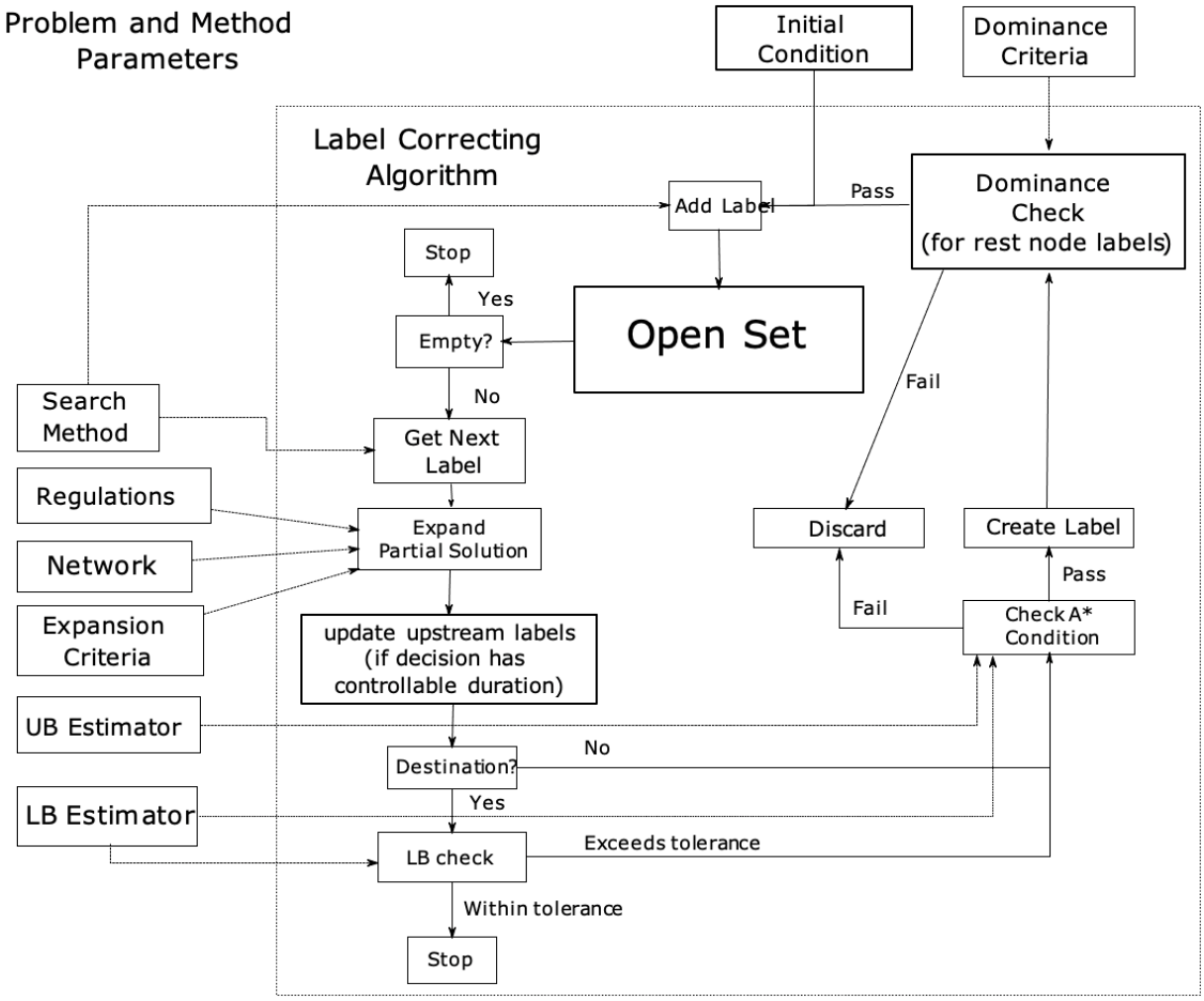


Figure 114. Label-correcting algorithm workflow diagram.

Partial Solution Expansion

Expansion Criteria

If the node being treated has no time duration decision, then there is only a finite number of possible decisions and all feasible decisions should be tested. If a time duration decision is required, i.e., at rest nodes and at the origin, then it is necessary to choose which decisions to test.

Let $A^d \subset A'$ represent the set of all arcs with *driving* as their assigned activity. For every node pair (p, q) such that there is a directed path from p to q, let $\mathcal{D}(p, q)$ and $\mathcal{D}_d(p, q)$ be, respectively, the minimum travel time (including service time) and minimum driving time between nodes p and q with all resource, time-window and HOS constraints relaxed:

$$\mathcal{D}(p,q) = \begin{cases} \min(\Delta_{pq}), & \text{if } (p,q) \in A' \\ \min\limits_{(p,k)\in A'} \left(\min(\Delta_{pk}) + \mathcal{D}(k,q)\right), & \text{o/w} \end{cases} \quad (100)$$

$$\mathcal{D}_d(p,q) = \begin{cases} 0, & \text{if } (p,q) \in A' \setminus A^d \\ \min(\Delta_{pq}), & \text{if } (p,q) \in A^d \\ \min\limits_{(p,k)\in A'} \left(\mathcal{D}_d(p,k) + \mathcal{D}_d(k,q)\right), & \text{o/w} \end{cases} \quad (101)$$

If there is no directed path from p to q, then $\mathcal{D}(p,q) = \mathcal{D}_d(p,q) = \infty$.

Consider the label $\theta_i = \left(\eta_i^0, c_i, \eta_i^b, \eta_i^r, \psi_i^r, \psi_i^w\right)$ relative to a partial solution ending at node v_i. Let $\mathcal{N}(x_i)$ represent the set of nodes that can be reached from state x_i without resting, given by:

$$\mathcal{N}(x_i) = \{v \in V' | \, \mathcal{D}(v_i, v) \leq \rho(\theta_i) \wedge \mathcal{D}_d(v_i, v) \leq t^{ar} - \psi_i^r\} \quad (102)$$

$$\rho(\theta_i) = \min\left(t^{eb} - \eta_i^b, t^{er} - \eta_i^r, t^{aw} - \psi_i^w\right) \quad (103)$$

where $\rho(\theta_i)$ represents how much time is left until one of the HOS constraints related to elapsed time (η_i^b, η_i^r) or on-duty (ψ_i^w) time is broken. As the constraints considered in $\rho(\theta_i)$ are affected by service time, $\rho(\theta_i)$ is compared to $\mathcal{D}(v_i, v)$. The accumulated driving time since last *daily rest* (ψ^r) is not affected by service time, so $t^{ar} - \psi_i^r$ is considered separately and compared to $\mathcal{D}_d(p,q)$. Let $\tilde{x}_i = \left(v_i, \tilde{\theta}_i\right)$ represent a modified x_i where the resources already account for potential label improvements. This is used to avoid ignoring decisions that are only feasible after label improvement. Label improvement is described in the section Label Improvement, and $\tilde{\theta}_i$ is defined in (110). When expanding θ_i, for each time-window of each node in $\mathcal{N}(\tilde{x}_i)$, the shortest decision that can generate a path that will satisfy the time-window is tested. The set $\widehat{U}(x_i)$ of feasible decisions to be tested is described by:

$$\widehat{U}(x_i) = \left\{(v,\delta) \in U(\tilde{x_i}) \middle| \exists l \in \mathcal{N}(\tilde{x_i}), \tau \in \{1, \cdots, T_l\}, \delta = min\left(B(x_i, v, l, \tau)\right)\right\} \quad (104)$$

$$B(x_i, v, l, \tau) = \left\{y \in \Delta_{v_i v} \middle| \eta_i^0 + y + \mathcal{D}(v, l) \in \left[t_{l\tau}^{min}, t_{l\tau}^{max}\right]\right\} \quad (105)$$

where $B(x_i, v, l, \tau)$ represents the duration of decisions passing through node v that, given the current state x_i, can generate paths that reach node l within the time-window $[t_{l\tau}^{min}, t_{l\tau}^{max}]$.

Label Improvement

This step targets specifically nodes that have a time duration decision, i.e., rest nodes and the origin. These nodes have an outgoing arc with controllable length, which is used to define if and how much the driver will wait at the rest stop after completing the minimum required off-duty period, and when to depart from the origin. While drivers may need to extend their rest duration to accommodate time-window constraints, it is important to notice that the waiting times at different nodes are not equivalent. One example is that extending a *break* is more 'expensive' than a *weekly/daily rest*, because *breaks* affect more resources. In terms of wait

time 'cost' we can rank the node types as follows: *origin* < *weekly/daily rests*<*breaks*. So, when a waiting time is needed, it is more efficient to try to extend the waiting time at the last upstream node of a 'cheaper' type. In order to do so we keep track of how much we can extend the wait times of the origin node and of the last upstream *weekly/daily rest* node without affecting the feasibility of solutions generated.

Let N_d denote the set of nodes that have a time duration decision, and N_t the set of nodes that have time-window constraints. Consider a partial solution with path $(v_0, v_1, \cdots, v_j)$, let δ_i be the duration of the decision taken at node v_i, η_i^0 the arrival time at node v_i, and $[a_i, b_i]$ the time-window being used at node v_i, such that $a_i \leq \eta_i^0 \leq b_i$. The sets of slack variables $\{\sigma_{i,j}\}$ and $\{\lambda_{i,j}\}$ are used to track by how much δ_i can be updated. The variable $\sigma_{i,j}$ stores the surplus of off-duty time between nodes v_i and v_j, which can be interpreted as by how much δ_i can be increased without affecting η_j^0. $\lambda_{i,j}$ stores by how much δ_i can be extended without pushing η_k^0 outside of the time-window $[a_k, b_k]$ for all $i < k \leq j$. They are defined as follows:

$$\sigma_{i,j} = \begin{cases} 0, & if\ \ j = i+1 \\ \sigma_{i,j-1} + \delta_{j-1}, & if\ j > i+1\ and\ v_{j-1} \in N_d \\ \sigma_{i,j-1}, & if\ j > i+1\ and\ v_{j-1} \notin\ N_d \end{cases} \tag{106}$$

$$\lambda_{\mathrm{i,j}} = \begin{cases} \infty, & if\ \ \mathrm{j = i+1\ \ and\ v_j} \notin \mathrm{N_t} \\ \lambda_{i,j-1}, & if\ j > i+1\ and\ v_{j-1} \notin N_t \\ b_j - \eta_j^0 + \sigma_{i,j}, & if\ j = i+1\ \ and\ \ v_j \in N_t \\ min(\lambda_{i,j-1}, b_j - \eta_j^0 + \sigma_{i,j}), & o/w \end{cases} \tag{107}$$

where $v_i \in N_d$ and $i < j$. We define $\lambda_{i,i} = \sigma_{i,i} = 0$. For each generated label, the σ's and λ's stored are the ones relative to the origin and the last *weekly* or *daily rest*. As can be seen in the REFs on Table 22, waiting at these types of nodes spends fewer resources than waiting at *break* nodes. In terms of resource expenditure when extending an off-duty period, *weekly* and *daily rests* are equivalent, so only the information regarding the most recent one is stored.

Let $(v_l, H) \in \widehat{U}(x_j)$ be one of the decisions chosen to expand state $x_j = (v_j, \theta_j), v_j \in N_d$, and let $\theta_l = (\eta_l^0, c_l, \eta_l^b, \eta_l^r, \psi_l^r, \psi_l^w)$ be the label generated by this decision. If $\min(\sigma_{0,j}, \lambda_{0,j}) > 0$, then the waiting time δ_0 at node v_0 can be increased to δ_0^* as follows:

$$\delta_0^* = \min(\lambda_{0,j}, \sigma_{0,j} + H) + \delta_0 \tag{108}$$

A new label $\theta_1{}'$ is created for node v_1 using the new decision (v_1, δ_0^*). Instead of putting this label in *OPEN*, it is treated immediately and separate from the others. This label will be expanded passing by the same nodes as the partial solution that is being improved, but taking the shortest decisions that will not decrease the arrival time at any intermediate node. Note that the intermediate labels generated are not stored in *OPEN*, and only one decision is used in their expansion. We want to generate an updated version of the partial solution being improved, not a new tree of partial solutions. This updated path will generate a label $\theta_l{}'$ with

cost resource $c_l' = c_l + (\alpha_0 - \alpha_r)\min(\lambda_{0,j}, \sigma_{0,j} + H)$ that dominates θ_l. Before storing the label in *OPEN*, we check if other slacks can be removed.

Let v_r be the last *weekly* or *daily rest* node visited, with $v_r = v_0$ if none was visited. If $\min(\sigma_{0,j}, \lambda_{0,j}) = 0$, but $r \neq j$ and $\min(\sigma_{r,j}, \lambda_{r,j}) > 0$, then the same procedure can be used to update δ_r using $\sigma_{r,j}$ and $\lambda_{r,j}$. Updating δ_r reduces η_j^r (elapsed time since last *daily rest*) by $\min(\lambda_{r,j}, \sigma_{r,j} + H)$. The improved label θ_l' is given by:

$$\theta_l' = (\eta_l^0,\ c_l + (\alpha_0 - \alpha_r)\min(\lambda_{0,j}, \sigma_{0,j} + H),\ \eta_l^b, \eta_l^r - \min(\lambda_{r,j}, \sigma_{r,j} + H),\ \psi_l^r,\ \psi_l^w) \quad (109)$$

After both slacks have been removed, θ_l' is stored in *OPEN*. A pseudocode representation of this path update process is presented in Appendix L. The $\tilde{\theta}_i$ used in the section Expansion Criteria considers the maximum improvements that could be obtained (i.e., large H) and is defined as:

$$\tilde{\theta}_i = (\eta_i^0,\ c_i + (\alpha_0 - \alpha_r)\lambda_{0,i},\ \eta_i^b,\ \eta_i^r - \lambda_{r,i},\ \psi_i^r,\ \psi_i^w) \quad (110)$$

In summary, the algorithm will, at first, only explore the minimum waiting times necessary to satisfy the time-windows of nodes reachable without resting, not considering how that affects the stops that follow. When the algorithm reaches a state that indicates that earlier decisions can be improved, the current partial solution is updated according to the decisions that will be tested at the current state. This method reduces the number of unnecessary labels generated and also handles the continuous variables without discretizing and testing the whole decision space, which would be computationally expensive. Appendix K presents an optimality proof for the algorithm proposed.

A^* : The A^* algorithm is used during expansion to discard labels that cannot generate solutions with cost lower than the current upper-bound. Let v_t be the destination node, UP be an upper-bound for the optimum solution's cost, and $LOW(v_i, v_t, \theta_i)$ be a lower-bound for the cost of a trip from v_i to v_t with initial resources θ_i. If $c_i + LOW(v_i, v_t, \theta_i) > UP$, then θ_i is discarded. Faster termination can be achieved by setting a tolerance $\epsilon > 0$, and replacing the condition by $c_i + LOW(v_i, v_t, \theta_i) > UP - \epsilon$. By accepting only labels that can improve the current upper bound by at least ϵ, running speed is reduced while the solution obtained is kept within ϵ of optimality.

In order to find a lower bound for the cost of solutions that can be generated from a given state $x_i = (v_i, \theta_i)$, a relaxed version of the problem is solved. The minimum driving time from the current node to the destination is used to calculate the minimum trip duration until the destination, relaxing the time-window constraints and allowing the driver to rest anywhere. Let $D_{HOS}(d, \theta)$ represent the minimum duration of a HOS-compliant trip with d driving hours and initial resource vector θ, assuming the driver can rest anywhere, and without considering service time and time-window constraints, i.e., if a driver were at the beginning of an empty straight road of length d km where he/she can rest anywhere, given an initial resource vector

θ, how long would he/she take to reach the end of the road without breaking the HOS regulations. Whenever a new state x_i is generated, we use the minimum driving time from v_i to v_t, $\mathcal{D}_d(v_i, v_t)$, and the label's resource vector θ_i to calculate a lower bound $D_{HOS}(\mathcal{D}_d(v_i, v_t), \theta_i)$ for the time spent driving or resting until v_t considering only HOS restrictions. Appendix I describes the method used to calculate $D_{HOS}(d, \theta)$ with and without considering the current HOS resource values. Let $D_s(v_i, v_t)$ and $D_l(v_i, v_t)$ represent the service time and minimum distance between nodes v_i and v_t, given by:

$$\mathcal{D}_s(p,q) = \begin{cases} 0, & \text{if } (p,q) \in A' \setminus A^s \\ \min(\Delta_{pq}), & \text{if } (p,q) \in A^s \\ \min\limits_{(p,k)\in A'} (\mathcal{D}_s(p,k) + \mathcal{D}_s(k,q)), & \text{o/w} \end{cases} \quad (111)$$

$$\mathcal{D}_l(p,q) = \begin{cases} 0, & \text{if } (p,q) \in A' \setminus A^d \\ l_{pq}, & \text{if } (p,q) \in A^d \\ \min\limits_{(p,k)\in A'} (\mathcal{D}_l(p,k) + \mathcal{D}_l(k,q)), & \text{o/w} \end{cases} \quad (112)$$

where $A^s \subset A'$ is the set of *service* arcs. The lower bound $LOW(v_i, v_t, \theta_i)$ for the cost of a trip from v_i to v_t with initial resource θ_i is given by:

$$LOW(p,q,\theta) = \alpha_r D_r(p,q,\theta) + \alpha_s D_s(p,q) + \alpha_d \mathcal{D}_d(p,q) + \beta_d \mathcal{D}_l(p,q) \quad (113)$$

$$D_r(p,q,\theta) = D_{HOS}(\mathcal{D}_d(p,q), \theta) - \mathcal{D}_d(p,q) \quad (114)$$

where $D_r(\cdot)$ represents how much time from $D_{HOS}(\cdot)$ would be spent resting.

As $D_{HOS}(d, \theta)$ is independent of the network topology and time-windows, it can be calculated beforehand for different values of d and θ, and used in any problem instance. However, discretizing the possible values of d and θ and storing the results for every combination would require a lot of storage space. Setting all HOS related resources to zero and discretizing only d when calculating and storing $D_{HOS}(\cdot)$ greatly reduces storage space, but also generates a looser bound. Another option is to store the results for a limited number of (d, θ) combinations, and use those results to approximate others or accelerate their computation during run time. The formulation presented in Appendix I uses $D_{HOS}\big(d, (0,0,\dots,0)\big)$ to calculate $D_{HOS}(d, \theta)$ for a general θ. We tested storing only the results of $D_{HOS}\big(d, (0,0,\dots,0)\big)$ and calculating the others during run time, but the lower-bound improvement was not enough to compensate for the extra computations. Therefore, in our experiments, the lower-bound is calculated without considering current HOS resources.

The initial upper bound can be taken from a known sub-optimal solution, or calculated based on the planning horizon and max distance to destination, and updated as the algorithm finds better solutions. If the initial upper bound is set too high, the solve time can increase significantly. However, we noticed that, as the algorithm can quickly determine that the problem is infeasible when the upper bound is too low, it is efficient to set a low upper bound and increase it gradually until the problem becomes feasible.

Dominance Rules

Let $\theta_i = (\eta_i^0, c_i, \eta_i^b, \eta_i^r, \psi_i^r, \psi_i^w)$ and $\theta_i' = (\eta_i^{0\prime}, c_i', \eta_i^{b\prime}, \eta_i^{r\prime}, \psi_i^{r\prime}, \psi_i^{w\prime})$ denote labels for two different partial solutions ending at the same node v_i. If θ_i dominates $\theta_i{}'$, then for every solution that can be generated by expanding $\theta_i{}'$, there is a better solution that can be generated by expanding θ_i. For example, assume $\theta_i = (10,9,7,10,10,10)$ and $\theta_i{}' = (10,10,7,10,10,10)$, and that neither label can be improved. As the arrival times are equal, i.e., $\eta_i^0 = \eta_i^{0\prime}$, any time-window that can be satisfied starting from $\theta_i{}'$ can also be satisfied starting from θ_i. As all HOS related resources are also equal, no label has an advantage regarding when a rest stop will be required. Therefore, as θ_i has a smaller cost ($c_i = 9$, $c_i' = 10$), this cost advantage will be carried to all paths generated from θ_i, making them cheaper than paths generated from $\theta_i{}'$. The example portrays the base case for dominance check, when $\eta_i^0 = \eta_i^{0\prime}$ and neither label can be improved. In this case, if all resources in θ_i are smaller or equal to the resources in $\theta_i{}'$, with at least one being strictly smaller, θ_i dominates $\theta_i{}'$. Including the effects of possible label improvements makes the conditions a little more complicated. Similar dominance rules were used in (Asvin Goel 2012). However, as they consider that early arrivals are allowed and that label improvement is always performed before dominance checks, these rules cannot be used in our model.

As we do not allow early arrivals, when $\eta_i^0 < \eta_i^{0\prime}$, paths generated from θ_i may be unable to satisfy time-window constraints of downstream nodes, otherwise satisfied by paths generated from $\theta_i{}'$, due to arriving too early. However, by using the particular structure of the problem, dominance rules for when $\eta_i^0 \leq \eta_i^{0\prime}$ were established for *weekly rest*, *daily rest* and *break* nodes. These are the only types of nodes with controllable outgoing arc duration, which can be used to equalize the arrival times at the next node. Thus, the dominance conditions were derived by using their REFs to define when every label generated from $\theta_i{}'$ through a decision (j, δ) is dominated (after label improvements) by the label generated from θ_i through the decision $(j, \delta + \gamma)$, where $\gamma = \eta_i^{0\prime} - \eta_i^0 \geq 0$. In order to reduce the number of dominance checks performed, only labels assigned to rest nodes ($v_i \in V_b \cup V_r \cup V_w$) are checked. As rest nodes' dominance rules are not restricted to labels with matching arrival times, they have a greater potential for identifying and discarding inferior solutions. The dominance rules are given by:

$$v_i \in V_b \cup V_r \cup V_w \tag{115}$$

$$\eta_i^0 + \gamma = \eta_i^{0\prime} \tag{116}$$

$$\gamma \geq 0, \tag{117}$$

$$\eta_i^b \leq \eta_i^{b\prime} \tag{118}$$

$$\psi_i^r \leq \psi_i^{r\prime} \tag{119}$$

$$\psi_i^w \leq \psi_i^{w\prime} \tag{120}$$

$$c_i - c_i' \leq g_c \tag{121}$$

$$\eta_i^r - \eta_i^{r\prime} \leq g_r, \quad \text{if } v_i \in V_b \tag{122}$$

$$c_i - c_i' \leq \alpha_0 w_0 - \alpha_r(\gamma + w_0) \tag{123}$$

$$\eta_i^r - \eta_i^{r\prime} \leq -\gamma - w_r, \quad \text{if } v_i \in V_b \tag{124}$$

where g_c, g_r, w_0 and w_r are defined as:

$$\mathrm{g_c} = \begin{cases} \alpha_0 w_0 - \alpha_r(\gamma + w_0), & if \ \ \lambda_{0,i} \leq \sigma_{0,i} + \gamma \\ \alpha_0(\mathrm{y_0} - \gamma) - \alpha_\mathrm{r} \mathrm{y_0}, & if \ \ (\lambda_{0,\mathrm{i}} > \sigma_{0,\mathrm{i}} + \gamma) \wedge (\lambda'_{0,\mathrm{i}} \leq \sigma'_{0,\mathrm{i}}) \\ \min\big(\alpha_0(\mathrm{z_0} - \gamma) - \alpha_\mathrm{r} \mathrm{z_0}, \alpha_0 \mathrm{w_0} - \alpha_\mathrm{r}(\gamma + \mathrm{w_0})\big), & o/w \end{cases} \tag{125}$$

$$\mathrm{g_r} = \begin{cases} -\gamma - \mathrm{w_r}, & if \quad \lambda_{r,i} \leq \sigma_{r,i} + \gamma \\ -\mathrm{y_r}, & if \quad (\lambda_{r,\mathrm{i}} > \sigma_{r,\mathrm{i}} + \gamma) \wedge (\lambda'_{r,\mathrm{i}} \leq \sigma'_{r,\mathrm{i}}) \\ \min(-\mathrm{z_r}, -\gamma - \mathrm{w_r}), & \mathrm{o/w} \end{cases} \tag{126}$$

$$w_0 = \lambda'_{0,i} - \lambda_{0,i}, \qquad y_0 = \lambda'_{0,i} - \sigma_{0,i}, \quad z_0 = \sigma'_{0,i} - \sigma_{0,i} \tag{127}$$

$$w_r = \lambda'_{r,i} - \lambda_{r,i}, \qquad y_r = \lambda'_{r,i} - \sigma_{r,i}, \quad z_r = \sigma'_{r,i} - \sigma_{r,i} \tag{128}$$

where (115) defines for which nodes these rules are applicable. Conditions (116) and (117) define γ and restrict its allowed values, so that labels can only be dominated by labels with a lower or equal arrival time. Conditions (118)-(124) describe the conditions on the cost and HOS resources. (118) is always valid as this is checked only at rest nodes, where the elapsed time since last *break* resource is reset, i.e., $\eta^b = 0$. As the resources representing cost (c) and elapsed time since last *daily rest* (η^r) are affected by label improvements, the dominance rule must check if the solutions generated from $\theta_i{}'$ are inferior to ones generated from θ_i even if label improvement is performed at the current node, or at a downstream node. (121) and (122) include the effects of the currently available label improvements. (123) and (124) consider the impact of potential label improvements if the slacks cannot be completely used at the next decision. A partial solution might appear better at the current node, but be less flexible to adapt to downstream constraints, e.g., if θ_i was generated by a path with very loose time-windows, it will have more flexibility to adjust upstream arrival times to reduce the amount of unnecessary off-duty time at downstream nodes. (125)-(128) define the auxiliary variables used to define (121)-(124). We can say that θ_i dominates $\theta_i{}'$ if conditions (115)-(124) are satisfied, with at least one among (118)-(124) being a strict inequality. Appendix J describes in detail how the dominance rules were obtained.

Case Study

To evaluate the proposed algorithm a set of test scenarios is created using the network shown in Figure 115. The network is based on a route going from San Diego to Seattle via the I-5 freeway, and includes some possible detours. For easier visualization, the parking lots (p_i

nodes) are displayed along the routes. On the actual graph, they are outside the routes, as in Figure 112, to allow for them to be bypassed. For simplicity, the distances between the routes and the parking lots are set to 0.

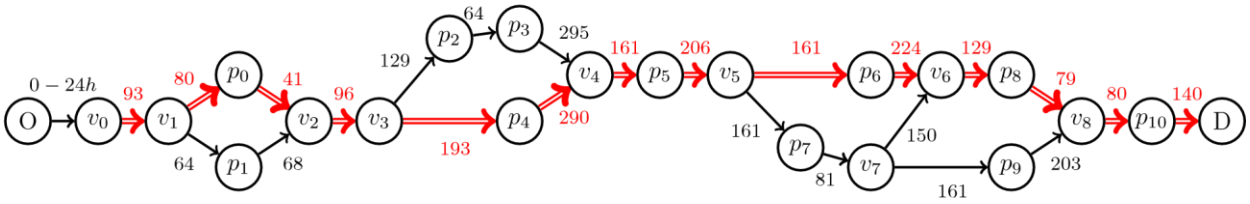


Figure 115. Network used for experiments. Based on a main route going from San Diego to Seattle via the I-5 freeway indicated in red with double arrows, together some possible detours indicated with black arrows.

We want to study the impact of HOS rules and parking constraints on the shortest path to the destination and compare the cost of taking parking availability information into account with the estimated costs of not doing so. Furthermore, to illustrate the importance of including parking information early in the planning process as opposed to only doing so at the scheduling stage, we also study scenarios where the driver is restricted to the main route. As mentioned in (Sun et al. 2013), not all drivers are allowed to choose their routes freely. If the carrier determines the route and does not consider parking availability information during route assignment, the trip's duration and cost may be significantly larger than estimated by the carrier, especially if the driver is not allowed to adjust the route. In this case study, we take the trip duration as the cost, and we assume that there is no extra cost for waiting at the origin, i.e., $\beta = \alpha_0 = 0$ and $\alpha^d = \alpha^s = \alpha^b = \alpha^r = \alpha^w = 1$.

The following aspects are considered when evaluating the algorithm:

- What is the estimated cost of not using parking availability information during planning?
- How often the minimum duration path includes one or more alternative routes?
- Average driving times under varying alternative routes' speed and in their absence.
- Average trip duration under varying alternative routes' speed and in their absence.

Scenarios

Taking longer routes (in terms of driving time) is beneficial only if the time saved due to better parking availability conditions is larger than the increase in driving time. The experiments focus on studying how the problem's solution is affected by the usage of parking availability information, by the main route's parking availability time-windows' distribution and by the difference in driving time between the alternative routes and the main one (by varying the average speed on the alternative routes).

Table 23 shows the distributions used to sample the start and end times of the three types of parking availability time-windows used in the tests. It is assumed that each day has a single time-window. The daily parking availability time-windows are defined by sampling a start and

an end time from normal distributions with the parameters listed in the table. The parameters given are in hours and follow the 24h format. The route drawn in red with double arrows on Figure 115 is taken as the main route. This route has the shortest driving time from origin to destination, so, in the absence of HOS and parking constraints, the optimal solution would go through this path. All road sections not included in the main route are considered alternative routes. For simplicity, we set all main route links' average speed to 75km/h. The average speed of alternative route links vary with the scenario. Table 24 shows the parameters used for each set of scenarios, and how many instances were tested. Scenarios 10 to 12 restrict the driver to the main route, so alternative routes' parameters are shown as *blocked*. The time-window configurations for departure and delivery times are shown in the lower part of Table 23. The departure time has a single time-window on the first day (0h to 24h), whereas the delivery time at the destination (D) has daily time-windows (8h to 16h). Besides the single client case just described, the twelve scenarios were also tested for the case when node v_4 is a client with a daily time-window (12h to 16h) and zero service time. All other parameters are the same as for the twelve original scenarios.

Table 23. Experiment Parameters

Parking Time-windows Distribution		
	Start Time (h)	**End Time (h)**
Narrow	N(9,1)	N(16,1)
Medium	N(7,1)	N(19,1)
Wide	N(5,1)	N(22,1)

Other Time-windows		
	Start Time (h)	**End Time (h)**
Departure (unique)	0	24
Delivery at D (daily)	8	16
Delivery at v_4 (daily, 2 clients case)	12	16

Table 24. Experiment Scenarios

		Time-windows		Average Speed (km/h)	
Scenario	**Instances**	**Main Route**	**Alternative Routes**	**Main Route**	**Alternative Routes**
1	100	Wide	Wide	75	75
2	100	Medium	Wide	75	75
3	100	Narrow	Wide	75	75
4	100	Wide	Wide	75	70
5	100	Medium	Wide	75	70
6	100	Narrow	Wide	75	70
7	100	Wide	Wide	75	65
8	100	Medium	Wide	75	65
9	100	Narrow	Wide	75	65
10	100	Wide	blocked	75	blocked
11	100	Medium	blocked	75	blocked
12	100	Narrow	blocked	75	blocked

Estimating the cost of disregarding parking information

As previously stated, the truck parking shortage has multiple negative consequences for drivers, industry and society. While it is hard to accurately estimate the costs involved, in this section we propose a recourse function used to simulate how the driver would react if he/she were to arrive at a parking facility and find it unavailable, as well as the costs incurred. For each one of the scenarios explained previously, a solution is generated without including parking availability constraints. Then the recourse function is applied to each scenario's solution in order to simulate the driver's reactions. Three situations are considered: parking is available, parking is

unavailable and there is another facility within reach, parking is unavailable and no facility is within reach.

Parking is available: In this case the driver follows the schedule.

Parking is unavailable and no other facility is within reach: A new route and schedule are generated (also without considering parking availability) starting from the exit of the parking facility where the driver is currently at. No cost or time penalties are applied.

Parking is unavailable and no other facility is within reach: It is assumed that the driver spends 0.5 hours searching for parking, so a 0.5h time penalty is imposed. It is also considered that the driver is likely to park at an unofficial parking location, so a cost penalty is applied. A new route and schedule are generated (also without considering parking availability) starting from the entrance of the parking facility where the driver is currently at. The required rest is taken at the current location as if it was an unofficial parking.

In these experiments, the cost was measured in hours, so the cost penalty is also converted to hours. As it is hard to estimate the cost of the financial risk a driver is undertaking every time he/she parks illegally, a sensitivity analysis is performed with multiple penalty values. The values tested are: 2h, 4h, 6h, 8h, and 10h. For comparison, considering an hourly marginal cost of operation of $71.78 (Murray and Glidewell 2019), the 4h (~$287) penalty is comparable to the fine for non-emergency stops on freeways in California (~$238) (Judicial Council of California 2019), and factoring in accident risks would increase penalties further. The risk of getting involved in an accident or being fined varies with the region, route, time of day, and type of vehicle, as do accidents' average severity and cost. Nevertheless, when accidents do happen, costs can be substantial. In 2005, the average cost of truck crashes was $91 thousand overall, $195 thousand for injury crashes, and $3.6 million for crashes involving fatalities (Zaloshnja and Miller 2007); around 1.2 thousand, 2.7 thousand and 50 thousand hours of operation, respectively. The costs presented in (Zaloshnja and Miller 2007) are in 2005 dollars, and were calculated using 2002 values for the Value of Statistical Life; current values should be significantly higher. Therefore, a 1% accident probability would result in penalties higher than 12 hours, already exceeding the tested penalties.

Other indirect costs may include a drop in driver satisfaction, higher insurance premiums, loss of clients, etc. Data on all these factors are required for a company to accurately estimate the financial risks of illegal parking. In our experiments, we assigned the same penalties to all locations, but, with enough data, appropriate values can be set for each region/location.

Results

Estimated Cost of Disregarding Parking Information

Figure 116 shows the average costs (after adjustments by the recourse function) of trips that are planned without taking into account parking availability information. Both the single client (Figure 116a) and the two client (Figure 116b) cases are shown. In the scenarios tested, the average costs did not vary with the alternative routes' average speed, so only the plots for

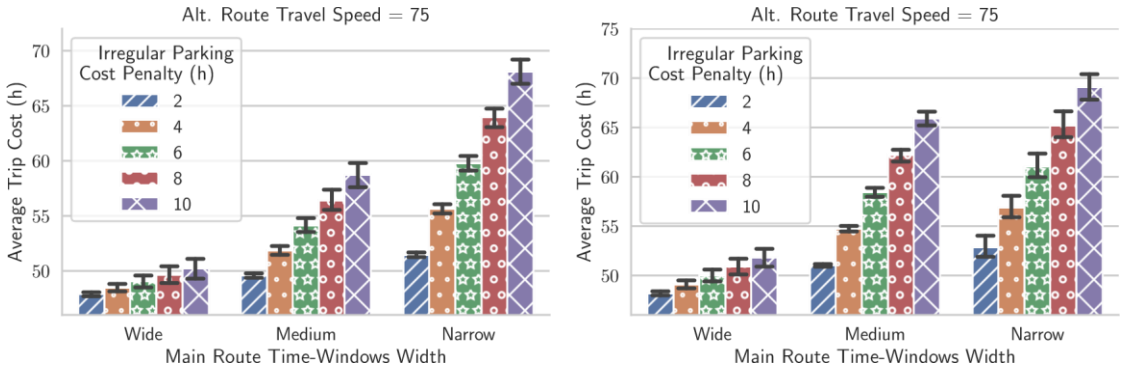
75km/h are shown, the others can be assumed equal. This result is likely caused by the network used, and is discussed later in Rerouting behavior. Figure 117 shows the average trip cost when parking availability information is considered during planning. In this case, as we assumed that parking availability is guaranteed within the time-windows, the schedules generated are always feasible and no penalty or rerouting is required. Therefore, the trip cost equals the trip duration.

When comparing the plots in Figure 116 and Figure 117, it can be seen that the advantage of considering parking availability information varies substantially with the scenario and penalties considered. As previously stated, stopping early to guarantee appropriate parking affects driver productivity (Boris and Brewster 2018). The same is true for extending rest periods to guarantee parking at the next facility. The following question can be raised. Which one is more expensive: waiting longer to guarantee appropriate parking, or the financial risks of allowing irregular parking?

When parking is abundant (wide time-windows), accounting for parking availability information is always beneficial, as the extra waiting times are not significant. However, as parking becomes scarce, the effect of the waiting times required to guarantee parking can be considerable, especially if alternative routes are slow or nonexistent. Stricter constraints cause longer waiting times and, consequently, higher costs. This effect can be seen on the cost increase caused by narrowing time-windows, and by adding a second client.

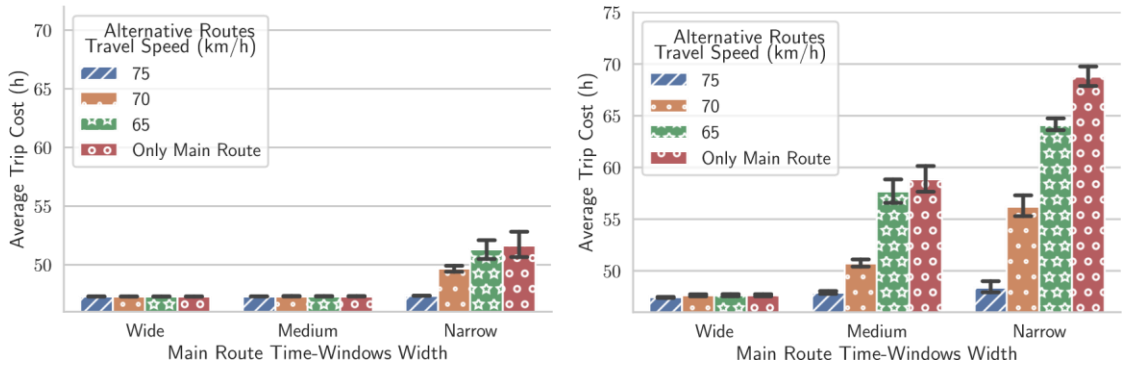
Nevertheless, situations with strict constraints are when drivers need the most help. In these cases, the financial risk of illegal parking is the defining factor. Focusing on the more restrictive two-clients scenarios of Figure 116b and Figure 117b, it can be seen that whether using parking information is cheaper or not, depends on the penalty values considered. For example, if drivers are fined every time they park illegally, but the region they work at is very safe and no accidents or robberies happen, it would put the penalty for illegal parking at around 4h of operation (assumptions explained in the section Estimating the cost of disregarding parking information), i.e., the yellow bar with small circles on Figure 116b. In this particular case, if parking is very scarce along the main route (narrow time-windows) it is cheaper to consider parking information only if the travel speed of the alternative routes is 75km/h. If the alternative routes are slow, the driver would likely prefer to park illegally and pay the fines.

The difficulty in accurately gauging this risk may lead drivers and companies to mistakenly assume that it is cheaper to disregard parking constraints. Although these penalties are very hard to estimate, the litigious environment that the trucking industry has been facing is pushing these costs up (Lysiak 2019). Some insurance companies are already pushing their clients for the adoption of safety technologies, such as collision avoidance or camera systems, and legal parking planning practices may be a good complement to the industry's safety standards.



(a) Single client **(b) Two clients**

Figure 116. Average trip cost (when disregarding parking information) for different irregular parking penalties according to the type of parking availability time-windows considered for the main route. In this experiment, the results did not vary with the alternative routes' travel speed so the plots for other speeds were omitted.



(a) Single client **(b) Two clients**

Figure 117. Average trip cost/duration (when using parking information for planning) according to parking availability time-windows (main route) and travel speed (alternative routes) used.

Rerouting behavior: We expected that in some cases the HOS constraints would force the driver to take an alternative route when rerouting. However, in the tested scenarios, whenever parking is unavailable the driver does not have enough remaining driving time to reroute and head for a different parking facility. This causes the driver to use irregular parking instead of rerouting. After resting, the driver is less likely to take an alternative route due to HOS constraints. As this happens for all alternative route speeds, the average cost does not vary with the speed parameter.

Effect on Shortest Path

Single Client

Figure 118, Figure 119 and Figure 117a show how often alternative routes are used and how the average driving time and average trip duration vary with the parameters used for each scenario. It can be seen that when the main route has wide or medium parking time-windows, there is little or no benefit in taking alternative routes. In these cases, the main route has enough parking availability and alternative routes are rarely used. However, this does not hold anymore when the parking availability time-windows are narrow. When parking is scarce on the main route, alternative routes can significantly lower costs. The usage of alternative routes is more pronounced when their travel speed does not differ much from the main one's, but it can still be seen even when the speed is lower.

It is also important to note that the averaging dilutes the contribution of the instances that used the alternative paths. For example, Figure 118 shows that alternative routes are used only 60% of the time when the time-windows are narrow and the speed is 70km/h, so the $\sim 2h$ improvement in cost seen in Figure 117a is actually caused by only half of the instances. Similarly, the $\sim 0.5h$ increase in driving time seen in Figure 119 is also caused by only half of the instances. So, in the end, there are 60 instances where paths are, on average, $\sim 1h$ longer, but still generate costs $\sim 4h$ shorter on average.

When more restrictions are applied and the problem complexity increases, the benefits of considering multiple paths become more evident, as can be seen on the case with two clients.

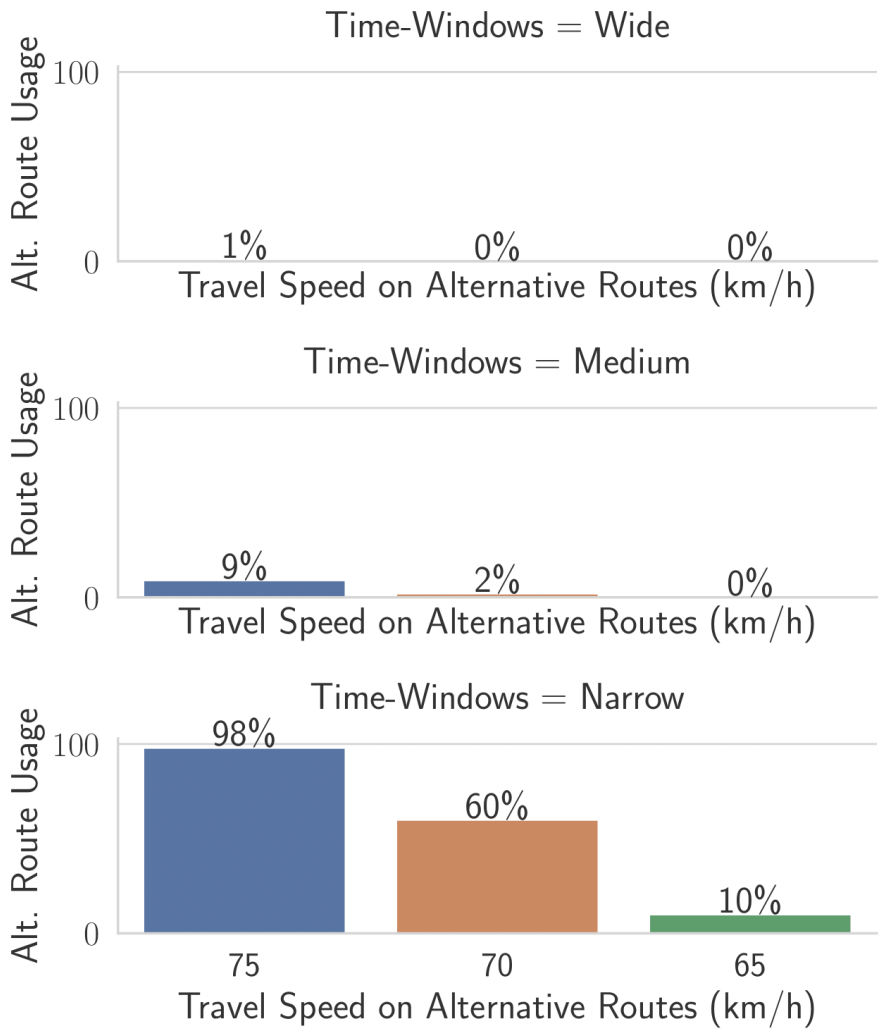


Figure 118. Number of instances that used alternative routes for the single client case, according to the time-windows (main route) and travel speed (alternative routes) used.

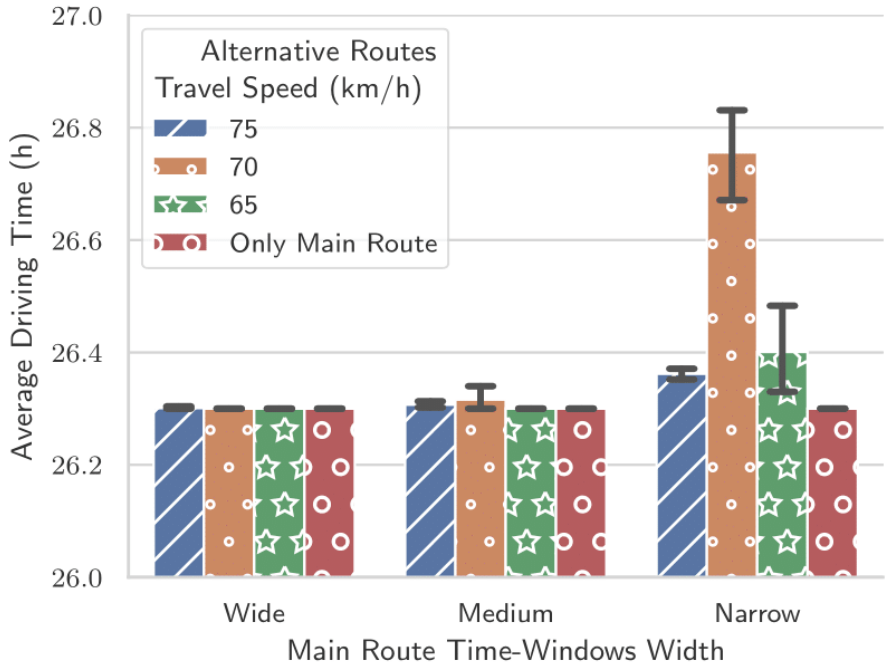


Figure 119. Average driving time of solutions for the single client case, according to the time-windows (main route) and travel speed (alternative routes) used.

Two Clients

Figure 120, Figure 121 and Figure 117b show how often alternative routes are used and how the average driving time and average trip duration vary with the parameters used for each scenario. In this case, the advantage of considering alternative routes is clear and they are used significantly more often when the time-windows are medium or narrow. Even in the case where the travel speed of the alternative routes is 65km/h and the average driving time is increased by more than 1 hour, there is an improvement of almost 5 hours to the average trip duration.

As expected, the benefits of considering alternative routes increase when they are not significantly longer than the main route and when the main route has limited parking availability. Moreover, it is important to note that alternative routes can also impact problem feasibility. In these experiments, the planning horizon is set high enough so that all instances would have a feasible solution. However, many of the solutions with larger costs could become infeasible with a shorter planning horizon.

When solving the VRP, the shortest paths are usually assumed known, so no alternative paths are considered when running the scheduling subroutines. As parking availability is usually not considered, these methods would be expecting the trip cost/duration to be similar to the values for wide parking availability time-windows shown in Figure 117. However, depending on the severity of the region's parking shortage, a driver that attempted to plan the trip accounting for parking availability could face considerably larger costs like the ones shown for narrow parking time-windows. With the worst cost happening if the driver tried to stick to the exact path considered by the company (bars for 'Only Main Route' case in Figure 117). Drivers that do not

consider parking when planning would be subjecting themselves to safety risks as the ones estimated in Figure 116. In either case, the trip costs used by VRP algorithms that do not consider parking availability and that assume a fixed path between any two clients can be a complete misrepresentation of the costs the driver or company will actually experience.

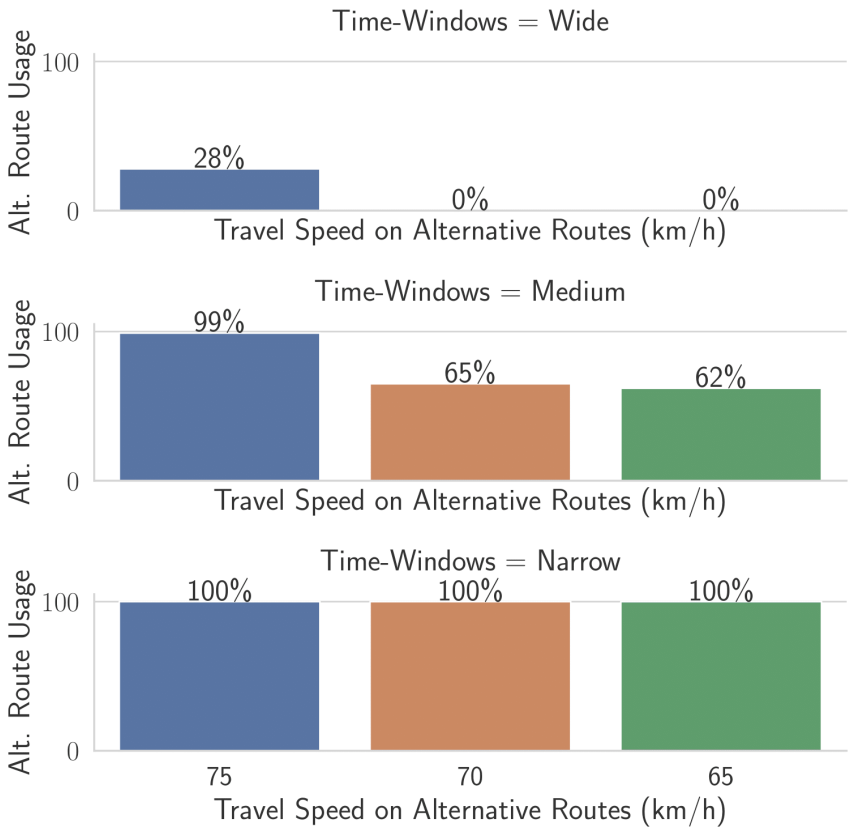


Figure 120. Number of instances that used alternative routes for the two clients case, according to the time-windows (main route) and travel speed (alternative routes) used.

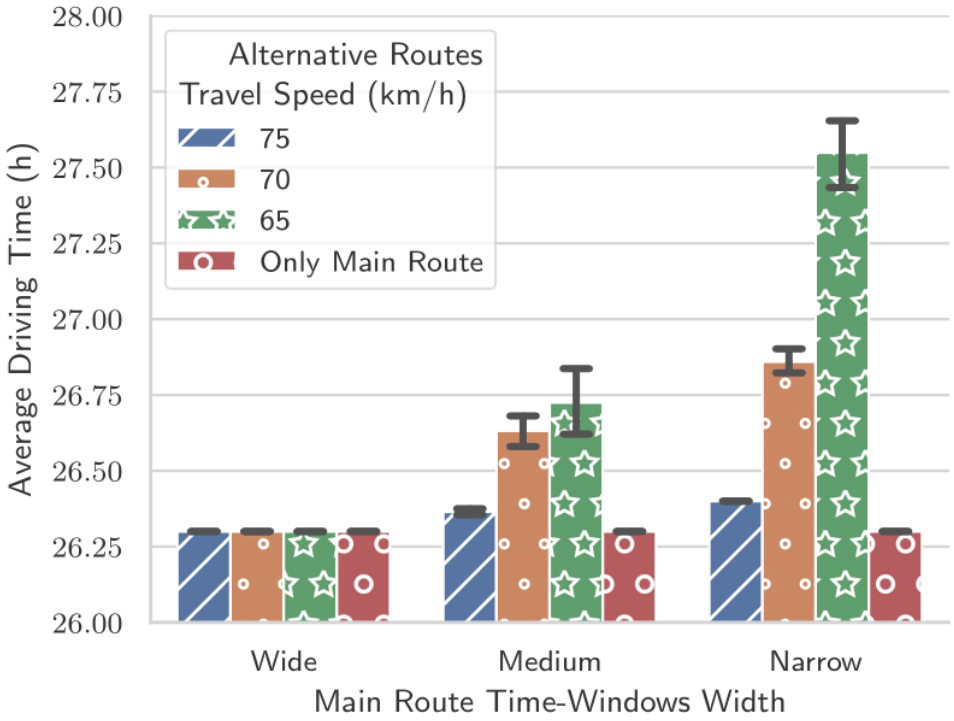


Figure 121. Average driving time of solutions for the two clients case, according to the time-windows (main route) and travel speed (alternative routes) used.

Randomized Networks Experiments

In this section we present experiments measuring the performance of the algorithm on randomly generated networks. The first experiment focuses on networks with a small number of clients and a large variation on the driving time required to reach the last client. The second experiment focuses on varying the number of clients with narrower range of driving times.

Setup

The graphs were created with the following characteristics

- Network organized in layers.
- Random number of layers between consecutive clients.
- Random number of nodes per layer.
- Every node has a fixed probability of having an edge linking it to each node of the following layer.
- Every node that has an incoming edge, has at least one outgoing edge.
- Branches leading to TPLs are inserted randomly along each edge according to a Poisson process.
- Multiple values for the average spacing between TPLs were tested.
- Multiple values for the probabilities of a TPL having a *narrow*, *medium* or *wide* availability window were tested.

For each graph, five possible parking shortage levels, with increasingly strict parking conditions, were tested. The parking shortage levels in Table 25 define the probability of each TPL having a *Narrow*, *Medium* or *Wide* time-window. These three types of time-windows are defined in Table 23. For each graph and parking shortage level, 20 problem instances were generated. All clients have daily time-windows from 9:00 to 17:00. The algorithm's tolerance is set to find solutions within 0.25 hours of optimality.

Table 25. Parking Shortage Level

	Probability of each type of time-window		
Parking Shortage Level	**Narrow**	**Medium**	**Wide**
1	0.1	0.2	0.7
2	0.2	0.3	0.5
3	0.33	0.33	0.34
4	0.5	0.3	0.2
5	0.7	0.2	0.1

Table 26 Random Networks Configurations

Avg. Spacing (km)	Graphs	Total Instances
50	84	8400
100	86	8600
150	67	6700

Experiment 1

In this experiment, we studied how the parking shortage affects total trip duration, and our algorithm's performance in networks of varying sizes. Note that our focus is long-haul trucking; drivers travel long distances to visit a relatively small number of clients. When studying larger networks, our interest is in varying the trip duration, the number of possible paths and the number of TPLs along each path. The number of TPLs varies from 0 to 288, the total driving time of the solutions found for each instance varies from 3.3 to 60.7 hours, and the total number of clients varies from 1 to 5. Table 26 presents the average spacing between TPLs used to generate the test networks, as well as how many networks were created with each spacing value.

Performance

We implemented our algorithm in Python 3.8, and all experiments were run on a Intel Core i5, 3.1GHz CPU with 8Gb of RAM. We would like to note that the obtained running times could be reduced by implementing the algorithm in faster languages, such as C, C++ or Java, however, this is not the focus of this work. Figure 122 shows how the average running time varies with the solutions' total driving time and the total number of TPLs present in each instance. Both the number of TPL choices and the number of rests that need to be scheduled increase the

problem's complexity, so we concluded that this was the most meaningful way of presenting the results. The method still presents some scalability issues, with running times increasing sharply when trips approach the weekly driving limit and a large number of TPLs is considered. Nevertheless, most scenarios tested have an average running time below 200 seconds, which is reasonable for a truck driver planning his itinerary for the following week. Also, although the current performance does not allow for this method to be used at every iteration of a VRTDSP algorithm, it can still be used as a method to refine the final routes chosen.

We also noticed that the time windows can significantly affect running time. Depending on the parking shortage level and instance considered, the same network may have vastly different running times.

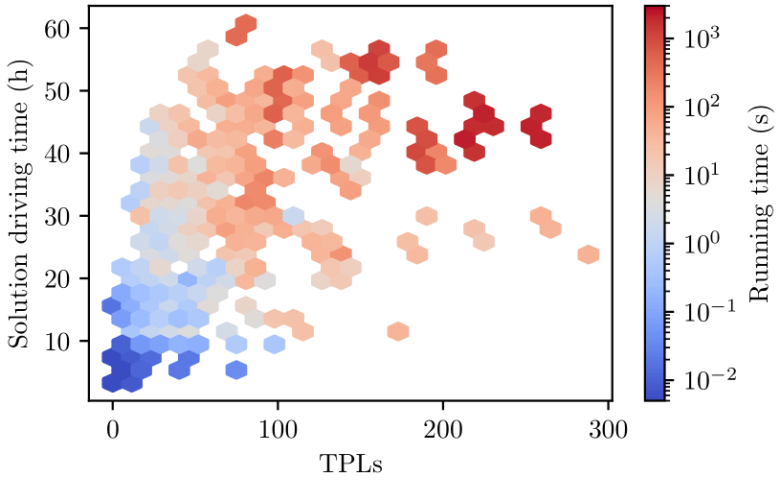


Figure 122. Average running time over randomized networks.

Parking Shortage's Impact

Table 27 presents the percentual increase in trip duration of instances with parking shortage level 2 to 5 relative to level 1 instances. Level 5 instances have 5.1% higher trip duration overall, and even considering only instances with an average spacing of 50km between TPLs the increase is 2.3%, which is not negligible given the trucking industry's size. The case study presented in the section Case Study illustrated how parking availability conditions can have a significant impact on trip durations. This experiment over randomized networks further supports this claim, by showing that significant impacts are seen even on more complex networks.

Table 27. Average Trip Duration Increase

	Parking Shortage Level				
Avg. Spacing (km)	**1 (baseline)**	**2**	**3**	**4**	**5**
50	0%	+0.2%	+0.8%	+1.3%	2.3%
100	0%	+0.7%	+1.4%	+2.5%	4.2%
150	0%	+0.6%	+2.7%	+5.9%	8.3%
200	0%	+0.9%	+3.5%	+5.6%	7.1%
Overall	0%	+0.6%	+1.9%	+3.4%	+5.1%

Experiment 2

This experiment focuses on studying how the algorithm behaves when the number of clients is increased. As stated previously, our focus is on long-haul, which implies a small number of clients visited over a long trip, so the scenarios tested in this experiment diverge from our intended use case. In order to avoid the effects of driving time (and number of TPLs) on the running time, shown in Figure 122, the networks used have driving times around 30-40 hours, and the number of clients varies from 1 to 46 with zero service time. Therefore, the average distance between clients decreases as the number of clients increases. These networks have only 1 intermediate layer between clients, so there are not many possible routes before accounting for the detours necessary to reach TPLs. The average spacing between TPLs is set to 100 km and most networks tested have between 20 and 40 TPLs. Experiment 2 was run in the same computer as experiment 1.

Figure 123, Figure 124, and Figure 125 show how the running time varies with the number of clients, trip duration, off-duty time, and driving time. In Figure 123, we see that having to accommodate constraints of a larger number of clients in a relatively short trip causes large increases in trip duration despite similar driving times. We believe that the number of clients' effect on running time is caused mostly by this increase in trip duration. Figure 123 and Figure 124 show that the scenarios with larger trip duration and off-duty times are also the ones with higher running times, possibly due to needing to schedule a larger number of rest stops or testing longer rest duration values for each stop. Similarly, Figure 125 shows that running time increases with the off-duty time ratio, and also that the off-duty time ratio increases with the number of clients.

A point worth noting is that this type of scenario with a large number of clients close to each other does not fit long-haul trucking usual jobs, it might be closer to what happens in local trucking, where drivers stay within a smaller region. However, in the U.S., short-haul drivers that operate within a 150 air-mile radius of their normal work reporting location are subject to different, less restrictive, regulations (Code of Federal Regulations, n.d.). As clients are expected to be widely spaced and require non-negligible service time in the context of long-haul trucking, we judge that the limiting factor for our problem is not the number of clients, but the trips' expected driving/on-duty time, duration and the number of TPLs and paths considered during planning (as seen in Figure 122).

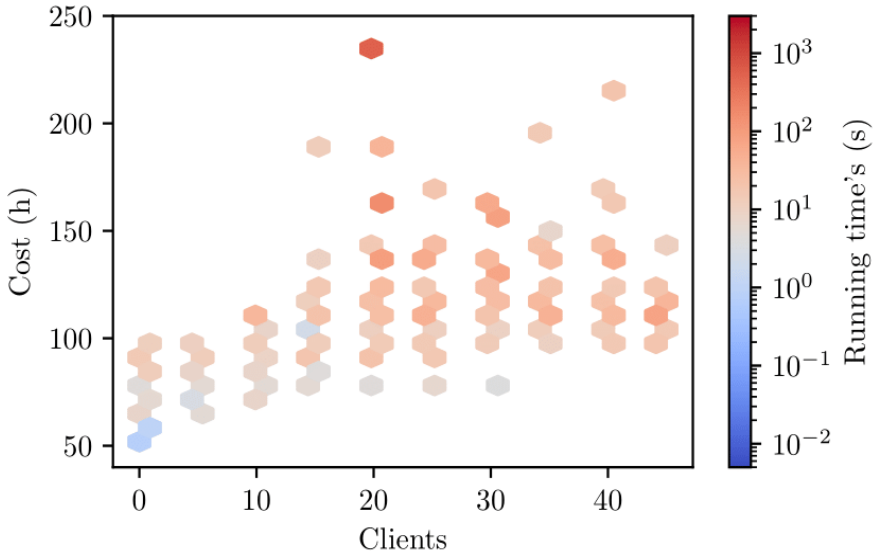


Figure 123. Average running time for instances with varying number of clients and cost (trip duration in hours).

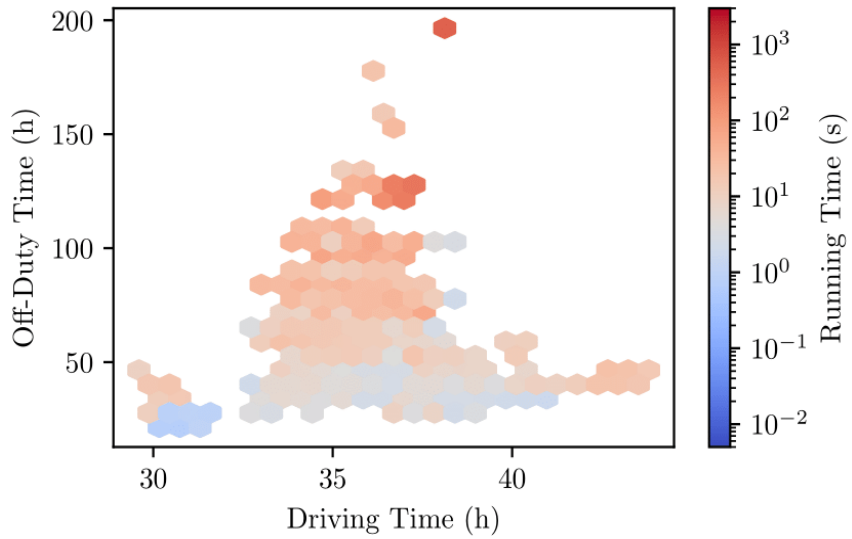


Figure 124. Average running time for instances with varying driving and off-duty time.

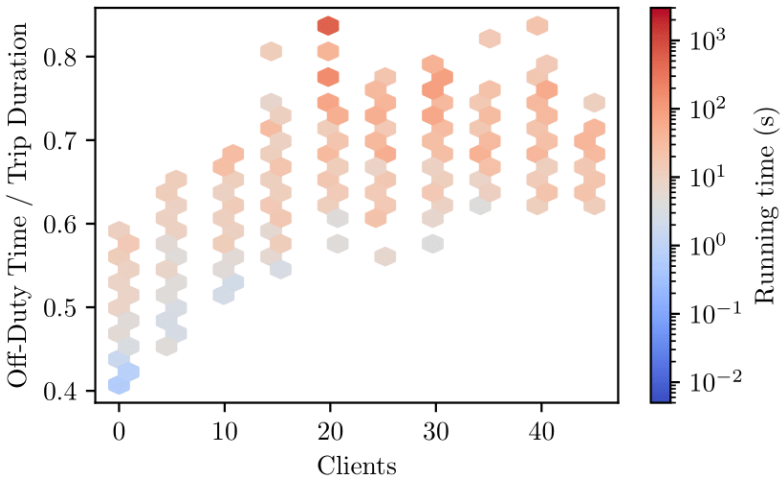


Figure 125. Average running time for instances with varying number of clients and off-duty time ratio (off-duty time/trip duration).

Conclusion

In this study, we introduced the shortest path and truck driver scheduling problem with parking availability constraints (SPTDSP-PA), which integrates parking availability information into the path planning and scheduling of long-haul truck shipments. A resource-constrained shortest path formulation is proposed, and a tailored label-correcting algorithm is developed to solve it by efficiently exploiting the problem's particularities.

We performed a case study on a network based on the US I-5 freeway, under various parking availability conditions and alternative routes' speed. The results show that, as expected, the cost of imposing parking constraints increases as parking availability decreases. Therefore, illegal parking costs are the deciding factor when evaluating the profitability of imposing parking constraints. Depending on the illegal parking penalties considered, the cost savings from preventing said penalties can exceed, or at least offset, the cost increase caused by parking constraints. Due to the high direct and indirect costs of truck-related accidents, we believe that these penalties are likely high, and that safer parking practices would benefit the trucking industry not only from a safety standpoint, but also from an economic standpoint. To illustrate the importance of considering parking information when defining the paths between clients, we studied how parking availability and alternative paths affect trip costs. Considering alternative paths by using the SPTDSP-PA formulation instead of only solving a truck driver scheduling problem can mitigate cost increases due to parking constraints and significantly change estimated trip costs.

Experiments on randomly generated networks showed that, even on networks with more alternative paths and parking options, limited parking availability can have a non-negligible impact on trip duration. In addition, this impact increases substantially in networks where parking facilities are scarcer. The cost/duration increase caused by imposing parking availability constraints can be seen as an estimate of how much drivers and companies would need to spend in order to ensure safe itineraries for the drivers, and prevent accidents and other costs

related to the difficulty of finding appropriate rest locations. By simulating how parking availability can affect trip duration, costs, and illegal parking, our model can aid in infrastructure and policy decisions. In future work, we intend to extend the model to include time-dependent travel times in order to account for the effects of traffic congestion.

Long Haul Battery Electric Truck Planning

The electrification of transport is seen as one of the main paths for emissions reduction. With the increase in performance and variety of EV models in the market, it becomes vital to develop ways to incorporate them into our society efficiently. However, despite continuing interest and investments to promote electric vehicles in goods movement, electric trucks still face many challenges to achieve meaningful market penetration, including limited range, long refueling time, and reduced maximum payload. Research on how to coordinate electric trucks' range and recharging limitations with regulation requirements and charging/parking infrastructure availability can reduce the performance gap between electric and diesel trucks, and promote the adoption of electric vehicles by the trucking industry.

Among the concerns regarding heavy-duty vehicles electrification are the reduced range, longer recharge time and the reduction in maximum payload due to added battery weight. However, it is often overlooked that commercial drivers are required to stop and rest regularly regardless of the vehicle's range. Even if long-haul trucks drive on average 600 miles per day (Smith et al. 2019), they are required to stop for at least 30 minutes every 8 hours and for at least 10 hours after 11 hours of driving time. Adequate infrastructure and efficient trip planning can mitigate the range issue by recharging the vehicle during mandatory stops. Nevertheless, recharging does take longer than refueling and charging stations are not abundant, making charging station availability an issue for battery electric trucks (BETs).

In this section, we integrate availability information (both for parking and charging) in the planning process for BET, thus coordinating the needs of off-duty and recharge time for BETs. In particular, we investigate how the effects of parking availability constraints on BETs compare to their effects on diesel trucks. Our goals are:

- Incorporate battery charge constraints in our shortest path and truck driver scheduling model.
- Study how BETs and diesel trucks' trip cost and duration compare under different parking conditions.

Section based on the publication:

F. Vital, and P. Ioannou, "Effects of Working Hour Regulations and Parking Shortages on Truck Electrification," 2021 IEEE 24th International Conference on Intelligent Transportation Systems (ITSC).

Introduction

Transportation electrification is seen as one of the main paths for emissions reduction. Electric vehicles (EV) have been receiving increasing attention in recent years. With the increase in performance and variety of EV models in the market, it becomes essential to study how to efficiently incorporate them into our society.

BETs are expected to be more efficient than diesel trucks and have lower operational costs (Sripad and Viswanathan 2019). Studies regarding the viability of battery electric trucks (BETs)

show promising results; however, concerns regarding reduced payload, limited range, long recharge time, and the required supporting infrastructure were raised (Mareev, Becker, and Sauer 2018; Earl et al. 2018; Çabukoglu et al. 2018; Liimatainen, van Vliet, and Aplyn 2019; Sripad and Viswanathan 2019; Smith et al. 2019). Current powertrain efficiency and battery density make it so large batteries are required for long-range trips. As trucks are subject to weight constraints, the battery weight reduces max payload. Moreover, large capacity batteries take longer to charge.

When accounting for the fact that drivers need to rest due to regulations, the recharge time issue might not be as pronounced. Drivers are required to rest regularly due to hours-of-service (HOS) regulations. If drivers can use the mandatory rest stops to recharge, the increase in trip duration due to recharge times can be reduced or eliminated. The synchronization of rest and recharge times was studied for single-day trips by Schiffer et al. (Schiffer et al. 2017), showing improvements for BETs. A study by Mareev et al. (Mareev, Becker, and Sauer 2018) used EU regulations to help estimate costs for long-haul BETs. Mareev used the EU regulation to generate a regulation-compliant baseline driving cycle. However, in practice, the trip's schedule depends on both clients and charging stations; thus, it can differ significantly from the regulation's minimum requirements.

Studies also overlooked the problem of parking availability. Due to limited infrastructure, BETs may face difficulties finding available recharging stations. However, diesel trucks are not entirely free of such worries. Currently, the large number of trucks already causes truck parking shortages, forcing drivers to adjust their schedules to find appropriate parking or face the risks of illegal parking (U.S. Department of Transportation 2015). In (Vital and Ioannou 2019; 2020), Vital and Ioannou studied the problem of including both parking availability information and HOS regulations in the planning of long-haul transportation, but those studies do not cover electric vehicles.

Planning for freight transport electrification requires a good understanding of BET operations' performance and cost under realistic scenarios and how each scenario affects BET usage's minimum requirements. These scenarios include the need for drivers to adapt their schedules to fulfill client restrictions, HOS regulations, parking availability, and charging needs. In this project, we study how the performance gap between BETs and diesel trucks is affected by practical constraints such as HOS regulations and limited parking availability. The paper is organized as follows: Section USA's Hours of Service Regulations describes the HOS regulations considered. Section Consumption Models describes the models used to estimate energy/fuel consumption and emissions. Section Problem Description describes the mathematical model used for BET trip planning under HOS and parking constraints. Sections Case Study and Experiments on Random Networks describe the experiments performed and their results. Section Conclusion presents the conclusion.

USA's Hours of Service Regulations

The USA HOS regulation restricts for how long drivers can drive/work, and how long they should rest before being allowed to drive again. We refer to the off-duty periods required by

the regulation based on their minimum duration: *breaks* (0.5 h), *daily rests* (10 h) and *weekly rests* (34 h). The USA HOS regulation can be summarized as follows (Federal Motor Carrier Safety Administration 2021):

- 11-hour Driving Time Limit: A driver may drive at most 11 hours between 2 consecutive *daily rests*.
- 14-Hour Elapsed Time Limit: A driver cannot drive after 14 hours have elapsed since the last *daily rest* ended.
- Rest Breaks: A driver must take a *break* after 8 cumulative hours of driving time. Recent changes in the regulation allow this constraint to be satisfied by any non-driving period of 30 consecutive minutes.
- 60-Hour Limit: A driver cannot drive after having been on duty for 60 hours in any period of 7 consecutive days. The 7 days period can be reset by taking a *weekly rest*.

We do not consider the sleeper berth provision, which allows *daily rests* to be split. And, for the 60-hour limit, instead of restricting the on-duty time over any period of 7 consecutive days, the on-duty time between two consecutive *weekly rests* was restricted to 60 hours.

Consumption Models

The energy/fuel consumption depends on the activity being considered, so we separate the model in three cases: driving, idling and charging (BET only). The model parameters considered are listed in Table 28.

Driving

The consumption models used are based on models found in the literature ((Bektaş and Laporte 2011; J. Lin, Zhou, and Wolfson 2016; Earl et al. 2018; Sripad and Viswanathan 2017; Gao, Lin, and Franzese 2017) for BETs, and (Wang and Rakha 2017) for diesel trucks). Both models first estimate the vehicle's power demand due to resistance forces acting on the vehicle, then estimate the consumption rate based on the power demand. We consider the average travel speed over each road section, and terms relative to acceleration and road grade were omitted.

BETs

Let $P_B(v)$ be the power demand (in kW) to the battery due to the forces acting against the truck's movement, accounting for the battery to wheel efficiency, and $\zeta_B(v)$ be the rate of energy consumption per unit of distance traveled (in kWh/km) defined as follows:

$$P_B(v) = \left(\frac{\rho A C_D}{25.92} v^2 + mgC_R\right)\frac{v}{3600\eta_{bw}} \tag{129}$$

$$\zeta_B(v) = \frac{P_B(v) + P_a}{v} \tag{130}$$

where C_D and C_R are the coefficients of drag and rolling resistance, respectively. The air density (kg/m^3) is given by ρ, and the acceleration due to gravity is given by g. The terms v, m and A represent the truck's speed (km/h), mass (kg) and frontal area (m^2), respectively. η_{bw} represents the battery-to-wheels efficiency, and P_a is the power demand from the vehicle's accessories and support systems, e.g., A/C, lighting, electric steering system. The battery level cannot be negative, so any displacement requiring more energy than currently stored in the battery is considered infeasible.

Diesel Trucks

For diesel trucks, we used the model presented by Wang and Rakha in (Wang and Rakha 2017). More specifically, the parameters used are the ones for a convex model of a Freightliner/FLD 120, year 2001, labeled as "HDDT8" in their paper. This model characterizes fuel consumption as a second-order polynomial function of the power demand, as follows:

$$P_D(v) = \left(\frac{\rho A C_D}{25.92} v^2 + mgC_R(c_1 v + c_2)\right)\frac{v}{3600\eta_d} \tag{131}$$

$$\zeta_D(v) = (\alpha_0 + \alpha_1 P_D(v) + \alpha_2 P_D(v)^2)\frac{3600}{v} \tag{132}$$

where, similarly to (129) and (130), $P_D(v)$ represents the power demand (kW), and $\zeta_D(v)$ represents the fuel consumption per distance (L/km). C_R, c_1 and c_2 are the rolling resistance parameters (unitless), η_d is the driveline efficiency (unitless), α_0, α_1 and α_2 are vehicle-specific model coefficients calibrated in (Wang and Rakha 2017) using empirical data. The remaining parameters are defined as in (129).

Idling

When 'idling', we consider a fixed consumption rate P_I (kW) for BETs and F_I (L/h) for diesel trucks. We assume that P_I is smaller than P_a as some systems, such as electric steering, might be inactive when the vehicle is stopped. For BETs, we assume that chargers can provide power to these support systems on top of charging needs and idling consumption is not subtracted from the charging rate at charging stations. However, when estimating emissions, idling consumption is included in the energy expenditure.

Charging

BETs can recharge their batteries at charging stations located on the road network (and possibly clients with charging infrastructure). We consider a finite battery capacity B (in kWh), and, for each charging station ℓ, a constant charging rate γ_ℓ (in kW). The battery cannot store more energy than its capacity, so we assume that the battery will stop charging when full.

Table 28. Model Parameters

Battery Electric Truck		
Parameter	**Description**	**Value**
C_D (Sripad and Viswanathan 2017)	coefficient of drag	0.63
C_R (Sripad and Viswanathan 2017)	coefficient of rolling resistance	6.3E-3
η_{bw} (Sripad and Viswanathan 2017)	battery-to-wheels efficiency	0.85
m **(kg)** (Sripad and Viswanathan 2017)	truck's total mass	3.6E4
P_a **(kW)** (Smith et al. 2019)	support systems power demand	10
P_I **(kW)**	idling power demand	3
A **(**m^2**)** (Sripad and Viswanathan 2017)	truck's frontal area	7.2
Diesel Truck		
C_D (Wang and Rakha 2017)	coefficient of drag	0.78
C_R (Wang and Rakha 2017)	coefficient of rolling resistance	1.25E-3
c_1 (Wang and Rakha 2017)	coefficient of rolling resistance	0.0328
c_2 (Wang and Rakha 2017)	coefficient of rolling resistance	4.575
η_d (Wang and Rakha 2017)	driveline efficiency	0.94
m **(**kg**)** (Wang and Rakha 2017)	truck's total mass	3.6E4
A **(**m^2**)** (Wang and Rakha 2017)	truck's frontal area	10
α_0 (Wang and Rakha 2017)	vehicle-specific model coefficient	2.16E-3
α_1 (Wang and Rakha 2017)	vehicle-specific model coefficient	7.98E-5
α_2 (Wang and Rakha 2017)	vehicle-specific model coefficient	1.0E-8
F_I **(**L/h**)** (U.S. Department of Energy 2015)	idling fuel consumption	3
General		
$g(m/s^2)$ (Wang and Rakha 2017)	gravity	9.8066
$\rho(kg/m^3)$ (Sripad and Viswanathan 2017)	air density	1.2256
$\beta_d(kg/L)$ (U.S. Energy Information Administration 2016; Argonne National Laboratory 2020)	CO_2 emission factor for diesel	3.13
$\beta_e(kg/kWh)$ (Argonne National Laboratory 2020)	CO_2 emission factor for electricity in California	0.2

Problem Description

The problem consists of planning the path, travel speed and duration of rest and recharge stops of a single battery-electric truck from origin to destination with required stops at an ordered set of client locations. The solution must comply with the USA HOS regulations, and satisfy battery level, delivery time and parking availability constraints. This problem is a variant of the SPTDSP-

PA (shortest path and truck driver scheduling problem with parking availability constraints), which was introduced in (Vital and Ioannou 2020). This variant differs mainly in the inclusion of factors relevant to EV planning, i.e., finite battery capacity, partial recharges, speed-dependent energy consumption, and speed control.

The problem is solved over a simplified road network that includes only the main routes the truck can take between two consecutive client locations, and the rest areas and charging stations around them. The simplified road network is defined as an acyclic directed graph $G = (V, A)$, where V is the set of nodes of the graph and A is the set of edges. Each road section $(i, j) \in A$ has a fixed length d_{ij} and an allowed speed range $\left[s_{ij}^{-}, s_{ij}^{+}\right]$, thus setting the allowed travel time to $\left[\frac{d_{ij}}{s_{ij}^{+}}, \frac{d_{ij}}{s_{ij}^{-}}\right]$. The speed limit is considered constant within each road section, but the average travel speed can be adjusted within the allowed range to control the travel time and energy consumption. The vehicle has a finite battery capacity and the energy models used to calculate the energy consumption and recharge rates are described in the section Consumption Models. The battery can only be recharged at charging stations, which can be either rest areas or client facilities with charging infrastructure.

During long trips, HOS regulations require drivers to rest along the way. Rest stops are restricted to rest areas and their minimum durations are defined by the regulation. We do not allow for rests to be taken at client locations. However, note that service times longer than 30min can reset the 8h driving limit constraint despite counting as on-duty time for other constraints. Each parking location has a set of time-windows representing the intervals when parking spaces are expected to be available. These time-windows restrict the vehicle's arrival time. The vehicle is not allowed to arrive early and wait. The regulation sets a minimum duration for the rest stops, but it does not set a maximum duration, so the driver is allowed to extend the stay when convenient. Similarly, each client has a set of time-windows constraints and a service time, which define when the truck can arrive at the client and the duration of stay. However, drivers cannot extend the service time at the client. As rest areas are not required stops, the graph G is built so that rest areas can be bypassed. Clients are mandatory stops, so all considered routes go through the client nodes.

Our model is based on the resource constrained shortest path problem formulation presented by Vital and Ioannou for the SPTDSP-PA (Vital and Ioannou 2020). Time and counters for the different HOS regulations are treated as resources, and an extended network is used to explicitly represent drivers' possible activities. We use the extended network proposed in (Vital and Ioannou 2020), which is described in the section Extended Network. The extended network edges also store charging parameters. Subsection System Equations describes how the system equations were updated to include battery level as a resource. Section Dynamic Programming Formulation and Rollout Algorithm describes the rollout algorithm used to approximately solve the problem.

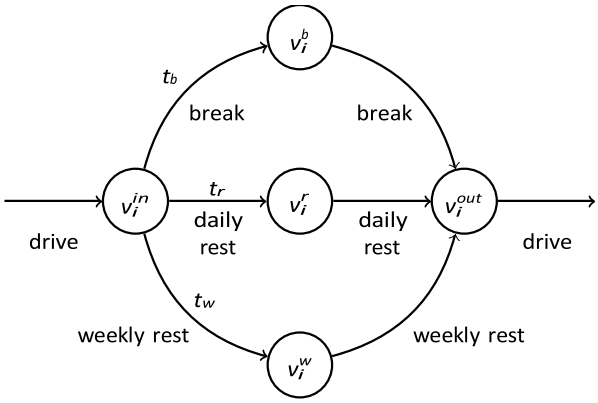
Extended Network

At the origin, client locations and rest areas/charging stations, drivers perform non-driving activities. These nodes are expanded according to the subnetworks in Figure 126, forming an extended network $G' = (V', A')$ that includes non-driving activities explicitly in the graph. Figure 126 shows the subnetworks that replace the nodes being expanded. Each edge has its activity indicated below the arrow. Edges that have a fixed duration have their duration indicated above the arrow. The incoming/outgoing edges of the subnetwork are the incoming/outgoing edges of the node being expanded.

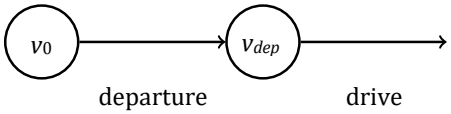
Rest Areas/Charging Stations: HOS regulations are affected differently by rests of different durations, so each branch of this sub-network models rests with similar effects. They are *break*, *daily rest* and *weekly rest*, which have minimum durations of t_b, t_r and t_w, respectively. Figure 126a shows the sub-network representing the node after expansion. The duration of the second half of each path is a decision variable used to model rest time beyond the minimum required. Each edge also stores information regarding charging capabilities, i.e., whether it is a charging station or not, and the charging rate.

Origin: This expansion, shown in Figure 126b, is used to model the vehicle's departure time. Departure time constraints are modeled as time-window constraints at the departure node v_{dep}.

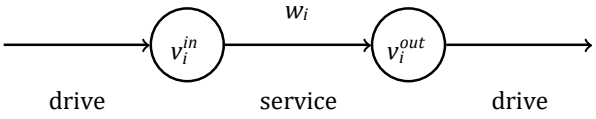
Clients: This expansion is used to model the service time at each client, and is shown in Figure 126c. A priori we consider that vehicles cannot recharge at client locations and that the idling consumption rate is the same for all client locations, but charging and energy consumption information can also be stored in the service edges to customize behavior at each client.



(a) Sub-network used to expand rest area nodes.



(b) Sub-network used to expand the origin node.



(c) Sub-network used to expand client nodes. w_i is the service time.

Figure 126. Sub-networks used to model non-driving activities.

System Equations

We consider the system's state as being a vector $x_k = (v_k, \theta_k)$, where $\theta_k = (\eta_k^0, \eta_k^b, \eta_k^r, \psi_k^r, \psi_k^w, b_k)$ containing the current location of the truck (v_k) and the current resource values (θ_k). The resources are responsible for tracking the HOS restrictions, battery level and arrival time at each node. The resources used are:

- Time when node was visited
- Accumulated driving time since last *break*
- Elapsed time since last *daily rest*
- Accumulated driving time since last *daily rest*
- Accumulated on-duty time since last *weekly rest*
- Battery level

The evolution of the system is described by $x_{k+1} = f(x_k, u_k)$, where x_k is the current state, x_{k+1} is the next state, and u_k is the decision taken. The decision u_k is composed by an edge $e_k = (v_k, v_{k+1}) \in A'$, with length μ_k, and a duration δ_k included in e_k's allowed duration set. When dealing with edges related to driving, this set is defined by the length and allowed speed values of e_k. The function $f(x_k, u_k)$ defines how each element of x_k is affected by a decision u_k. As different activities have different impacts on each resource, each edge of the extended network has an activity assigned to it. Table 29 shows how the resources are updated

depending on the activity. The functions f^d, f^s, f^b, f^r, f^w and f^0 describe the update rules for activities *drive*, *service*, *break*, *daily rest*, *weekly rest* and *departure*, respectively. Figure 126 shows how the activities are assigned to each edge. Note that η^b's and f^b's definitions differ from (Vital and Ioannou 2020) due to recent changes in the regulation. Now the 8h limit is applied to driving time instead of elapsed time, and any non-driving period longer than 30 minutes can satisfy this constraint.

Table 29. Resource Extension Functions

	f^d	f^s	f^b	f^r	f^w	f^0
$\eta^0_{k+1} =$	$\eta^0_k + \delta_k$					
$c_{k+1} - c_k =$	$\alpha_d\delta_k + \beta_d\mu_k$	$\alpha_s\delta_k$	$\alpha_b\delta_k$	$\alpha_r\delta_k$	$\alpha_w\delta_k$	$\alpha_0\delta_k$
$\eta^b_{k+1} =$	$\eta^b_k + \delta_k$		0			η^b_k
$\eta^r_{k+1} =$	$\eta^r_k + \delta_k$			0		η^r_k
$\psi^r_{k+1} =$	$\psi^r_k + \delta_k$	ψ^r_k		0		ψ^r_k
$\psi^w_{k+1} =$	$\psi^w_k + \delta_k$		ψ^w_k		0	ψ^w_k
$b_{k+1} =$	$b_k - \mu_k\zeta_B\left(\frac{\mu_k}{\delta_k}\right)$	$\begin{cases} max(B, b_k + \delta_k\gamma), & if\ charging\ b_k \\ -\delta kPI, & o/w \end{cases}$				b_k

Dynamic Programming Formulation and Rollout Algorithm

Let $J(x_k)$ be the minimum cost to go from state x_k to the destination, and X_d the set of feasible states at the destination node. This cost-to-go function is defined as:

$$J(x_k) = \begin{cases} 0, \text{ if } x_k \in X_d \\ \min\limits_{u \in U(x_k)} g(x_k, u) + J\big(f(x_k, u)\big), \text{o.w.} \end{cases} \quad (133)$$

where $g(x_k, u)$ is the cost accrued by decision u at state x_k, and $U(x_k)$ is the set of decisions u for which $f(x_k, u)$ is a feasible state. A state is considered feasible if all resources are within their respective feasible ranges. If $U(x_k)$ is empty, we say that the destination cannot be reached from x_k and $J(x_k)$ is infinite. The choice of $g(\cdot)$ determines what is being minimized. In this project, we use the trip duration as the objective to be minimized, but more complex objective functions can be used. For example, we can include the cost of energy and perform a multi-objective optimization that minimizes a weighted sum of the cost generated by the trip duration (driver's hourly salary) and costs generated by energy expenditure.

Although any node has only a finite number of outgoing edges, the decision space $U(x_k)$ can have uncountably many elements if the allowed duration set of one or more of these edges is a continuous interval. In order to mitigate this issue, we first propagate the constraints of each

node to all upstream nodes. This reduces the feasible space at each node and the decision space to be considered for each decision. During execution, the algorithm uses the preprocessed feasible ranges to generate a reduced decision space, which is then discretized, generating a finite set of decisions. Nevertheless, due to the curse of dimensionality, this approach does not scale well for large instances. Using a coarse decision space discretization can bring significant improvements to computation time, but will also cause the cost to deteriorate. Therefore, we use a rollout algorithm (D. Bertsekas 2017) to find suboptimal solutions while keeping the computational demand in check. The general idea is to use the cost obtained from applying a base policy as an approximate cost function, then use this approximation to generate a one-step lookahead policy. One-step lookahead policies choose the decision that minimizes the following expression:

$$\min_{u_k \in U(x_k)} g(x_k, u_k) + \tilde{J}\big(f(x_k, u_k)\big) \tag{134}$$

where $\tilde{J}(x_k)$ is the approximated cost-to-go of state x_k. Let the policy π be a function that returns a feasible decision $\pi(x_k) \in U(x_k)$ for every state x_k. $J_\pi(x_k)$ is the cost-to-go when the policy π is used to take decisions at every state, and it can be described as:

$$J_\pi(x_k) = \begin{cases} 0, \text{ if } x_k \in X_d \\ g\big(x_k, \pi(x_k)\big) + J_\pi\Big(f\big(x_k, \pi(x_k)\big)\Big), \quad \text{o.w.} \end{cases} \tag{135}$$

In this project, we used $\tilde{J}(x) = J_\pi(x)$, where π is the policy generated by solving the problem with a coarser discretization of the decision space. The strategy used to propagate constraints is included in the section Constraint Propagation and Feasible Decision Space. Section Graph Preprocessing describes how the graphs were preprocessed to reduce issues with short links. Section Analytical Solutions and Section Cost Lower Bound show, respectively, analytical solutions and cost lower bounds that can be used to speed-up the algorithm. The cost lower bounds presented include cost functions that are a weighted sum of time and energy consumption, which are more general then the cost functions considered in our experiments.

Constraint Propagation and Feasible Decision Space

Consider the following expression describes how the states are updated:

$$x_{i+1} = f(x_i, u_i),\ u_i \in U_i(x_i) \subset U_i \tag{136}$$

Let F_i represent the set of feasible states at node v_i. We define $U_i(x_i)$ as:

$$U_i(x_i) = \{u \in U_i |\ f(x_i, u) \in F_{i+1}\} \tag{137}$$

When choosing the decisions to test, we can either sample U_i and check the feasibility of each decision or calculate the feasible decision space with an inverse function $f^{-1}(F_{i+1}, x_i)$ that returns the elements of U_i that can generate a next state in F_{i+1}. As most edges update the resources by adding its duration to the current resource, in general this operation consists of shifting the intervals representing the constraint for each resource, then taking the intersection

between all of them, e.g., if the next node has a time-window [10,15] and the current time is 5, then the decision duration must be in the interval [5,10] to be feasible. Different resources will generate different intervals, and feasible decisions must satisfy all of them.

Originally, F_i represents only the feasibility regarding the local constraints at node v_i, however, if we consider constraints from other nodes, we may be able to reduce F_i, and consequently reduce $U_i(x_i)$. Each node's local constraints can be propagated downstream and upstream to reduce other nodes' feasible spaces.

Forward Propagation

Let $\mathcal{F}^*\left(F_i,\ F_j, U_i(\cdot)\right)$ represent a function that returns which states in F_j can be reached from F_i, i.e.,

$$\mathcal{F}^*\left(F_i,\ F_j, U_i(\cdot)\right) = \left\{x_j \in F_j | \exists x_i \in F_i,\ \exists u \in U_i(x_i),\ f(x_i, u) = x_j\right\} \tag{138}$$

The set $\mathcal{R}_j^*$ of states that can be reached at node v_j is given by:

$$\mathcal{R}_j^* = \bigcup_{i,\,(v_i,v_j)\in A} \mathcal{F}^*\left(F_i, F_j, U_i(\cdot)\right) \tag{139}$$

$\mathcal{R}_j^*$ can be overly complex due to the coupling between resources, so we try to approximate it by propagating the constraints for each resource separately. Let $F_i^{(r)}$ be the projection of F_i on the axis representing resource r, and $f^{(r)}$ the component of f that defines the evolution of resource r. Let $\mathcal{F}^{(r)}\left(F_i,\ F_j, U_i\right)$ be a function that returns which values of resource r can be reached at node v_j, defined as follows:

$$\mathcal{F}^{(r)}\left(F_i,\ F_j, U_i\right) = \left\{x_j^{(r)} \in F_j^{(r)} | \exists x_i^{(r)} \in F_i^{(r)},\ \exists u \in U_i,\ f^{(r)}\left(x_i^{(r)}, u\right) = x_j^{(r)}\right\} \tag{140}$$

Let $\hat{\mathcal{R}}_j$ approximate $\mathcal{R}_j^*$ as follows:

$$\begin{aligned} \mathcal{R}_{i,j} &= \prod_r \left(\mathcal{F}^{(r)}\left(\hat{\mathcal{R}}_i,\ F_j, U_i\right)\right) \\ \hat{\mathcal{R}}_j &= \prod_r \bigcup_{j,\,(v_i,v_j)\in A} \mathcal{R}_{i,j}^{(r)} \end{aligned} \tag{141}$$

Where $\mathcal{R}_{i,j}$ is approximation accounting only for the constraints of upstream node v_i, and ${\mathcal{R}_{i,j}}^{(r)}$ is its projection on the axis representing resource r. Note that, for a given v_i , if $\exists r$ such that $\mathcal{F}^{(r)}\left(\hat{\mathcal{R}}_i, F_j, U_i\right) = \emptyset$, then $\mathcal{R}_{i,j} = \emptyset$. That is, if states from v_i cannot satisfy the constraints for 1 or more resources, then v_i will not be counted when calculating the reachable states at v_j. Furthermore, the edge $\left(v_i, v_j\right)$ can be removed from the problem. At the origin node we have that $\mathcal{R}_0^* = \hat{\mathcal{R}}_0$. If the initial state is known, it is the only reachable state at the origin, otherwise $\mathcal{R}_0^*$ is the set of possible initial states.

Backward Propagation

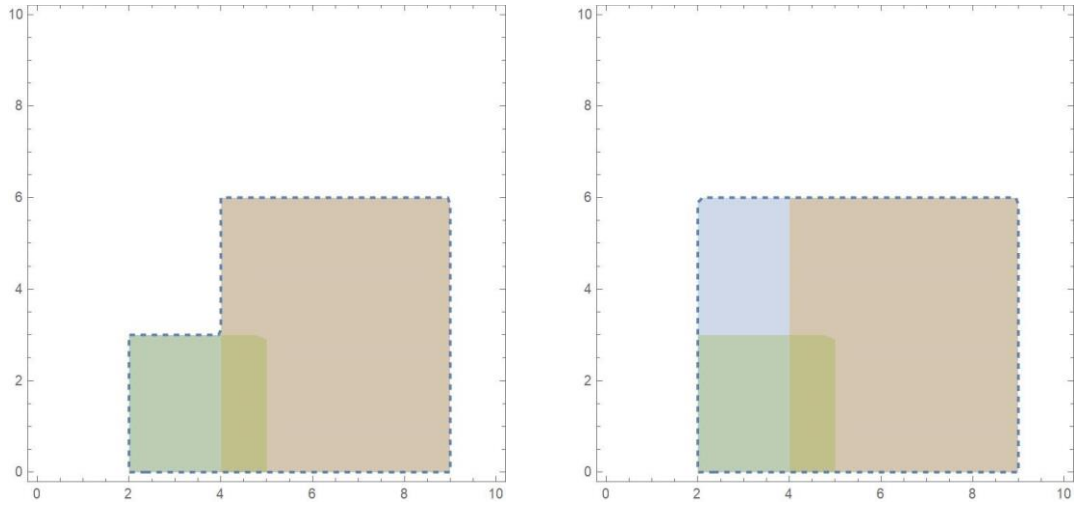
Backward propagation follows the same general idea as forward propagation. Let $\bar{F}_i$ be the reduced feasible space. Like how we calculated $U_i(x_i)$, we need a function $\mathcal{B}\left(F_i,\ F_j, U_i(\cdot)\right)$ that can calculate the values of x_i at node v_i that can lead to at least one feasible state x_j at one of the successors v_j , i.e,

$$\mathcal{B}\left(F_i, F_j, U_i(\cdot)\right) = \left\{x_i \in F_i | \exists u \in U_i(x_i),\ f(x_i, u) \in F_j\right\} \tag{142}$$

However, this function is hard to compute and generates complex regions that will require more space to store, and more time to check during execution. Therefore, we calculate separate regions for each resource and use it to generate an approximate feasible state space $\hat{F}_i$ as follows:

$$\begin{aligned} \mathcal{C}^{(r)}\left(F_i,\ F_j, U_i\right) &= \left\{x_i^{(r)} \in F_i^{(r)} | \exists u \in U_i,\ f^{(r)}\left(x_i^{(r)}, u\right) \in F_j^{(r)}\right\} \\ \hat{F}_{i,j} &= \prod_r \left(\mathcal{C}^{(r)}\left(F_i,\ \hat{F}_j, U_i\right)\right) \\ \hat{F}_i &= \prod_r \bigcup_{j,\,(v_i,v_j)\in A} F_{i,j}^{(r)} \end{aligned} \tag{143}$$

As in $\mathcal{R}_{i,j}$, $\hat{F}_{i,j}$ is the empty set if any resource constraint cannot be satisfied. In this case, the edge $\left(v_i, v_j\right)$ can be removed from the graph as it cannot generate feasible states. We do the forward propagation before the backward, so, at the destination node v_n, we have that $\bar{F}_n = \hat{F}_n = \hat{\mathcal{R}}_n$. For example, if a node v_j has a time-window $[10,15]$ and the edge (i,j) can have a duration in the interval $[2,5]$, then v_i must be visited in the time-window $[10-5, 15-2] = [5,13]$. If a different edge (v_i, v_k) generated a propagated time-window of $[7,17]$ on v_i, we would consider the union of both time-windows, i.e., $[5,17]$. Then we would take the intersection ofv_i's original time-window, say $[0,15]$, and the time-windows obtained from propagating downstream constraints to obtain an estimated feasible time-window of $[5,17]$. Note that the interval $[5,15]$ can be divided into an interval feasible for paths passing through v_i, $[5,13]$, and one feasible for paths through v_k, $[7,15]$. The same can happen to other resource constraints. Therefore, it is possible that a state in $\hat{F}_i$ only satisfies the time-window for a certain path but satisfies the HOS resource constraints only for a different path. As all constraints are satisfied by some path, the state is included in $\hat{F}_i$, but, in practice, that state cannot generate feasible successors. So, we have that $\hat{F}_i$ might contain states that cannot satisfy downstream solutions, i.e., $\bar{F}_i \subseteq \hat{F}_i \subseteq F_i$. Figure 127 shows a 2D example of the difference between reduced feasible state space $\bar{F}_i$ and its approximation $\hat{F}_i$. The blue region in Figure 127b belongs to $\hat{F}_i$, but not to $\bar{F}_i$.



(a) Blue dashed line: correct feasible space. (b) Blue dashed line: approximate feasible space. Blue region: infeasible.

Figure 127. The green and brown regions are examples of possible feasible regions in a 2D space. The figures show how the exact (a) and approximate (b) feasible spaces are calculated.

Propagating the resources

We separate the resource extension functions according to how they affect the resource being updated. The resource extension functions either add a value to the resource (ADD), maintain the current resource value (NoEff), or set the resource value to 0 (RESET). Let $e = (v_i, v_j)$ be an edge, $[\delta_e^-, \delta_e^+]$ be edge e's possible durations defined in U_i. Let $[\eta_i^-, \eta_i^+]$ be the feasible values for resource r at node v_i. The approximate propagation functions described previously are defined as follows for the 3 types of REF:

Forward Propagation

ADD: $\mathcal{F}^{(r)}(F_i,\ F_j, U_i) = [\eta_i^- + \delta_e^-, \eta_i^+ + \delta_e^+] \cap [\eta_j^-, \eta_j^+]$

NoEff: $\mathcal{F}^{(r)}(F_i,\ F_j, U_i) = [\eta_i^-, \eta_i^+] \cap [\eta_j^-, \eta_j^+]$

RESET: $\mathcal{F}^{(r)}(F_i,\ F_j, U_i) = \{0\} \cap [\eta_j^-, \eta_j^+]$

Backward Propagation

ADD: $\mathcal{C}^{(r)}(F_i,\ F_j, U_i) = [\eta_j^- - \delta_e^+, \eta_j^+ - \delta_e^-] \cap [\eta_i^-, \eta_i^+]$

NoEff: $\mathcal{C}^{(r)}(F_i,\ F_j, U_i) = [\eta_i^-, \eta_i^+] \cap [\eta_j^-, \eta_j^+]$

RESET: $\mathcal{C}^{(r)}(F_i,\ F_j, U_i) = \{0\} \cap [\eta_j^-, \eta_j^+]$

When a resource's feasible range is a set of disjoint intervals, the functions above can be applied to each interval separately and we take the union of the resulting sets. Note that, in these REFs, the decision's duration is directly used to update the resource values. When energy/fuel consumption is included as a resource, the update value will be a function of the duration, so the propagation function will depend on the consumption model used.

Reduced Decision Space

The reduced decision space is generated following the same idea.

$$\bar{\mathrm{U}}_{\mathrm{i,j}}\left(\mathrm{x_i}, \mathrm{U_i}(\cdot), \hat{\mathrm{F}}_{\mathrm{j}}\right) = \left\{\mathrm{u} \in \mathrm{U_i}(\mathrm{x_i}) \middle| \mathrm{f}(\mathrm{x_i}, \mathrm{u}) \in \hat{\mathrm{F}}_{\mathrm{j}}\right\} = \bigcap_{\mathrm{r}} \left\{\mathrm{u} \in \mathrm{U_i}(\mathrm{x_i}) \middle| \mathrm{f}^{(\mathrm{r})}(\mathrm{x_i}, \mathrm{u}) \in \hat{\mathrm{F}}_{\mathrm{j}}^{(\mathrm{r})}\right\} \quad (144)$$

$$\bar{\mathrm{U}}_{\mathrm{i}}\mathrm{x_i}, \mathrm{U_i}(\cdot)\left(, \left\{\hat{\mathrm{F}}_{\mathrm{j}}\right\}\right) = \bigcup_{\mathrm{j},\, (\mathrm{v_i}, \mathrm{v_j}) \in \mathrm{A}'} \bar{\mathrm{U}}_{\mathrm{i,j}}\left(\mathrm{x_i}, \mathrm{U_i}(\cdot), \hat{\mathrm{F}}_{\mathrm{j}}\right) \quad (145)$$

Let η_i be the current value of resource r. The other symbols are defined as in the previous section.

ADD: $\bar{U}_{i,j}^{(r)}\left(x_i, U_i(\cdot), \hat{F}_j\right) = \left\{u \in U_i(x_i) | f^{(r)}(x_i, u) \in \hat{F}_j^{(r)}\right\} = \left[\eta_j^- - \eta_i, \eta_j^+ - \eta_i\right] \cap [\delta_e^-, \delta_e^+]$

NoEff: $\bar{U}_{i,j}^{(r)}\left(x_i, U_i(\cdot), \hat{F}_j\right) = \begin{cases} \emptyset, if\ \eta_i \notin \hat{F}_j^{(r)} \\ [\delta_e^-, \delta_e^+],\ o.w. \end{cases}$

RESET: $\bar{U}_{i,j}^{(r)}\left(x_i, U_i(\cdot), \hat{F}_j\right) = \begin{cases} \emptyset, if\ 0 \notin \hat{F}_j^{(r)} \\ [\delta_e^-, \delta_e^+],\ o.w. \end{cases}$

Analytical Solutions

At nodes where the only possible next stop is the destination it is possible to analytically define the best decision so that the algorithm does not need to search over the remainder of that search tree branch. Naturally, the decision depends on the cost function and constraints being considered in the problem. Consider the following cost function for a decision of duration δ:

$$C(\delta) = \begin{cases} \alpha\delta + \beta\mu_e\zeta(\mu_e/\delta), & if\ driving \\ (\alpha + \beta\gamma + \theta)\delta, & o.w. \end{cases} \quad (146)$$

, where α is the trucks hourly operational cost (excluding fuel/energy) and β is the cost per unit of fuel/energy. For non-driving decisions, γ is the hourly idling fuel/energy consumption, θ represents hourly costs incurred while stopped from sources other than idle energy consumption and operational costs. For driving decisions, μ_e is the length of the road segment considered, and $\zeta(v)$ is the fuel/energy consumption per unit of distance. This cost function considers both time and energy/fuel related costs, and their relative importance can be adjusted using the parameters α, β, and θ. In this section, we study the optimal decisions for

the last driving and rest extension decisions. Although we focus on BETs, the solutions for diesel trucks can be obtained by ignoring the battery constraint.

Last driving decision

$$\frac{dC}{d\delta} = \alpha + \beta\mu_e \frac{d\zeta(v)}{d\delta} = \alpha + \beta\mu_e \frac{d\zeta(v)}{dv}\frac{dv}{d\delta} = \alpha - \beta\mu_e \frac{d\zeta(v)}{dv}\frac{\mu_e}{\delta^2} \\ = \alpha - \beta v^2 \frac{d\zeta(v)}{dv} = 0 \tag{147}$$

Cost is minimum for $\tilde{\delta} = \frac{\mu_e}{\tilde{v}}$, such that $\tilde{v}$ is the root of $v^2 \frac{d\zeta(v)}{dv} = \frac{\alpha}{\beta}$. Assuming that $\zeta(v)$ is a convex function, and, consequently, $\frac{d\zeta}{dv}$ is monotonically non-decreasing, we can say that $v^2 \frac{d\zeta(v)}{dv}$ is strictly increasing over $(\max(0, v'), \infty)$, where v' satifies $\frac{d\zeta(v')}{dv} = 0$. As α and β are positive, $\tilde{v}$ is unique. The function $v^2 \frac{d\zeta(v)}{dv}$ does not depend on the edge, so $\tilde{v}$ can be calculated beforehand. Let $\left[\underline{\delta}, \overline{\delta}\right]$ be δ's domain, the optimal decision is given by:

$$\delta = \begin{cases} \underline{\delta}, & if \quad \tilde{\delta} < \underline{\delta} \\ \overline{\delta}, & if \quad \tilde{\delta} > \overline{\delta} \\ \tilde{\delta}, & o.w. \end{cases} \tag{148}$$

Last rest extension

Let ρ be the recharge rate at the current location, $\delta_0 \in \left[\underline{\delta_0}, \overline{\delta_0}\right]$ the rest extension to be chosen, and $\delta_\ell \in \left[\underline{\delta_\ell}, \overline{\delta_\ell}\right]$ the duration of the decision at the following edge, which is the last driving edge. The cost from the rest node to the destination can be written as $C(\delta_0, \delta_\ell) = (\alpha + \beta\gamma + \theta)\delta_0 + \alpha\delta_\ell + \beta\mu_e\zeta(\mu_e/\delta_\ell)$. Assume that, due to the destination node's resource constraints and the current state's resource values, $\delta_0 + \delta_\ell \in \left[\underline{D}, \overline{D}\right]$. The optimization problem being solved at the last rest decision can be described as:

$$\min_{\delta_0, \delta_l} C(\delta_0, \delta_l) = (\alpha + \beta\gamma + \theta)\delta_0 + \alpha\delta_l + \beta\mu_e\zeta(\mu_e/\delta_l) \tag{149}$$

s.t.:

$$\mu_e\zeta(\mu_e/\delta_l) - \delta_0\rho - B_0 \leq 0 \tag{150}$$

$$\underline{D} \leq \delta_0 + \delta_l \leq \overline{D} \tag{151}$$

$$\underline{\delta_0} \leq \delta_0 \leq \overline{\delta_0} \tag{152}$$

$$\underline{\delta_l} \leq \delta_l \leq \overline{\delta_l} \tag{153}$$

, where (150) guarantees that the battery charge is non-negative when arriving at the destination. (151) restricts the time to reach the destination, and can be related to both HOS

and time-window constraints. (152) and (153) restrict the domains of δ_0 and δ_ℓ to the reduced decision space, which is affected by all constraints and the current state. Consider the following definitions:

$$
\begin{aligned}
H(v) = &\quad v^2 \frac{d\zeta(v)}{dv} \\
P(v) = &\quad \frac{\mu_e \zeta(v) - B_0}{\rho} \\
\nabla C = &\quad \left[(\alpha + \beta\gamma + \theta) \quad \left(\alpha - \beta H(v)\right)\right] \\
\nabla g_1 = &\quad \left[(-\rho) \quad \left(-H(v)\right)\right] \\
\tilde{v}, &\quad H(\tilde{v}) = \frac{\alpha}{\beta} \\
\hat{v}, &\quad H(\hat{v}) = -\frac{\theta}{\beta} - \gamma \\
\check{v}, &\quad H(\check{v}) = \rho \\
\breve{v}, &\quad H(\breve{v}) = 0 \\
v^*, &\quad H(v^*) = \frac{\alpha\rho}{\alpha + \beta(\rho + \gamma) + \theta}
\end{aligned}
$$

, where g_1 represents constraint (150). $H(v)$ and $P(v)$ are auxiliary functions defined to simplify the notation and represent, respectively, the derivative of the energy consumption with respect to δ_ℓ and the minimum feasible δ_0 given δ_ℓ. The v's with different accents are values used in the solution that can be calculated offline. $\tilde{v}$, $\hat{v}$, and v^* represent, respectively, the speeds at which the cost gradient ∇C is perpendicular to (152), (151), and (150). $\check{v}$ and $\breve{v}$ are the speeds at which (150) is parallel to (151) and (152), respectively. Note that, given a distance μ_e, each v also defines a duration δ_ℓ, e.g., $\hat{\delta}_\ell = \frac{\mu_e}{\hat{v}}$. The accents on the δ's indicate which v generate them. First, consider the case when (150) is not active (e.g., diesel trucks). The optimum point is given by:

$$
\delta_0 = \begin{cases} \underline{\delta}_0, & if \quad \underline{\delta}_0 + \underline{\delta}_\ell \geq \underline{D} \\ \min\left(\overline{\delta}_0, \underline{D} - \underline{\delta}_\ell, \max\left(\underline{\delta}_0,\ \underline{D} - \hat{\delta}_l,\ \underline{D} - \overline{\delta}_\ell\right)\right), & o.w. \end{cases}
$$

$$
\delta_\ell = \begin{cases} \underline{\delta}_\ell, & if \quad \underline{\delta}_0 + \underline{\delta}_\ell \geq \underline{D} \\ \underline{D} - \delta_0, & if \left(\underline{\delta}_0 + \underline{\delta}_\ell < \underline{D}\right) \wedge \left(\delta_0 \neq \underline{\delta}_0\right) \\ \min\left(\overline{\delta}_\ell,\ \max\left(\underline{\delta}_\ell,\ \tilde{\delta}_l,\ \underline{D} - \underline{\delta}_0\right)\right), & o.w. \end{cases}
$$

If the point (δ_0, δ_ℓ) satisfies (150), then it is optimal. Otherwise, it means that (150) must be active. In this case, we can define 7 candidate points and the sufficient conditions for them to be the optimum. The candidate points are given by the point along g_1 with minimum cost and the points where (150) intersects other constraints, and the conditions conditions are derived from each point's KKT conditions. Table 30 presents the candidate solutions and their conditions. Feasibility is a basic necessary condition for any solution, and was thus omitted from the table. $P^{-1}(v)$ refers to the inverse of $P(v)$ over the domain $v \in [\breve{v}, \infty)$. The points x_2

and x_3, representing the candidates where (150) and one of the constraints forming (151) intersect, might be computationally expensive to calculate, so we can leave testing them for last. We can also use approximate solutions instead of solving it exactly. Note that the conditions are generated from speeds that can be calculated beforehand. Therefore, we may be able to directly eliminate some candidate solutions based on δ_ℓ's domain.

Table 30. Solution Candidates

Point	Condition
$x_1 = (P(v^*), \delta_\ell^*)$	-
$x_2 = \left(\overline{D} - \delta_2, \delta_2\right),\ P(\mu_e/\delta_2) = \overline{D} - \delta_2$	$\check{\delta}_\ell < \delta_2 \le \delta_\ell^*$
$x_3 = \left(\underline{D} - \delta_3, \delta_3\right),\ P(\mu_e/\delta_3) = \underline{D} - \delta_3$	$\delta_\ell^* \le \delta_3 \le \hat{\delta}_\ell$
$x_4 = \left(\underline{\delta}_0, P^{-1}\left(\underline{\delta}_0\right)\right)$	$\tilde{\delta}_\ell \le P^{-1}\left(\underline{\delta}_0\right) \le \delta_\ell^*$
$x_5 = \left(\overline{\delta}_0, P^{-1}\left(\overline{\delta}_0\right)\right)$	$\delta_\ell^* \le P^{-1}\left(\overline{\delta}_0\right) \le \check{\delta}_\ell$
$x_6 = \left(P\left(\mu_e/\underline{\delta}_\ell\right), \underline{\delta}_\ell\right)$	$\underline{\delta}_\ell \ge \delta_\ell^*$
$x_7 = \left(P\left(\mu_e/\overline{\delta}_\ell\right), \overline{\delta}_\ell\right)$	$\overline{\delta}_\ell \le \delta_\ell^*$

Cost Lower Bound

Let $A^d \subset A'$ represent the set of all arcs with driving as their assigned activity. For every node pair (p, q) such that there is a directed path from p to q , let $\mathcal{D}(p,q)$, $\mathcal{D}_d(p,q)$, and $\mathcal{D}_\ell(p,q)$ be, respectively, the minimum travel time (including service time), minimum driving time and minimum travel distance between nodes p and q with all resource, time-window and HOS constraints relaxed:

$$\mathcal{D}(p,q) = \begin{cases} \min\left(\Delta_{pq}\right), & if\,(p,q) \in A' \\ \min\limits_{(p,k)\in A'} \left(\min\left(\Delta_{pk}\right) + \mathcal{D}(k,q)\right), & o.w. \end{cases}$$

$$\mathcal{D}_d(p,q) = \begin{cases} 0, & if\,(p,q) \in A' \setminus A^d \\ \min\left(\Delta_{pq}\right), & if\,(p,q) \in A^d \\ \min\limits_{(p,k)\in A'} \left(\mathcal{D}_d(p,k) + \mathcal{D}_d(k,q)\right), & o.w. \end{cases}$$

$$\mathcal{D}_\ell(p,q) = \begin{cases} 0, & if\,(p,q) \in A' \setminus A^d \\ \mu_{pq}, & if\,(p,q) \in A^d \\ \min\limits_{(p,k)\in A'} \left(D_\ell(p,k) + D_\ell(k,q)\right), & o.w. \end{cases}$$

If there is no directed path from p to q , then $\mathcal{D}(p,q) = \mathcal{D}_d(p,q) = \mathcal{D}_\ell(p,q) = \infty$.

Let $D_{HOS}(d, \psi)$ represent the minimum duration of a HOS-compliant trip with d driving hours and initial resource vector ψ, assuming the driver can rest anywhere, and without considering service time and time-window constraints, i.e., if a driver were at the beginning of an empty straight road with length equivalent to d driving hours where he/she can rest anywhere, given an initial resource vector ψ, how long would he/she take to reach the end of the road without breaking the HOS regulations. A method to calculate $D_{HOS}(d, \psi)$ is described in Appendix I. Let $\mathcal{D}_s(p, q)$ be the service time required between nodes p and q. If the objective were simply to minimize trip duration, the lower bound $\mathcal{L}_{dur}$ can be calculated as:

$$\mathcal{L}_{dur}(p, q, \psi) = D_{HOS}(\mathcal{D}_d(p, q), \psi) + \mathcal{D}_s(p, q)$$

However, when considering a combination of trip duration and energy/fuel consumption or emissions as the objective function, the lower bound generated using only the duration term ($\alpha \mathcal{L}_{dur}(p, q, \psi)$) is too loose and not as useful. Therefore, we need a lower bound on the fuel consumption/emissions.

Bound 1

Idling cost: Let γ the energy/fuel consumption rate when idle (resting or service). A lower bound on the idling cost is given by:

$$\mathcal{L}_{idl1}(p, q, \psi) = (\beta\gamma + \theta)\big(D_{HOS}(\mathcal{D}_d(p, q), \psi) - \mathcal{D}_d(p, q) + \mathcal{D}_s(p, q)\big) \qquad (154)$$

$\mathcal{D}_s$ is fixed as client visits are mandatory. $\mathcal{D}_d$ considers the minimum driving time of each edge, and $D_{HOS}(d, \psi) - d$ is monotonically increasing in d (required rest time cannot decrease when driving time increases), so $\mathcal{L}_{idl1}$ is a lower bound on idling cost. Note that if the cost/consumption parameters for rest and service time are different, the term $\mathcal{D}_s(p, q)$ will appear separately multiplying its own parameter.

Driving consumption: Let v_{min} be the minimum travel speed allowed in the network. We assume that the fuel consumption per time $FC(v)$ is monotonically increasing in the range of speeds used in the problem, as is the case for the model we use. Therefore, $FC(v_{min})$ gives a lower bound on the energy/fuel consumption rate when driving. A lower bound on the consumption due to driving is given by:

$$\mathcal{L}_{f\_dr1}(p, q) = 3600 \cdot FC(v_{min})\mathcal{D}_d(p, q)$$

An alternative is using the minimum travel distance $\mathcal{D}_\ell(p, q)$ and the speed v_ℓ that minimizes the fuel consumption per distance, $\zeta(v)$, (or the nearest feasible speed) to generate a energy/fuel consumption lower bound.

Cost: Consider the cost function defined in (146). A cost lower bound is given by:

$$\mathcal{L}_{cost1}(p, q, \psi) = \alpha \mathcal{L}_{dur}(p, q, \psi) + \beta \mathcal{L}_{f\_dr1}(p, q) + \mathcal{L}_{idl1}(p, q, \psi)$$

Note that the driving energy/fuel consumption bound is calculated using the minimum travel speed, whereas the idling cost and trip duration bounds are calculated using the maximum travel speed. Therefore, this bound is not tight.

Bound 2

When calculating analytical solutions in Last driving decision, we showed how to calculate the optimal speed based on energy/fuel and duration costs, and consumption model. We now use this information to refine the lower bound.

Driving time: Bound 1 used a driving time considering the maximum travel speed. However, depending on the cost function, the cost increase due to fuel consumption at higher speeds may exceed savings due to shorter trip duration. Optimal solutions are expected to tend towards using the optimal speed $\tilde{v}$ (limited by possible increases in required rest time). With this in mind, we scale the driving time so that it represents the travel time at the optimal speed (or the nearest feasible speed).

$$\begin{aligned} v_t &= max(min(\tilde{v}, v_{max}), v_{min}) \\ \widetilde{\mathcal{D}}_d(p,q) &= \mathcal{D}_d(p,q)\frac{v_{max}}{v_t} \end{aligned}$$

This scaling assumes that all edges have the same speed limits and optimum speed. An alternative (but still assuming that all edges have the same optimum speed) would be to use the length of the minimum length path, $\mathcal{D}_\ell(p,q)$, to estimate a lower bound on the driving cost when traveling with speed v_t. A more general approach would be to, when building the graph, calculate v_t for each edge, and store in each edge the travel time and cost associated with v_t. The stored costs can be used to calculate a minimum cost path and its driving time. In both alternatives, the minimum cost (we refer to it as $\mathcal{L}_{dr\_cost}(p,q)$) can be used as a lower bound on the driving related costs (due to both emissions and duration) and we would require only to complement it with a lower bound on the idling costs (due to both emissions and duration).

It is important to remember that, due to HOS regulations, increasing driving time may end up increasing required rests. The extra rest time caused by driving time scaling is given by:

$$\Lambda = D_{HOS}\big(\widetilde{\mathcal{D}}_d(p,q),\psi\big) - \widetilde{\mathcal{D}}_d(p,q) - \big(D_{HOS}(\mathcal{D}_d(p,q),\psi) - \mathcal{D}_d(p,q)\big)$$

Trip duration and fuel consumption are calculated following the same ideas as Bound 1 but using the scaled driving time and correcting trip duration and idling time to remove the extra rest time.

Trip Duration: The trip duration is calculated as follows:

$$\mathcal{L}_{dur2}(p,q,\psi) = D_{HOS}\big(\widetilde{\mathcal{D}}_d(p,q),\psi\big) - \Lambda + \mathcal{D}_s(p,q)$$

Idling cost: The idling cost lower bound is given by the same expression as (154) due to the rest time correction, i.e.,

$$\mathcal{L}_{idl2}(p,q,\psi) = \mathcal{L}_{idl1}(p,q,\psi).$$

Driving consumption: Energy/fuel consumption due to driving is given by:

$$\mathcal{L}_{f\_dr2}(p,q) = 3600 \cdot FC(v_t)\widetilde{\mathcal{D}}_d(p,q)$$

Cost: A cost lower bound is given by:

$$\mathcal{L}_{cost2}(p,q,\psi) = \alpha\mathcal{L}_{dur2}(p,q,\psi) + \beta\mathcal{L}_{f\_dr2}(p,q) + \mathcal{L}_{idl2}(p,q,\psi)$$

Note that while $\mathcal{L}_{dur2}(p,q,\psi) \geq \mathcal{L}_{dur}(p,q,\psi)$ and $\mathcal{L}_{f\_dr2}(p,q) \geq \mathcal{L}_{f\_dr1}(p,q)$, $\mathcal{L}_{dur2}$ and $\mathcal{L}_{f\_dr2}$ are consistent with respect to the travel speed used for their calculation, and use a speed that minimizes cost (not accounting for mandatory rests). As the rest (idling) time is kept as the one from the minimum duration path, the rest time is the minimum feasible. Decreasing $\mathcal{L}_{dur2}$ would imply that one or more edges are using a speed greater than the optimal, causing an increase in fuel consumption costs that exceeds the savings in trip duration costs. Similarly, decreasing $\mathcal{L}_{f\_dr2}$, would cause an increase in trip duration costs, and increase overall cost. Therefore, $\mathcal{L}_{cost2}$ is a lower bound. Each term is not a lower bound for the value it approximates, but they are calculated so that they generate a cost lower bound. If the driving cost lower bound $\mathcal{L}_{dr\_cost}(p,q)$ is calculated directly, then the cost lower bound is given by:

$$\mathcal{L}_{cost2}(p,q,\psi) = \mathcal{L}_{dr\_cost}(p,q) + (\alpha + \gamma\beta + \theta)\underbrace{\left(D_{HOS}(\mathcal{D}_d(p,q),\psi) - \mathcal{D}_d(p,q) + \mathcal{D}_s(p,q)\right)}_{idling\ time}$$

Graph Preprocessing

In the approximate dynamic programming algorithm used, we store the decision and cost for several states at each node. Therefore, having a large number of intermediate nodes between rest areas increases both the number of decisions needed to reach the destination and the storage space required by the algorithm. Furthermore, when optimizing travel speed to reduce fuel consumption, the precision with which speed can be adjusted depends on the time resolution used in the decision space, but also on the length of any given edge. If an edge is too short, any change in duration might generate a travel speed outside of the allowed range. In order to reduce the number of nodes in the graph, we use a stop-based graph based on the road network and remove short edges between nearby rest areas (e.g., only consider rest areas that are at least 2h away from the current node). By stop-based graph we mean a graph that directly links possible stop locations (origin, rest areas, clients), analogous to customer-based graphs used for vehicle routing problems. However, the graph is not complete as each location is connected only to locations that were downstream in the original road network. As clients are mandatory stops and have a fixed order, nodes are not directly connected to nodes downstream of the next client. It can be seen as generating the stop-based graph based on the subnetworks connecting each pair of consecutive clients, as opposed to using the whole network directly. Figure 128 shows a graph representing a road network, whereas Figure 129

shows the stop-based graph that would be generated from that network. As our experiments set the same speed profile for all edges, each edge (i, j) of the stop-based graph was generated using the length of the minimum distance path between nodes i and j in the road network and setting the same speed profile used in the road network. We assume that a stop-based graph is known or can be obtained by the user, and do not cover the specifics of its construction for general networks. Algorithms to construct customer-based graphs for time-dependent road networks were proposed in (Ben Ticha et al. 2021).

Given a stop-based graph, we remove edges that have distance or minimum travel time shorter than chosen limits, except when one of the edge's nodes is a client, the origin, or the destination. In our experiments, the time and distance limits were set to 2h and 100km, respectively. In addition, as HOS regulations limit driving time, edges with minimum travel time greater than 8h were also removed. Although it is possible for the fastest path between locations to vary with time in time-dependent networks, we assume that edge lengths (distance) are fixed in the stop-based graph.

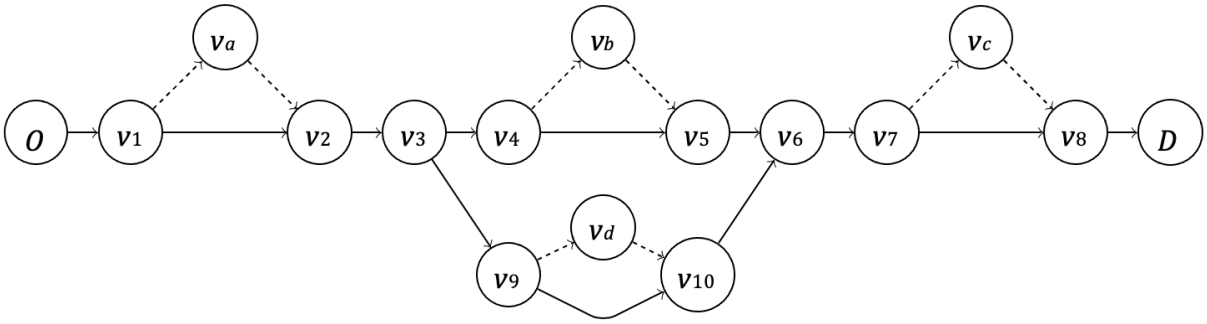


Figure 128. Example graph focusing on the road network. Focuses on rest area (nodes with letter indexes) placement along main roads. Easy to visualize but has a large number of intermediate nodes (nodes with number indexes).

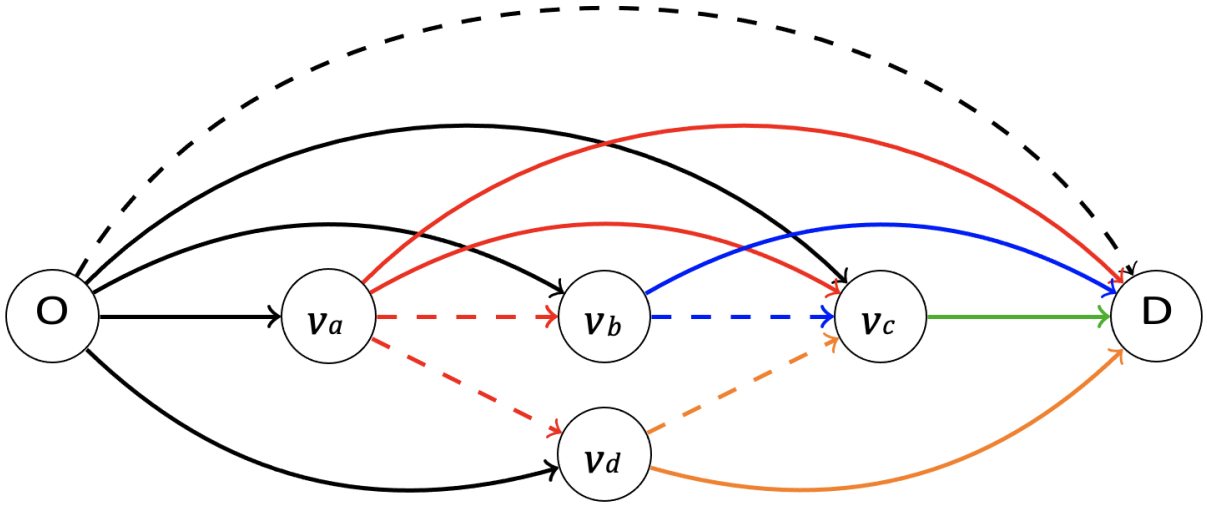


Figure 129. Stop-based graph generated from Figure 128 to focus on the connection between possible stops (rest areas, clients, origin, destination). Each possible stop is directly connected to downstream stops satisfying predetermined conditions. Dashed arrows exemplify edges that could be removed for being too short or too long.

Case Study

Experiments were performed on the graph network shown in Figure 130. Every node p_i is assumed to be a charging station that allows for long term parking. Nodes v_4 and D are clients with daily time-windows $[12,16]$ and $[8,16]$, respectively. All charging stations have the same charging power (100 kW or 50 kW, depending on the scenario), and are subject to availability time-windows. Within each scenario, probability distributions are defined for the start and end times of the availability time-windows. The same pair of distributions is used for all locations and days, but they are sampled separately. Table 31 describes the probability distributions used to generate the 3 different types of time-windows used. The battery capacity values tested were 400, 600, 800, and 1000kWh. Fuel constraints for diesel trucks were relaxed by setting a large fuel capacity. The average travel speed is set to 75km/h at all edges. As the energy consumption is affected by speed, speed can also affect vehicle range and problem feasibility. Therefore, for scenarios with feasibility issues, we performed experiments where the vehicle is allowed to reduce its speed to 70% of the average travel speed (around 52.5km/h). Our objective is to study how battery capacity, charging power and chargers/parking' availability affect BETs performance in terms of trip duration, CO2 emissions and route feasibility, and compare it with a diesel truck's performance.

Table 31. Experiment Parameters

Parking Time-windows Distribution		
	Start Time (h)	**End Time (h)**
Narrow	N(9,1)	N(16,1)
Medium	N(7,1)	N(19,1)
Wide	N(5,1)	N(22,1)

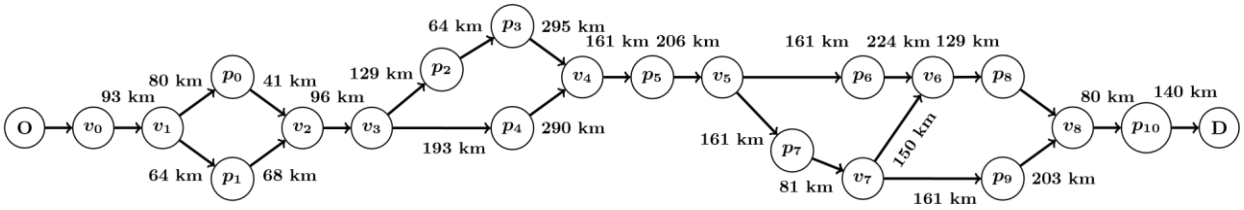


Figure 130. Network used for experiments. Arc lengths are given in kilometers.

Figure 131 shows how trip duration is affected by parking availability, charging power and battery capacity. All scenarios were infeasible for the 400kWh capacity, so no results are shown. It can be seen that when wide time-windows are used, diesel trucks hold a significant advantage in terms of trip duration, especially when lower charging power and battery capacities are used. This result is expected and likely caused by schedule adjustments needed to accommodate longer recharging stops. For example, due to the HOS regulations and parking time-windows, short delays might push drivers close to their driving limits and require them to add another *daily rest* (10h minimum) to the trip. However, as we consider more severe parking shortage scenarios by narrowing the time-windows, this advantage is gradually reduced. When narrow time-windows are considered, the BET's performance was comparable to the baseline diesel truck's. If parking availability is limited, diesel trucks would also need to adjust their schedules to guarantee appropriate parking regardless of having longer range and faster refueling times. Figure 132 shows results regarding CO2 emissions. In this aspect, BETs present a clear advantage in all feasible scenarios. In scenarios where 600kWh battery capacity was considered we encountered some feasibility issues, shown in Figure 133. As energy consumption is speed-dependent, we rerun these experiments allowing the speed to be adjusted between 70%-100% of the road's average speed. Figure 133 shows how this allowed speed reduction improves feasibility. The effects of the allowed speed reduction on the trip duration and CO2 emissions are shown in Figure 134 and Figure 135.

These results show that diesel trucks have an advantage (regarding trip duration) under ideal parking availability conditions, but that advantage is greatly reduced when we account for current parking shortage issues. Although the number of charging stations available is not comparable to the number of regular truck stops and gas stations, the number of BETs in operation will also be limited at first. Furthermore, integrating speed optimization into the planning mitigates range and feasibility problems. While larger battery capacity and charging power help narrowing the performance gap between BETs and diesel trucks, the battery and charger requirements are less restrictive when practical constraints are considered. It is true

that BETs present a series of limitations and that they cannot replace diesel trucks in every situation. However, it is important to remember that diesel trucks do not operate in ideal scenarios. Having a larger range does not mean that this range is always gonna be needed. Being able to refuel fast does not mean that breaks are limited to short refueling stops. Similarly, it is unrealistic to evaluate BETs as if they were supposed to be drop-in replacements for diesel trucks. BETs will require different itineraries, but, as seen in Figure 131, the impact of needed adjustments may be small depending on each application's constraints. Therefore, when comparing the performance of BETs and diesel trucks, it is important to account for the impacts of practical constraints such as HOS regulations, delivery and parking constraints.

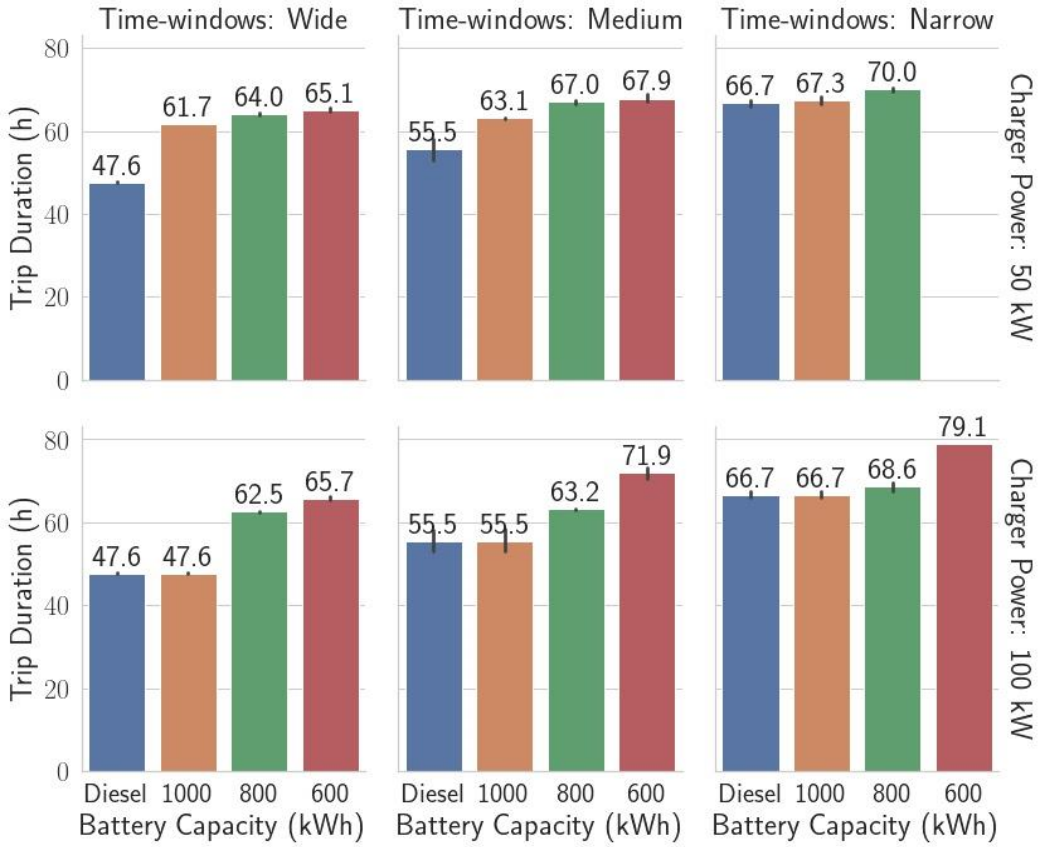


Figure 131. Trip duration under different parking availability and charging infrastructure conditions.

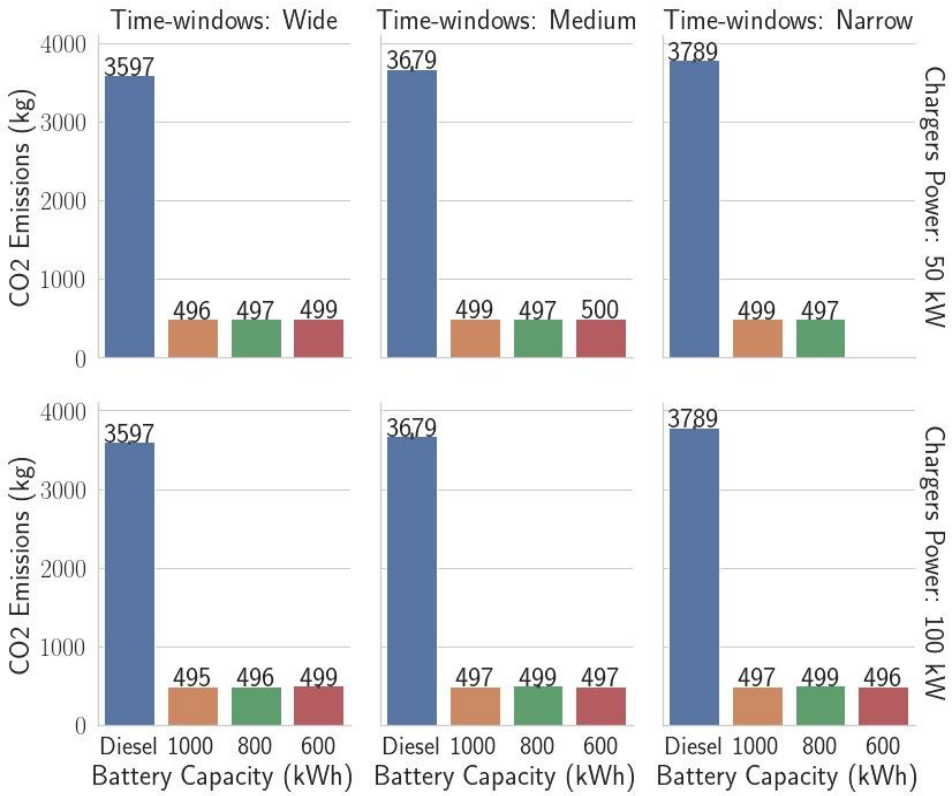


Figure 132. CO2 emissions under different parking availability and charging infrastructure conditions.

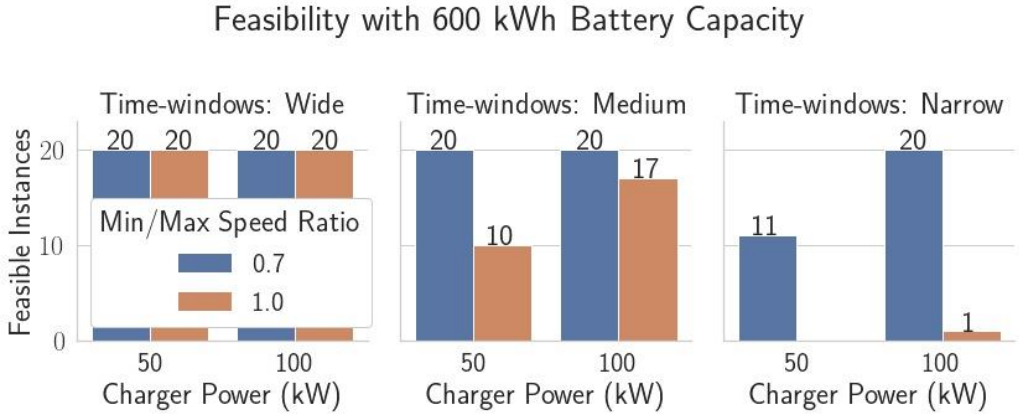


Figure 133. Trip feasibility under different parking availability and charging infrastructure conditions for scenarios with 600kWh battery capacity.

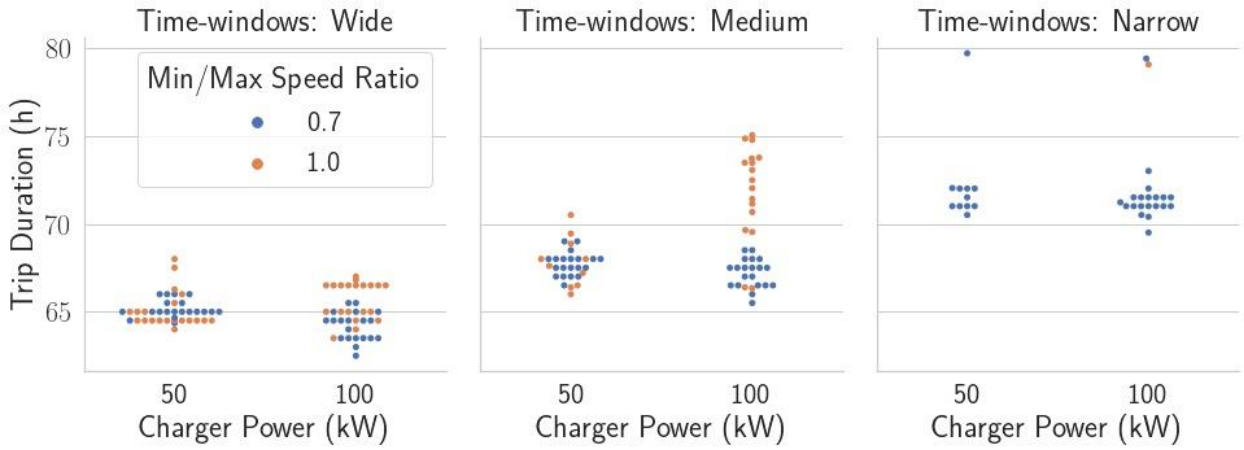


Figure 134. Trip duration under different parking availability and charging infrastructure conditions for scenarios with 600kWh battery capacity.

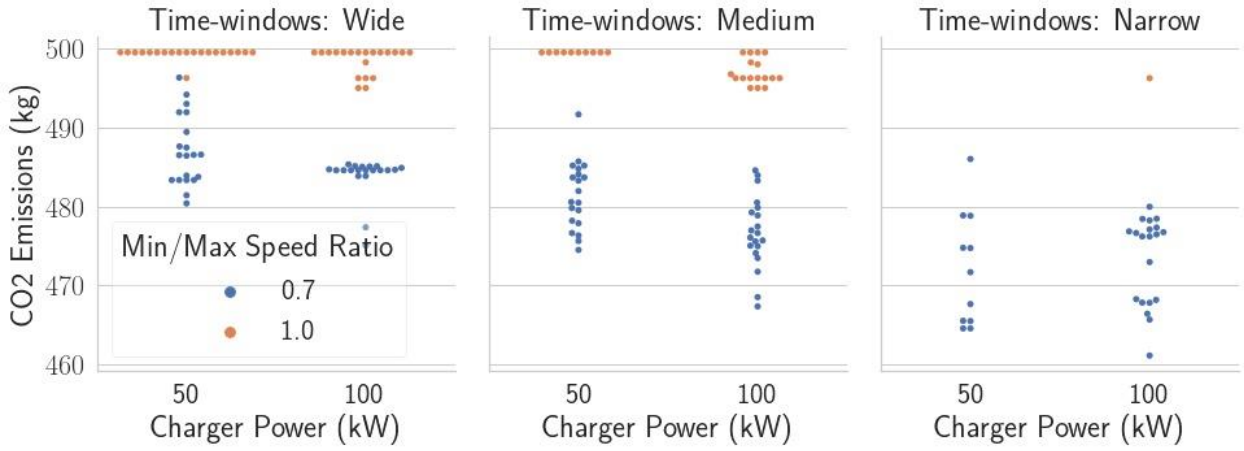


Figure 135. CO2 emissions under different parking availability and charging infrastructure conditions for scenarios with 600kWh battery capacity.

Experiments on Random Networks

Without fast chargers

Experiments were performed on 10 random networks with between 15 and 100 charging stations. The average distance between charging stations is set to 200km. All charging stations have the same charging power (100 kW or 50 kW, depending on the scenario), and are subject to availability time-windows. Within each scenario, probability distributions are defined for the start and end times of the parking availability time-windows. The same pair of distributions is used for all locations and days, but they are sampled separately. Table 31. Experiment Parameters describes the probability distributions used to generate the 3 different types of time-windows used. We consider a $[0,24]$ departure time-window (first day only), a daily $[8,16]$ delivery time-window for the destination, and 100h planning horizon. The battery capacity values tested were 500, 800, and 1000 kWh. Fuel constraints for diesel trucks were relaxed by

setting a large fuel capacity. The maximum speed limit is set to 80km/h at all edges. As the energy consumption is affected by speed, speed can also affect vehicle range and problem feasibility. Therefore, in the experiments, we also vary whether the speed limit is taken as the average travel speed or if the vehicle is allowed to reduce its speed to 70% of the speed limit (around 56km/h). Our objective is to study how battery capacity, charging power, speed control and chargers/parking' availability affect BETs performance in terms of trip duration, CO2 emissions and route feasibility, and compare it with a diesel truck's performance.

Figure 136 shows how average trip duration is affected by parking availability, charging power and battery capacity. In scenarios with wide time-windows and 50kW chargers, diesel trucks hold a significant advantage in terms of trip duration. This result is expected as schedule adjustments may be needed to accommodate longer recharging stops. HOS regulations and parking time-windows may also exacerbate this effect in some instances, as short delays can push drivers close to their working hours limits and require them to add another *daily rest* (10h minimum) to the trip. However, as we consider severe parking shortage scenarios (narrow time-windows), this advantage is reduced. If parking availability is limited, diesel trucks also need to adjust their schedules to guarantee appropriate parking regardless of having longer range and faster refueling times. When 100kW chargers are considered, the average trip duration of BETs with 1MWh battery is comparable to the baseline diesel trucks' in all scenarios. Showing that, in these instances, adequate synchronization of recharging and resting times completely negated the charging time and range issues. In terms of CO2 emissions, BETs present a clear advantage in all feasible scenarios, as shown in Figure 137. Figure 138 presents how route feasibility was affected by parking availability, charging power, battery capacity and speed flexibility. Note that the results shown are the aggregate of all 10 networks used. Although feasibility issues were mitigated by higher charging power and travel speed flexibility, they were not resolved, and some networks had zero feasibility rate for one or more scenarios. Nevertheless, the effect of parking availability and speed optimization show that these are also important factors to consider when assessing the viability of truck electrification.

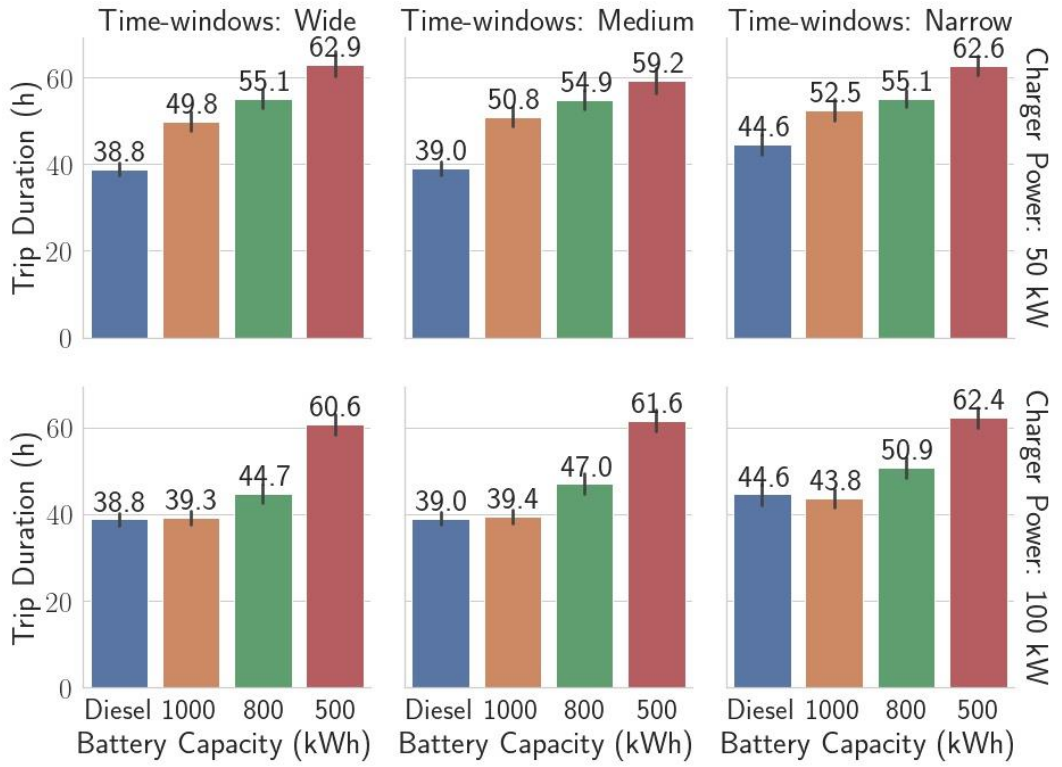


Figure 136. Average trip duration under different parking availability and charging infrastructure conditions. Includes only scenarios that allow speed reduction.

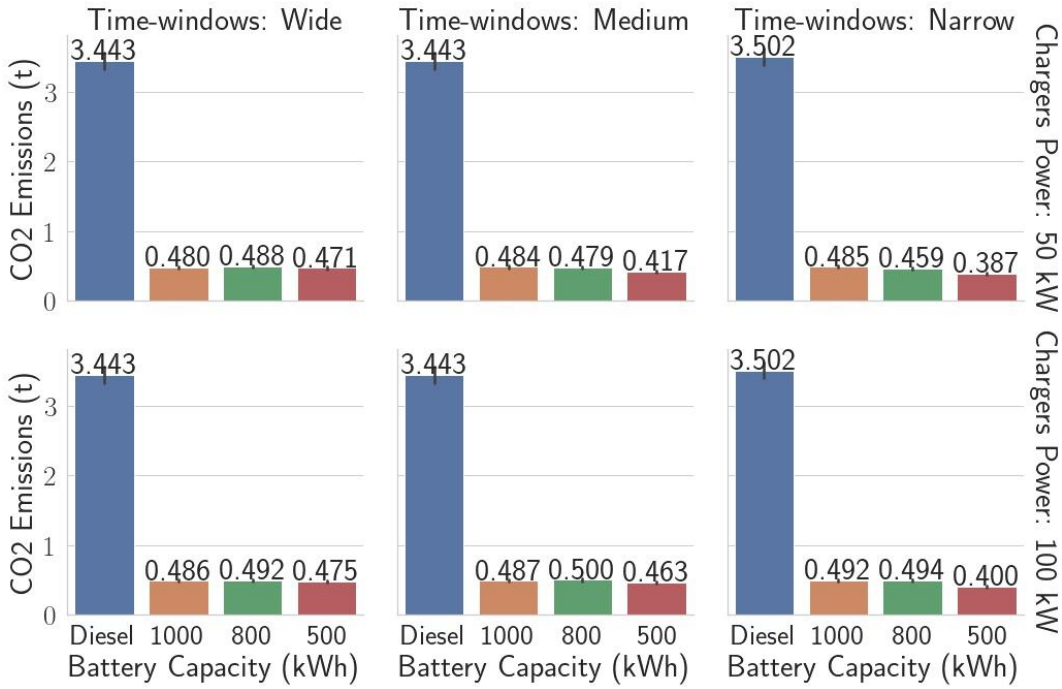


Figure 137. CO2 emissions under different parking availability and charging infrastructure conditions. Includes only scenarios that allow speed reduction.

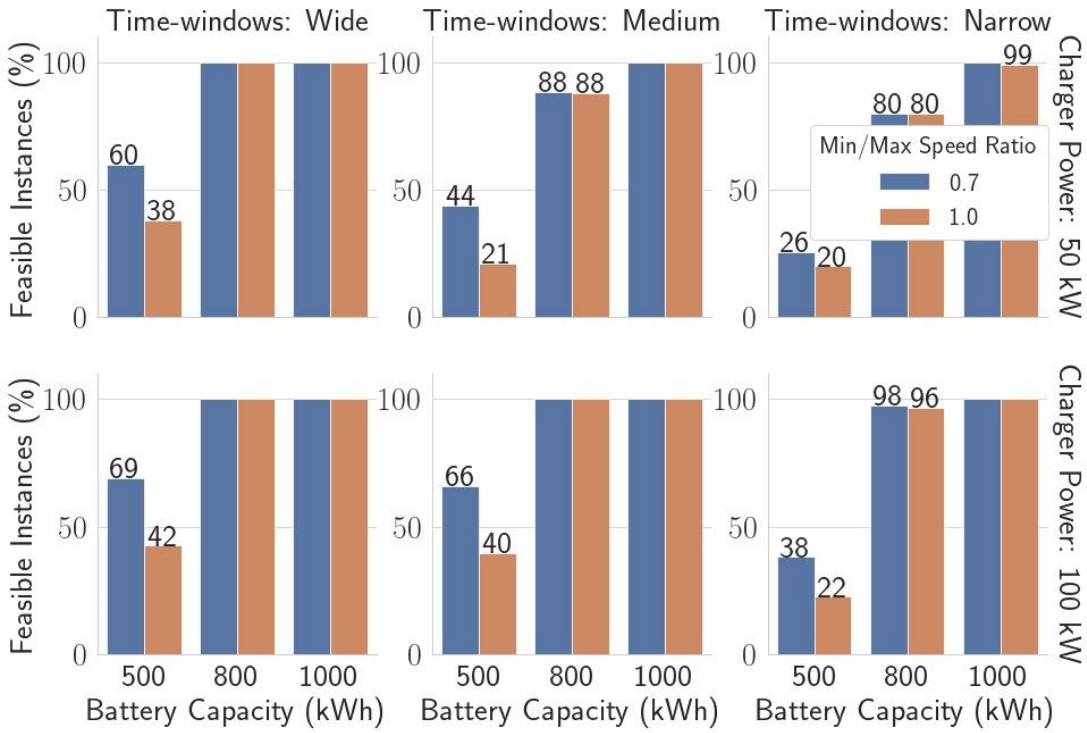


Figure 138. Percentage of instances that were feasible under different parking availability and charging infrastructure conditions. All diesel truck scenarios were feasible.

Fast chargers with fixed wait time

In these experiments, we use 10 randomly generated networks to study the effect of fast chargers. The average spacing between charging stations is set to 100km, the departure time constraint was narrowed to the interval $[6,12]$, and we fix the availability time-windows to the mean values of the distributions used previously. The scenarios considered vary the percentage of charging stations replaced by fast charging stations (0, 10, 20 or 30%), the charging power of the 2 types of chargers (regular chargers: 50 or 100kW, fast chargers: 150, 300, 500kW). Results for diesel trucks are also given as a baseline. The stop duration at fast chargers is limited to 3h. Charging stations with fast-chargers do not have availability time-windows, but, instead, have a fixed waiting time (0.5, 1 or 2h) before recharge starts. The waiting time is treated as service time, but it does not count towards the stop duration limit.

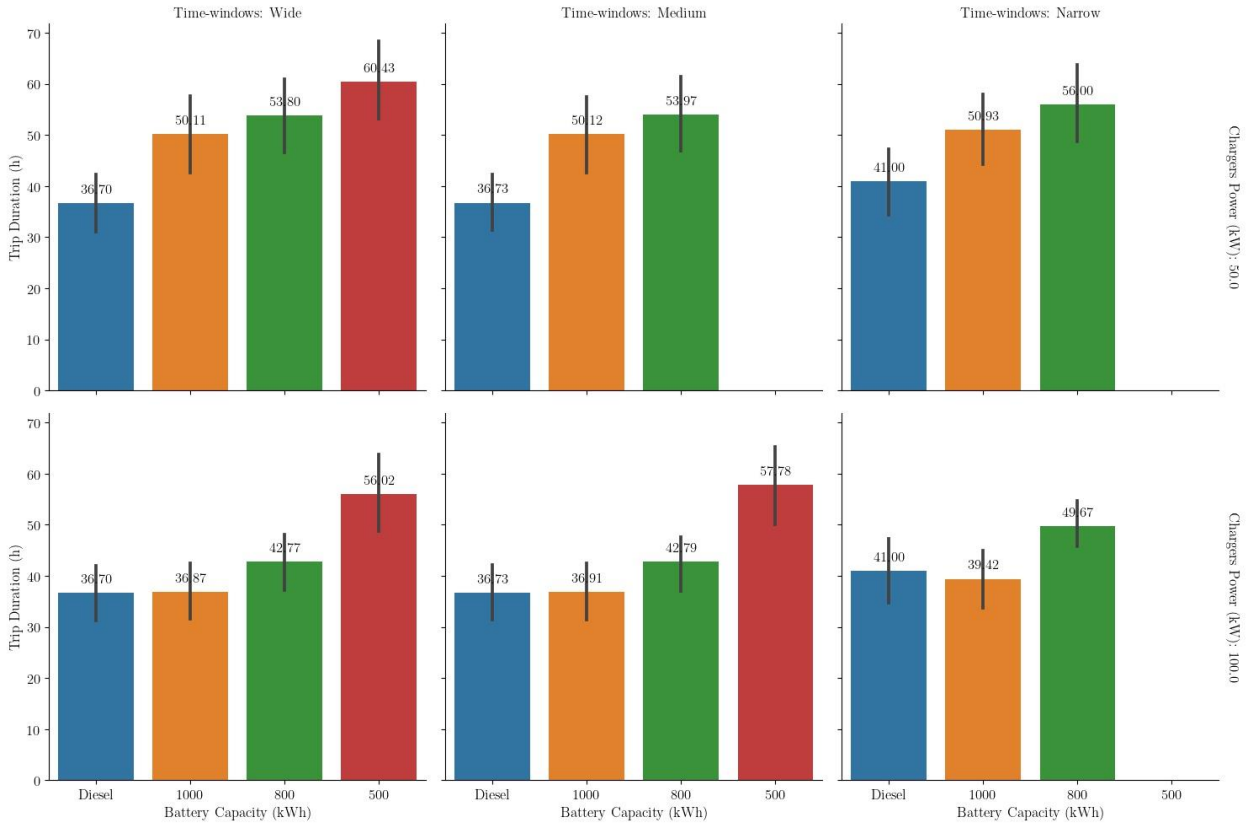


Figure 139. Average trip duration for scenarios with 100km avg. spacing between charging stations, and without fast chargers.

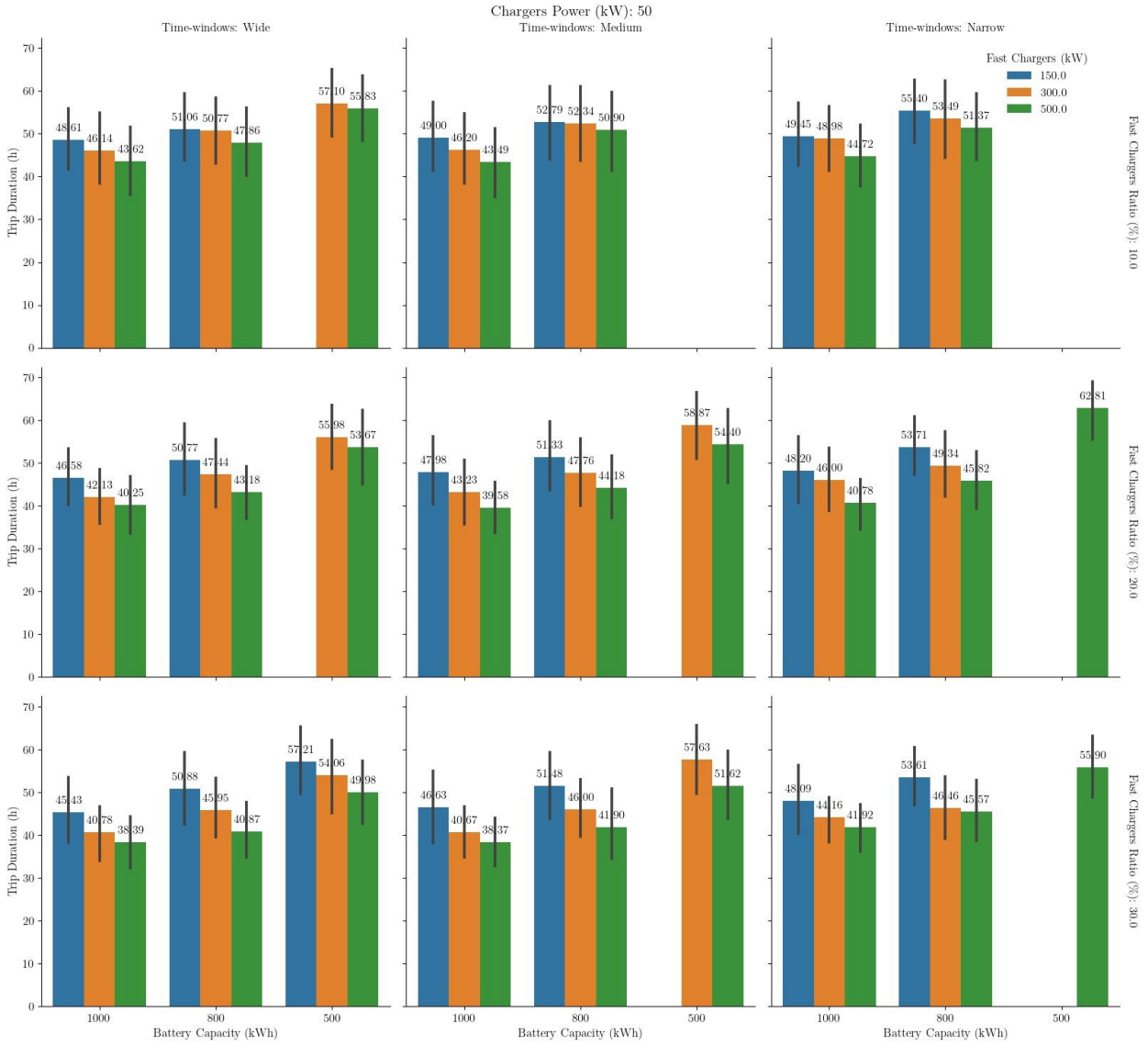


Figure 140. Average trip duration for scenarios with 100km avg. spacing between charging stations, 50kW chargers, and 0.5h wait.

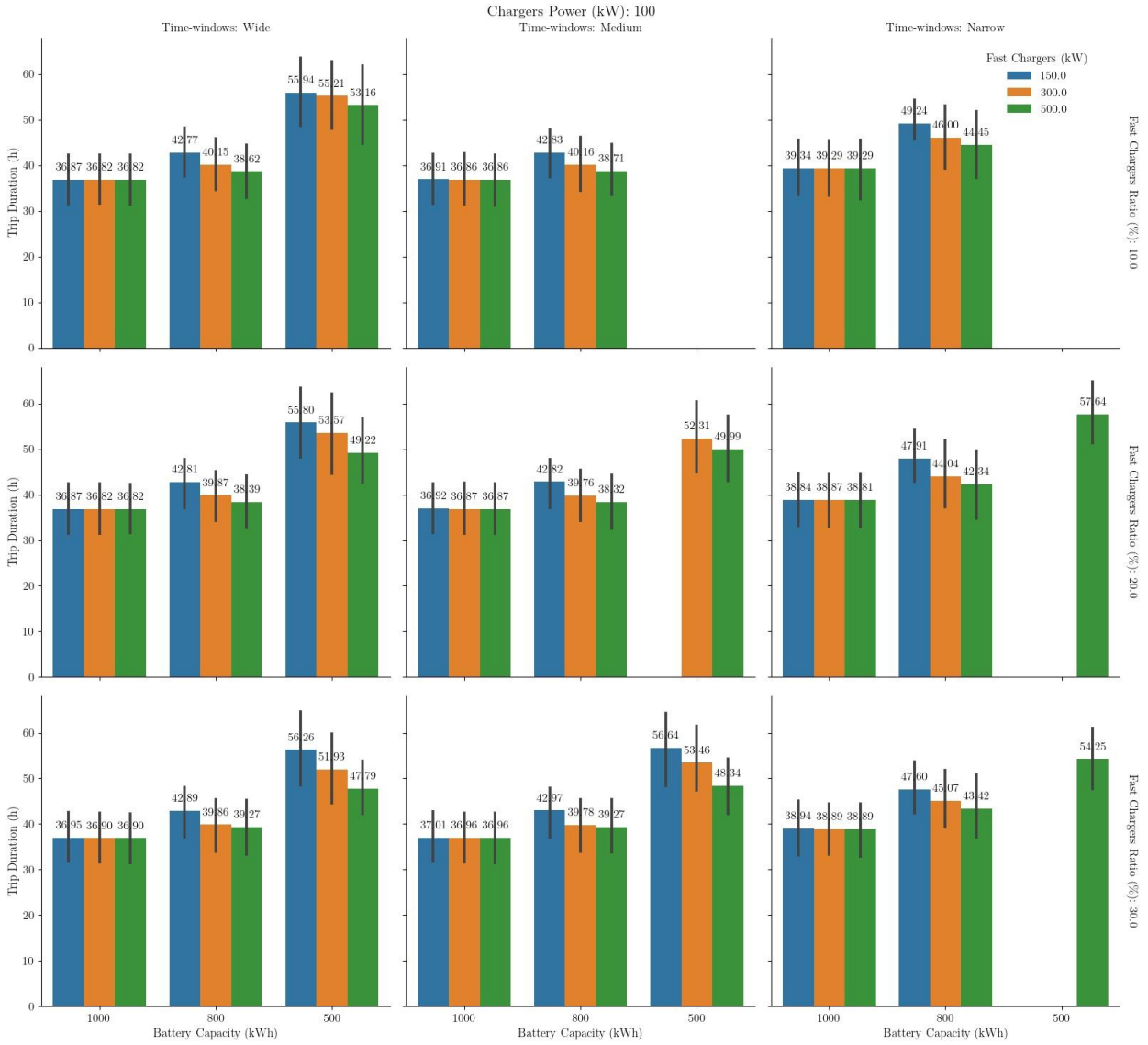


Figure 141. Average trip duration for scenarios with 100km avg. spacing between charging stations, 100kW chargers, and 0.5h wait.

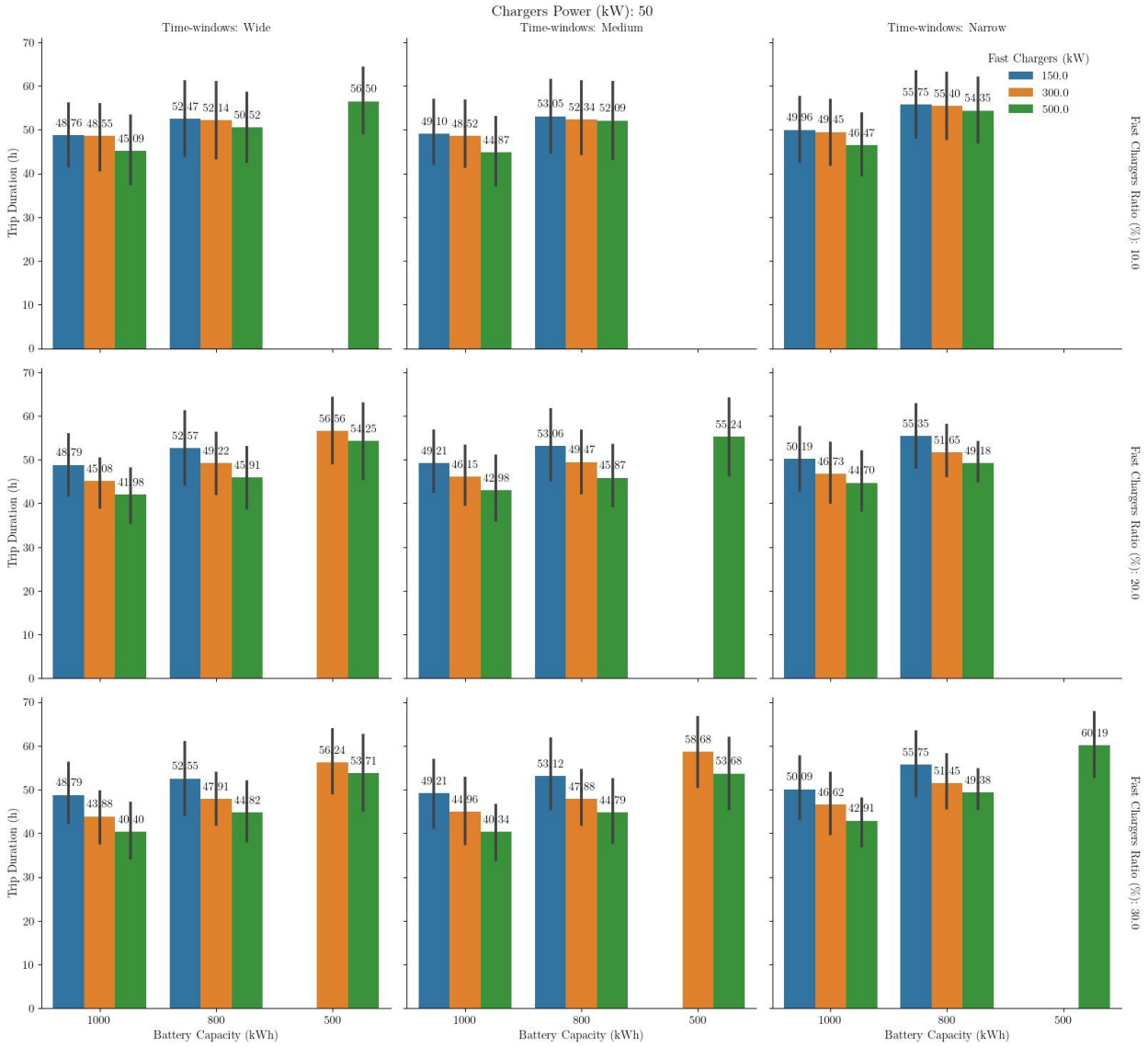


Figure 142. Average trip duration for scenarios with 100km avg. spacing between charging stations, 50kW chargers, and 1h wait.

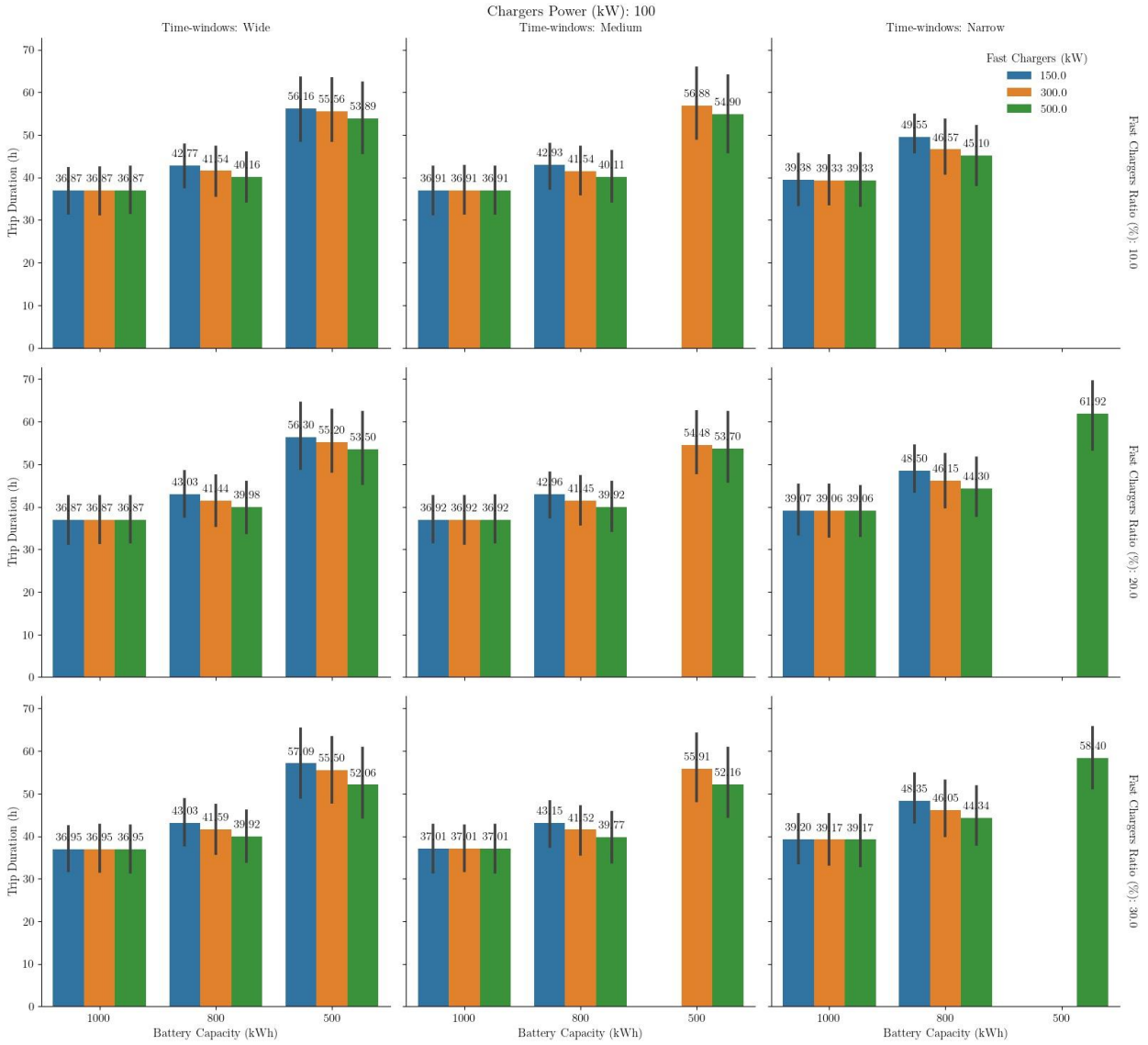


Figure 143. Average trip duration for scenarios with 100km avg. spacing between charging stations, 100kW chargers, and 1h wait.

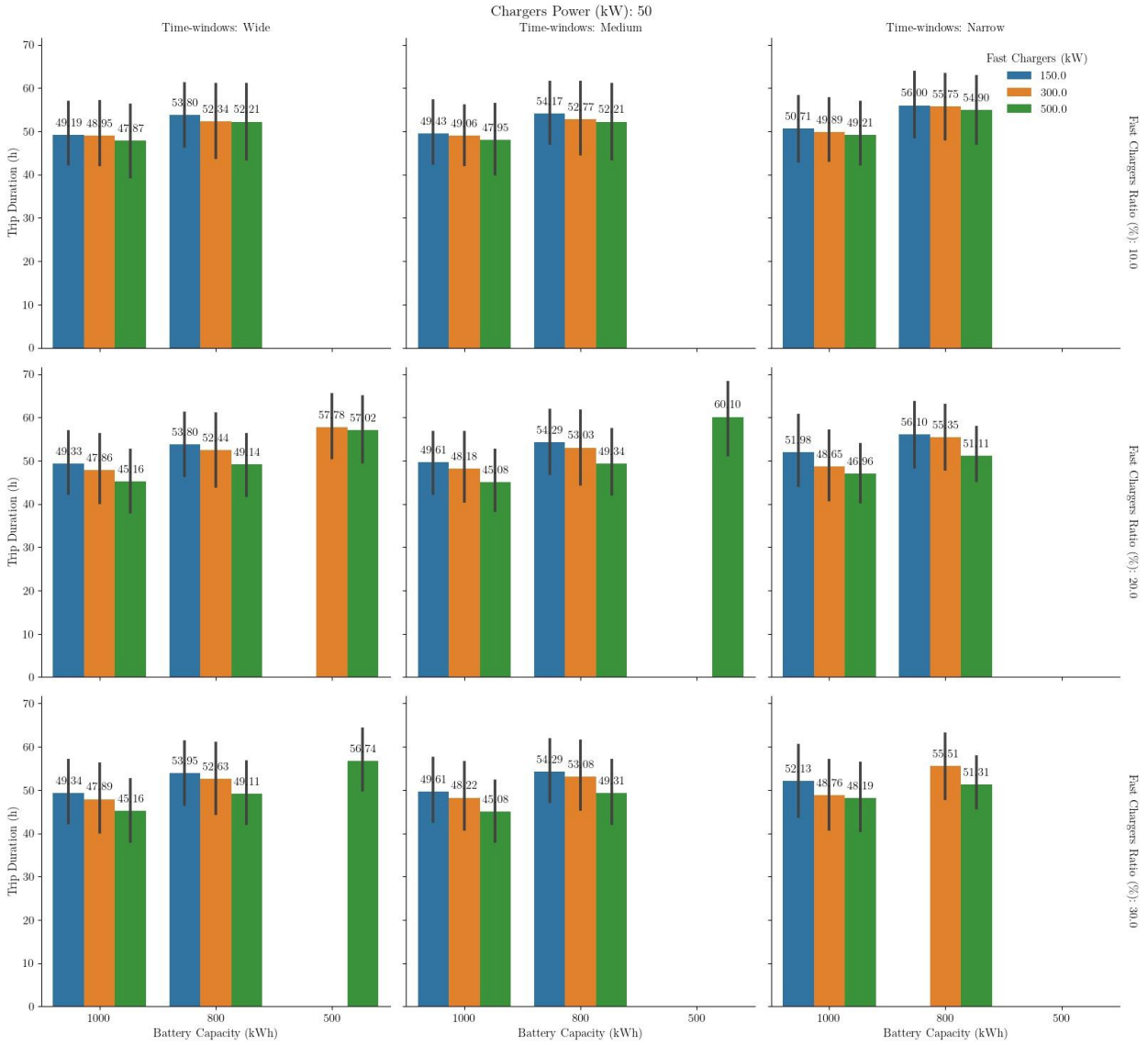


Figure 144. Average trip duration for scenarios with 100km avg. spacing between charging stations, 50kW chargers, and 2h wait.

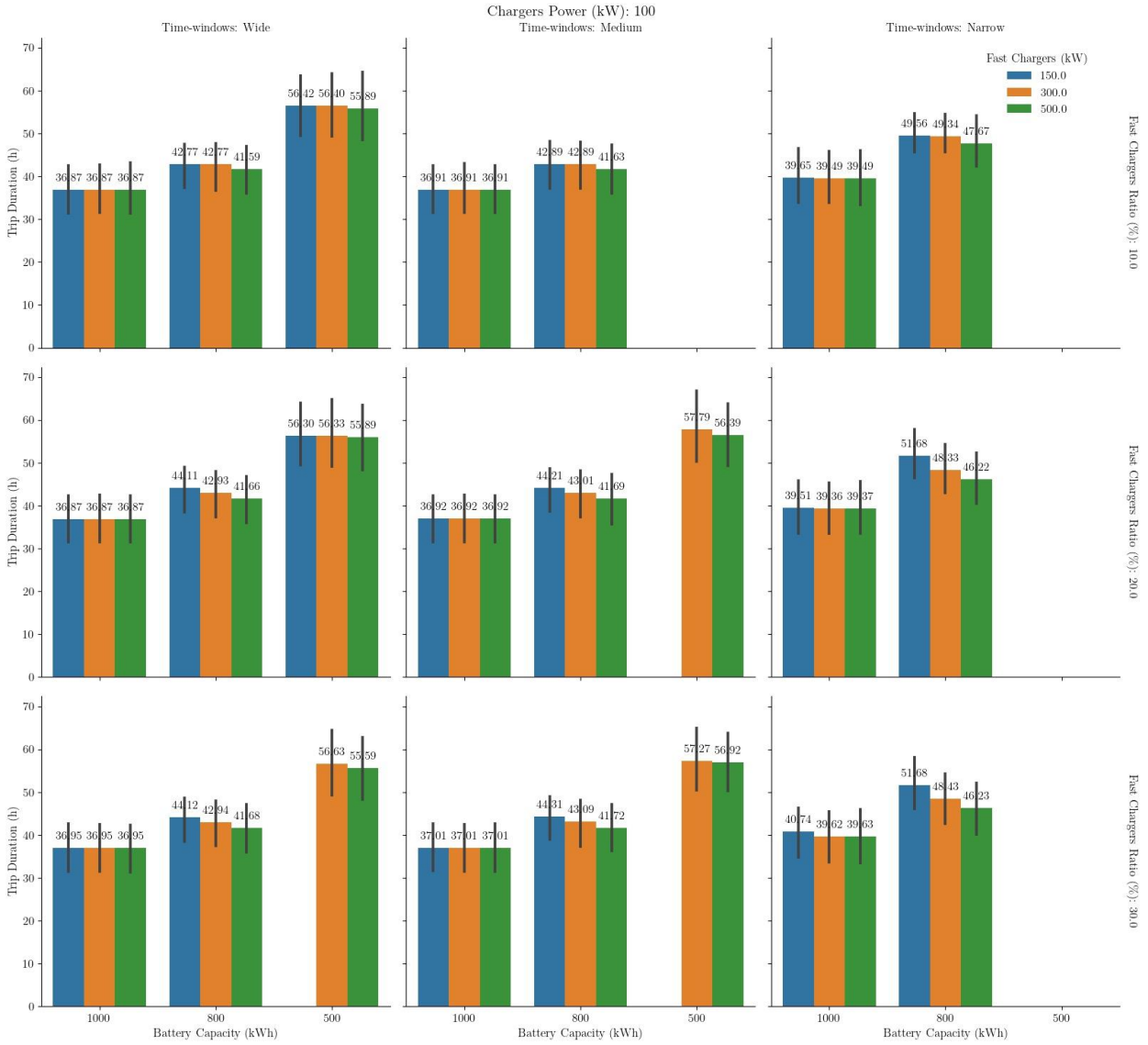


Figure 145. Average trip duration for scenarios with 100km avg. spacing between charging stations, 100kW fast chargers, and 2h wait.

Figure 139 shows the average trip duration for scenarios without fast chargers. The results obtained are similar to the ones from Figure 136. When 50kW chargers are used, the difference in trip duration between BETs and diesel trucks decreases when narrow time-windows are considered. When 100kW chargers are used, BETs with 1MWh batteries perform on the same level as diesel trucks. Figure 142 and Figure 143 present results when part of the charging stations are replaced by fast chargers with 1h wait. When regular chargers have only 50kW power, the inclusion of 300 and 500kW fast chargers show trip duration reductions of up to 20% compared to scenarios without fast chargers. In scenarios with 100 kW chargers, the improvement generated by fast chargers is significantly reduced, being completely eliminated for BETs with 1 MWh batteries. Furthermore, the improvements gained from increasing the percentage of fast chargers are also significantly smaller than in the scenarios with 50kW

chargers. This shows that although fast chargers are the intuitive way to address one of BET's main weak-points, depending on the context, it might be more advantageous to increase the number or power of regular chargers instead of increasing the number of fast chargers. Figure 140, Figure 141, Figure 144, and Figure 145 show the results for 0.5 and 2h waiting time.

These results shed some light on how the charging and parking infrastructure can impact BET performance and viability. Similar simulations can be used to determine appropriate infrastructure levels for a region given its expected BET population and available budget. For example, policymakers can study whether it is more beneficial to invest in increasing the number of chargers, or increasing the power of the installed chargers. It can also be used to advise truck drivers and trucking companies as to which types of vehicles will better fit the region's future infrastructure.

Conclusion

In this project, we studied how practical constraints, such as HOS (hours-of-service) regulations and limited parking availability, impact the performance gap between battery electric trucks (BETs) and diesel trucks. Both BETs and diesel trucks need to rest regularly due to HOS regulations, and, given the current truck parking shortage in the US and the risks associated with illegal truck parking, it is important to include parking information in the planning process. We include energy/fuel consumption constraints into the resource constrained shortest path formulation for the shortest path and truck driver scheduling problem with parking availability constraints proposed by the same authors in (Vital and Ioannou 2020). Experiments were performed on random networks to estimate the trip duration and C02 emissions of a baseline diesel truck and BETs with different battery capacities under different available charging power and parking shortage severity levels. In terms of emissions, BETs vastly outperformed the diesel truck in all feasible scenarios. Furthermore, when parking availability is limited, the performance gap (in terms of trip duration) between BETs and diesel trucks is greatly reduced in scenarios with 50kW chargers, and further reduced when 100kW chargers are considered. In addition, scenarios with fast charging stations show how the benefits of replacing some low-power charging stations by fast-chargers vary with the power and expected availability/wait of chargers from either type. This illustrates the importance of considering the various constraints to which drivers are subjected when evaluating the viability of BETs, and compare not if the solutions used by diesel trucks are feasible for BETs, but whether solutions tailored for BETs have a good enough performance in a particular region.

Conclusion

This project developed methods to integrate parking availability information into the planning process for long-haul trucking and studied truck parking shortages' potential impacts on the industry. First, we studied the truck driver scheduling problem (TDSP), which considers a fixed route and aims to determine a minimum duration regulation-compliant schedule. We proposed a mixed integer programming formulation that uses conditioned time-window constraints to model the parking availability at parking facilities and moving window constraints to model the hours-of-service (HOS) constraints for long trips. Simulation results illustrate that schedules

calculated without accounting for parking availability are often infeasible. Although parking constraints increased trip duration in some scenarios, these scenarios also showed lower feasibility rates when ignoring parking information.

We followed by extending the TDSP under parking availability constraints to include path planning. We proposed a resource-constrained shortest path formulation that uses a resource vector to keep track of the HOS and time constraints. The problem is solver over an auxiliary network that explicitly models the different activities available to drivers and how they affect each regulation constraint. We proposed a tailored label-correcting algorithm that solves the problem to optimality. Computational experiments showed that parking conditions could significantly affect the route choice, illustrating the importance of accounting for parking availability information early in the planning process. We also simulated the potential costs of disregarding parking information under different parking shortage severity levels and how they compare to the cost increase caused by imposing parking restrictions. The results vary greatly depending on the available routes' quality, the parking shortage severity, and the expected cost of illegal parking. The results underline the importance of including parking information as early as possible to increase the quantity and quality of available routes and schedules. In addition, it also elicits the importance of further research on estimating the potential costs and risks of illegal truck parking.

Finally, we extended the resource-constrained shortest path formulation to the case of battery-electric trucks (BETs). We studied the impact of coordinating rest and recharge needs on BETs' performance and its comparison to diesel trucks. Computational experiments were used to estimate the effects of different levels of charging and parking infrastructure. Although BETs generate only a small fraction of diesel trucks' CO2 emissions, BETs require longer trip durations in most scenarios. However, this gap in trip duration depends on battery capacity, charging infrastructure (power of regular and fast chargers), and parking/charging facilities' availability (regular chargers' time-windows, and fast chargers number and wait time). A common concern regarding the utilization of BETs for long-haul trucking is the infrastructure required to quickly charge large batteries, reducing the disparity to diesel trucks' refueling time. Nevertheless, our experiments show that, although fast-chargers can significantly reduce trip duration in many scenarios, trip duration is even more sensitive to the power and availability of the regular chargers used for long (overnight) rests. It is important to note that these results do not mean that fast chargers are without benefit. The advantages of particular infrastructure decisions will vary for each case. Our experiments illustrate that, while trying to have BETs operating in similar itineraries to current diesel trucks (e.g., using fast chargers to reduce recharging time) might be the instinctive way to approach truck electrification, it is not the only one. It is likely not the best approach either.

In this project, we exposed the importance of using truck parking availability information during planning, and proposed methods to do so. Besides helping individual truck drivers with trip planning, the methods developed in this project can simulate different scenarios and aid policymakers in estimating the impacts of infrastructure investment decisions.

References

Abdel-Aty, Mohamed, Jeremy Dilmore, and Albinder Dhindsa. 2006. “Evaluation of Variable Speed Limits for Real-Time Freeway Safety Improvement.” *Accident Analysis & Prevention* 38 (2): 335–45.

Alessandri, A, A Di Febbraro, A Ferrara, and E Punta. 1998. “Optimal Control of Freeways via Speed Signalling and Ramp Metering.” *Control Engineering Practice* 6 (6): 771–80.

Altan O., Wu, G., Barth, M., Boriboonsomsin, K., and Stark, J. (2017). “GlidePath: Eco-Friendly Automated Approach and Departure at Signalized Intersections.” *IEEE Transactions on Intelligent Vehicles*, 2(4), 266-277.

Andersson, Joel. 2013. “A General-Purpose Software Framework for Dynamic Optimization.” {P}h{D} thesis, Department of Electrical Engineering (ESAT/SCD); Optimization in Engineering Center, Kasteelpark Arenberg 10, 3001-Heverlee, Belgium: Arenberg Doctoral School, KU Leuven.

Baldi, S., I. Michailidis, E. B. Kosmatopoulos, A. Papachristodoulou, and P. A. Ioannou. 2014. “Convex Design Control for Practical Nonlinear Systems.” *IEEE Transactions on Automatic Control* 59 (7): 1692–1705. https://doi.org/10.1109/TAC.2014.2309271.

Banks, James H. 1991. “The Two-Capacity Phenomenon: Some Theoretical Issues.” *Transportation Research Record*, no. 1320.

Barth, M., and K. Boriboonsomsin (2009). Energy and emissions impacts of a freeway-based dynamic eco-driving system. *Transportation Research Part D: Transport and Environment*, 14(6): 400-410.

Baskar, Lakshmi Dhevi, Bart De Schutter, and Hans Hellendoorn. 2008. “Model-Based Predictive Traffic Control for Intelligent Vehicles: Dynamic Speed Limits and Dynamic Lane Allocation.” In *Intelligent Vehicles Symposium, 2008 IEEE*, 174–79. IEEE.

Bishop, G. A., Brent G. Schuchmann, Donald H. Stedman, and Douglas R. Lawson (2012). Emission Changes Resulting from the San Pedro Bay, California Ports Truck Retirement Program. *Environmental Science & Technology*, 46(1): 551-558.

Bock, Hans Georg, and Karl-Josef Plitt. 1984. “A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems.” In *Proceedings of the IFAC World Congress*.

Borek, J., B. Groelke, C. Earnhardt and C. Vermillion (2020). "Economic Optimal Control for Minimizing Fuel Consumption of Heavy-Duty Trucks in a Highway Environment," in IEEE Transactions on Control Systems Technology, vol. 28, no. 5, pp. 1652-1664.

Boriboonsomsin, K., M. J. Barth, W. Zhu, and A. Vu (2012). Eco-routing navigation system based on multisource historical and real-time traffic information. *IEEE Transactions on Intelligent Transportation Systems*, 13(4): 1694-1704.

Boriboonsomsin, K., Dean, J., and Barth, M. (2014). Examination of attributes and value of ecologically friendly route choices. *Transportation Research Record*, 2427, 13-25.

Boriboonsomsin, K. (2015). Reducing the Carbon Footprint of Freight Movement through Eco-Driving Programs for Heavy-Duty Trucks. White paper from the National Center for Sustainable Transportation, June.

Brunekreef, B., N. A. Janssen, J. de Hartog, H. Harssema, M. Knape, and P. van Vliet (1997). Air pollution from truck traffic and lung function in children living near motorways. *Epidemiology*, 298-303.

Butakov, Vadim A, and Petros Ioannou. 2015. "Personalized Driver/Vehicle Lane Change Models for ADAS." *IEEE Transactions on Vehicular Technology* 64 (10): 4422–31.

California Air Resources Board (2008). Heavy-Duty Diesel Engine Emission Standard Regulations. https://ww3.arb.ca.gov/regact/hdde2007/hdde2007.htm.

California Air Resources Board (2021). Heavy-Duty Low NOx. https://ww2.arb.ca.gov/our-work/programs/heavy-duty-low-nox.

Caligaris, Carlo, Simona Sacone, and Silvia Siri. 2007. "Optimal Ramp Metering and Variable Speed Signs for Multiclass Freeway Traffic." In *2007 European Control Conference*, 1780–85. IEEE.

Caltrans. 2016. "RAMP METERING DESIGN MANUAL." Sacramento, United States: California Department of Transportation.

Caltrans District 7. 2006. "Ramp Metering Annual Report." Los Angeles; Ventura Counties.

Carlson, Rodrigo Castelan, Ioannis Papamichail, and Markos Papageorgiou. 2011. "Local Feedback-Based Mainstream Traffic Flow Control on Motorways Using Variable Speed Limits." *IEEE Transactions on Intelligent Transportation Systems* 12 (4): 1261–76.

Carlson, Rodrigo C, Ioannis Papamichail, and Markos Papageorgiou. 2013. "Comparison of Local Feedback Controllers for the Mainstream Traffic Flow on Freeways Using Variable Speed Limits." *Journal of Intelligent Transportation Systems* 17 (4): 268–81.

Chang, Hwan, Yun Wang, Jianlong Zhang, and Petros A Ioannou. 2007. "An Integrated Roadway Controller and Its Evaluation by Microscopic Simulator VISSIM." In *Control Conference (ECC), 2007 European*, 2436–41. IEEE.

City of Los Angeles Harbor Department (2021). The Port of Los Angeles Eco-FRATIS Drayage Truck Efficiency Project. Final project report, prepared for California Energy Commission.

Coogan, Samuel, and Murat Arcak. 2016. "Stability of Traffic Flow Networks with a Polytree Topology." *Automatica* 66: 246–53.

Csikós, Alfréd, and Balázs Kulcsár. 2017. "Variable Speed Limit Design Based on Mode Dependent Cell Transmission Model." *Transportation Research Part C: Emerging Technologies* 85: 429–50.

Daganzo, Carlos F. 1994. "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory." *Transportation Research Part B: Methodological* 28 (4): 269–87.

Davis, S. C., Diegel, S. W., and Boundy, R. G. (2010). *Transportation Energy Data Book: Edition 29*. Report No. ORNL-6985, Oak Ridge National Laboratory.

Davis, S. C. and Boundy, R. G. (2020). *Transportation Energy Data Book: Edition 38*. Report No. ORNL/TM-2019/1333, Oak Ridge National Laboratory.

De León, K. (2012). Senate Bill No. 535. California Global Warming Solutions Act of 2006: Greenhouse Gas Reduction Fund. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201120120SB535.

EPA. 2014. "Motor Vehicle Emission Simulator (MOVES) User Guide." *US Environmental Protection Agency*.

European Commission (2021). "eCoMove – Cooperative Mobility Systems and Services for Energy Efficiency", http://www.ecomove-project.eu/.

Gao, Kejia. 2012. "Multi-Objective Traffic Management for Livability." PhD thesis, MS thesis, TU Delft, Delft, The Netherlands.

Garcia, C. (2017). Assembly Bill No. 617. Nonvehicular air pollution: criteria air pollutants and toxic air contaminants. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180AB617.

Gomes, Gabriel, Roberto Horowitz, Alex A Kurzhanskiy, Pravin Varaiya, and Jaimyoung Kwon. 2008. "Behavior of the Cell Transmission Model and Effectiveness of Ramp Metering." *Transportation Research Part C: Emerging Technologies* 16 (4): 485–513.

Gomez, J. (2016). Assembly Bill No. 1550. Greenhouse gases: investment plan: disadvantaged communities. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB1550.

Hadiuzzaman, Md, and Tony Z Qiu. 2013. "Cell Transmission Model Based Variable Speed Limit Control for Freeways." *Canadian Journal of Civil Engineering* 40 (1): 46–56.

Hadiuzzaman, Md, Tony Z Qiu, and Xiao-Yun Lu. 2012. "Variable Speed Limit Control Design for Relieving Congestion Caused by Active Bottlenecks." *Journal of Transportation Engineering*.

Hall, Fred L, and Kwaku Agyemang-Duah. 1991. "Freeway Capacity Drop and the Definition of Capacity." *Transportation Research Record*, no. 1320.

Hao, P., Wu, G., Boriboonsomsin, K., and Barth, M. (2019). "Eco-approach and departure application for actuated signals in real-world traffic." *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 30-40.

Hao, P., K. Boriboonsomsin, C. Wang, G. Wu, and M. Barth (2021). "Connected eco-approach and departure (EAD) system for diesel trucks", SAE International Journal of Commercial Vehicles 2(14).

Hegyi, A, SP Hoogendoorn, M Schreuder, H Stoelhorst, and F Viti. 2008. "SPECIALIST: A Dynamic Speed Limit Control Algorithm Based on Shock Wave Theory." In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, 827–32. IEEE.

Hegyi, Andreas, Bart De Schutter, and Hans Hellendoorn. 2005. “Model Predictive Control for Optimal Coordination of Ramp Metering and Variable Speed Limits.” *Transportation Research Part C: Emerging Technologies* 13 (3): 185–209.

Hooper A. and Murray, D. (2018). An Analysis of the Operational Costs of Trucking: 2018 Update. American Transportation Research Institute, October.

Huang, Y., Elvin C.Y. Ng, John L. Zhou, Nic C. Surawski, Edward F.C. Chan, Guang Hong (2018). Eco-driving technology for sustainable road transport: A review. *Renewable and Sustainable Energy Reviews*, 93: 596-609.

Ioannou, Petros A, and Jing Sun. 2012. *Robust Adaptive Control*. Dover Publications, Mineola, New York.

Ioannou, Petros, Yun Wang, Afshin Abadi, and Vadim Butakov. 2012. “Dynamic Variable Speed Limit Control: Design, Analysis and Benefits.”

Iordanidou, Georgia-Roumpini, Claudio Roncoli, Ioannis Papamichail, and Markos Papageorgiou. 2015. “Feedback-Based Mainstream Traffic Flow Control for Multiple Bottlenecks on Motorways.” *Intelligent Transportation Systems, IEEE Transactions on* 16 (2): 610–21.

Jha, Mithilesh, David Cuneo, and Moshe Ben-Akiva. 1999. “Evaluation of Freeway Lane Control for Incident Management.” *Journal of Transportation Engineering* 125 (6): 495–501.

Jiang, Y., Jiacheng Yang, David Cocker, Georgios Karavalakis, Kent C. Johnson, Thomas D. Durbin (2018). Characterizing emission rates of regulated pollutants from model year 2012+ heavy-duty diesel vehicles equipped with DPF and SCR systems. *Science of The Total Environment*, 619–620: 765-771.

Jin, Hui-Yu, and Wen-Long Jin. 2015. “Control of a Lane-Drop Bottleneck Through Variable Speed Limits.” *Transportation Research Part C: Emerging Technologies* 58: 568–84.

Jin, Wen-Long. 2013. “A Multi-Commodity Lighthill–Whitham–Richards Model of Lane-Changing Traffic Flow.” *Transportation Research Part B: Methodological* 57: 361–77.

Karafyllis, Iasson, and Markos Papageorgiou. 2015. “Global Exponential Stability for Discrete-Time Networks with Applications to Traffic Networks.” *IEEE Transactions on Control of Network Systems* 2 (1): 68–77.

Kejun, Long, Yun Meiping, Zheng Jianlong, and Yang Xiaoguang. 2008. “Model Predictive Control for Variable Speed Limit in Freeway Work Zone.” In *Control Conference, 2008. CCC 2008. 27th Chinese*, 488–93. IEEE.

Khalil, Hassan K, and JW Grizzle. 1996. *Nonlinear Systems*. Vol. 3. Prentice hall New Jersey.

Khondaker, Bidoura, and Lina Kattan. 2015. “Variable Speed Limit: A Microscopic Analysis in a Connected Vehicle Environment.” *Transportation Research Part C: Emerging Technologies* 58: 146–59.

Kontorinaki, Maria, Iasson Karafyllis, and Markos Papageorgiou. 2017. “Global Exponential Stabilisation of Acyclic Traffic Networks.” *International Journal of Control*, 1–21.

Kontorinaki, Maria, Anastasia Spiliopoulou, Claudio Roncoli, and Markos Papageorgiou. 2016. “Capacity Drop in First-Order Traffic Flow Models: Overview and Real-Data Validation.” In *Transportation Research Board 95th Annual Meeting*. 16-3541.

Kozawa, K.H., Fruin, S.A. and Winer, A.M. (2009). Near-road air pollution impacts of goods movement in communities adjacent to the Ports of Los Angeles and Long Beach. *Atmospheric Environment*, 43(18): 2960-2970.

Laval, Jorge A, and Carlos F Daganzo. 2006. “Lane-Changing in Traffic Streams.” *Transportation Research Part B: Methodological* 40 (3): 251–64.

Lebacque, J. 2003. “Two-Phase Bounded-Acceleration Traffic Flow Model: Analytical Solutions and Applications.” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1852: 220–30.

Lesage, J. (2013). “Volvo finds eco-driving improves fuel economy 10-15% in trucks”, https://www.autoblog.com/2013/12/16/volvo-finds-eco-driving-improves-fuel-economy-10-15-percent-trucks/, Accessed March, 2021.

Li, Zhibin, Pan Liu, Wei Wang, and Chengcheng Xu. 2014. “Development of a Control Strategy of Variable Speed Limits to Reduce Rear-End Collision Risks Near Freeway Recurrent Bottlenecks.” *Intelligent Transportation Systems, IEEE Transactions on* 15 (2): 866–77.

Lovisari, Enrico, Giacomo Como, Anders Rantzer, and Ketan Savla. 2014. “Stability Analysis and Control Synthesis for Dynamical Transportation Networks.” *arXiv Preprint arXiv:1410.5956*.

Lu, Xiao-Yun, Tony Z Qiu, Pravin Varaiya, Roberto Horowitz, and Steven E Shladover. 2010. “Combining Variable Speed Limits with Ramp Metering for Freeway Traffic Control.” In *Proceedings of the 2010 American Control Conference*, 2266–71. IEEE.

Lu, Xiao-Yun, and Steven E Shladover. 2014. “Review of Variable Speed Limits and Advisories.” *Transportation Research Record: Journal of the Transportation Research Board* 2423 (1): 15–23.

Lu, Xiao-Yun, Pravin Varaiya, Roberto Horowitz, Dongyan Su, and Steven Shladover. 2011. “Novel Freeway Traffic Control with Variable Speed Limit and Coordinated Ramp Metering.” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2229: 55–65.

Lu, Xiao-Yun, Pravin Varaiya, Roberto Horowitz, Dongyan Su, and Steven E Shladover. 2010. “A New Approach for Combined Freeway Variable Speed Limits and Coordinated Ramp Metering.” In *2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 491–98. IEEE.

Luo, Y, Y. Xiang, K. Cao, and K. Li (2016). ‘A dynamic automated lane change maneuver based on vehicle-to-vehicle communication,’’ *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 87–102.

McCaffery, C., Hanwei Zhu, Tianbo Tang, Chengguo Li, Georgios Karavalakis, Sam Cao, Adewale Oshinuga, Andrew Burnette, Kent C. Johnson, Thomas D. Durbin (2021). Real-world NOx emissions from heavy-duty diesel, natural gas, and diesel hybrid electric vehicles of different vocations on California roadways. *Science of The Total Environment*, 784: 147224.

Milanés, V., S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe and M. Nakamura (2014). "Cooperative Adaptive Cruise Control in Real Traffic Situations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 296-305.

Misra, C., Collins, J. F., Herner, J. D., Sax, T., Krishnamurthy, M., Sobieralksi, W., Burntizki, M., Chernich, D. (2013). In-use NOx emissions frommodel year 2010 and 2011 heavy-duty diesel engines equipped with aftertreatment devices. *Environmental Science & Technology*, 47 (14), 7892–7898.

Müller, Eduardo Rauh, Rodrigo Castelan Carlson, Werner Kraus, and Markos Papageorgiou. 2015. "Microsimulation Analysis of Practical Aspects of Traffic Control with Variable Speed Limits." *IEEE Transactions on Intelligent Transportation Systems* 16 (1): 512–23.

Muralidharan, Ajith, and Roberto Horowitz. 2015. "Computationally Efficient Model Predictive Control of Freeway Networks." *Transportation Research Part C: Emerging Technologies* 58: 532–53.

Oh, Heung-Un, and Virginia P Sisiopiku. 2001. "A Modified ALINEA Ramp Metering Model." In *Transportation Research Board 80th Annual Meeting*.

Papageorgiou, Markos, Habib Hadj-Salem, and Jean-Marc Blosseville. 1991. "ALINEA: A Local Feedback Control Law for on-Ramp Metering." *Transportation Research Record*, no. 1320.

Papageorgiou, Markos, Habib Hadj-Salem, and F Middelham. 1997. "ALINEA Local Ramp Metering: Summary of Field Results." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1603: 90–98.

Papageorgiou, Markos, Elias Kosmatopoulos, and Ioannis Papamichail. 2008. "Effects of Variable Speed Limits on Motorway Traffic Flow." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2047: 37–48.

Papageorgiou, M., and A. Kotsialos. 2002. "Freeway Ramp Metering: An Overview." *IEEE Transactions on Intelligent Transportation Systems* 3 (4): 271–81.

Papamichail, Ioannis, Katerina Kampitaki, Markos Papageorgiou, and Albert Messmer. 2008. "Integrated Ramp Metering and Variable Speed Limit Control of Motorway Traffic Flow." *IFAC Proceedings Volumes* 41 (2): 14084–89.

Port of Los Angeles (2018). Air Emissions Inventory. https://www.portoflosangeles.org/environment/air-quality/air-emissions-inventory.

Reis H, Reis C, Sharip A, Reis W, Zhao Y, Sinclair R, Beeson L. (2018). Diesel exhaust exposure, its multi-system effects, and the effect of new technology diesel exhaust. *Environ Int*, 114: 252-265.

Robinson RK, Birrell MA, Adcock JJ, Wortley MA, Dubuis ED, Chen S, McGilvery CM, Hu S, Shaffer MSP, Bonvini SJ, Maher SA, Mudway IS, Porter AE, Carlsten C, Tetley TD, Belvisi MG (2018). Mechanistic link between diesel exhaust particles and respiratory reflexes. *J Allergy Clin Immunol*, 141(3): 1074-1084.

Rodriguez, M. A. and H. K. Fathy (2018). “Speed Trajectory Optimization for a Heavy-Duty Truck Traversing Multiple Signalized Intersections: A Dynamic Programming Study”, 2018 IEEE Conference on Control Technology and Applications (CCTA).

Roncoli, Claudio, Markos Papageorgiou, and Ioannis Papamichail. 2015. “Traffic Flow Optimisation in Presence of Vehicle Automation and Communication Systems–Part I: A First-Order Multi-Lane Model for Motorway Traffic.” *Transportation Research Part C: Emerging Technologies* 57: 241–59.

Roncoli, C., I. Papamichail, and M. Papageorgiou. 2014. “Model Predictive Control for Multi-Lane Motorways in Presence of VACS.” In *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, 501–7. https://doi.org/10.1109/ITSC.2014.6957739.

Scariza, Joseph R. 2003. “Evaluation of Coordinated and Local Ramp Metering Algorithms Using Microscopic Traffic Simulation.” PhD thesis, Massachusetts Institute of Technology.

Schaefer, L, James Upchurch, and SA Ashur. 1998. “An Evaluation of Freeway Lane Control Signing Using Computer Simulation.” *Mathematical and Computer Modelling* 27 (9): 177–87.

Scora, G., Boriboonsomsin, K. and Barth, M. (2015). Value of eco-friendly route choice for heavy-duty trucks. *Research in Transportation Economics*, 52:3-14.

Scora, G., Boriboonsomsin, K., Vu, A., and Barth, M. (2019). Effects of Route Selection on NOx Emissions from Heavy-duty Diesel Trucks. 29th CRC Real-World Vehicle Emissions Workshop, Long Beach, CA, March 10-13.

Smaragdis, Emmanouil, and Markos Papageorgiou. 2003. “Series of New Local Ramp Metering Strategies: Emmanouil Smaragdis and Markos Papageorgiou.” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1856: 74–86.

South Coast Air Quality Management District (2017). Final 2016 Air Quality Management Plan. https://www.aqmd.gov/home/air-quality/clean-air-plans/air-quality-mgt-plan/final-2016-aqmp.

Srivastava, Anupam, and Nikolas Geroliminis. 2013. “Empirical Observations of Capacity Drop in Freeway Merges with Ramp Control and Integration in a First-Order Model.” *Transportation Research Part C: Emerging Technologies* 30: 161–77.

Srivastava, Anupam, and Wenlong Jin. 2016. “A Lane Changing Cell Transmission Model for Modeling Capacity Drop at Lane Drop Bottlenecks.” In *Transportation Research Board 95th Annual Meeting*, 16–5452. TRB.

Srivastava, Anupam, Wen-Long Jin, and Jean-Patrick Lebacque. 2015. “A Modified Cell Transmission Model with Realistic Queue Discharge Features at Signalized Intersections.” *Transportation Research Part B: Methodological* 81: 302–15.

Torne Santos, Josep Maria, Dulce Rosas, and Francesc Soriguera. 2011. “Evaluation of Speed Limit Management on c-32 Highway Access to Barcelona.” In *Transportation Research Board 90th Annual Meeting*. 11-2397.

Transportation, California Department of. 2015. “Caltrans Performance Measurement System (PeMS).” 2015. http://pems.dot.ca.gov/.

U.S. Department of Transportation (2014) “Applications for the Environment: Real-Time Information Synthesis (AERIS)”, https://www.its.dot.gov/research_archives/aeris/index.htm.

U.S. Energy Information Administration (2020). 2020 Annual Energy Outlook. https://www.eia.gov/outlooks/aeo/.

U.S. Environmental Protection Agency (2015). “MOVES (Motor Vehicle Emission Simulator)”. http://www.epa.gov/oms/models/moves/.

Van den Hoogen, Erick, and Stef Smulders. 1994. “Control by Variable Speed Signs: Results of the Dutch Experiment.” In *Road Traffic Monitoring and Control, 1994., Seventh International Conference on*, 145–49. IET.

Wächter, Andreas, and Lorenz T Biegler. 2006. “On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming.” *Mathematical Programming* 106 (1): 25–57.

Wang, J., A. Elbery, and H. A. Rakha (2019a). A real-time vehicle-specific eco-routing model for on-board navigation applications capturing transient vehicle behavior. *Transportation Research Part C: Emerging Technologies*, 104: 1-21.

Wang, Yun, and Petros Ioannou. 2011a. “Dynamic Variable Speed Limit Control: Design, Analysis and Benefits.” PhD thesis, University of Southern California.

Wang, Z., Hsu, Y.-P., Vu, A., Caballero, F., Hao, P., Wu, G., Boriboonsomsin, K., Barth, M., Kailas, A., Amar, P., Garmon, E., and Tangula, S. (2019b). “Early findings from field trials of heavy-duty truck connected eco-driving system.” Proceedings of the 22nd International IEEE Conference on Intelligent Transportation Systems, Auckland, New Zealand, October 27-30.

Wilson SJ, Miller MR, Newby DE (2018). Effects of Diesel Exhaust on Cardiovascular Function and Oxidative Stress. *Antioxid Redox Signal*, 28(9): 819-836.

Wright, Matthew, Gabriel Gomes, Roberto Horowitz, and Alex A Kurzhanskiy. 2015. “A New Model for Multi-Commodity Macroscopic Modeling of Complex Traffic Networks.” *arXiv Preprint arXiv:1509.04995*.

Xia, H., Boriboonsomsin, K., and Barth, M. (2013). “Dynamic ECO-driving for signalized arterial corridors and its indirect network-wide energy/emissions benefits.” *Journal of Intelligent Transportation Systems*, 17(1), 31-41.

Xie, Y., H. Zhang, N. H. Gartner, and T. Arsava (2017). ‘‘Collaborative merging strategy for freeway ramp operations in a connected and autonomous vehicles environment,’’ *J. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 136–147.

Zhang, Jianlong, Hwan Chang, and Petros A Ioannou. 2006. “A Simple Roadway Control System for Freeway Traffic.” In *2006 American Control Conference*, 6–pp. IEEE.

Zhang, Yihang, and Petros A Ioannou. 2015. “Combined Variable Speed Limit and Lane Change Control for Truck-Dominant Highway Segment.” In *2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, 1163–68. IEEE.

———. 2017a. “Combined Variable Speed Limit and Lane Change Control for Highway Traffic.” *IEEE Transactions on Intelligent Transportation Systems* 18 (7): 1812–23.

———. 2017b. “Combined Variable Speed Limit and Lane Change Control for Highway Traffic.” *IEEE Transactions on Intelligent Transportation Systems* 18 (7): 1812–23.

———. 2018. “Stability Analysis and Variable Speed Limit Control of a Traffic Flow Model.” *Transportation Research Part B: Methodological* 118: 31–65.

Zhang, Yihang, and Petros A. Ioannou. 2016. “Environmental Impact of Combined Variable Speed Limit and Lane Change Control: A Comparison of MOVES and CMEM Model.” *IFAC-PapersOnLine* 49 (3): 323–28. https://doi.org/http://dx.doi.org/10.1016/j.ifacol.2016.07.054.

Zhang, Yihang, Isik Ilber Sirmatel, Faisal Alasiri, Petros A Ioannou, and Nikolas Geroliminis. 2018. “Comparison of Feedback Linearization and Model Predictive Techniques for Variable Speed Limit Control.” In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 3000–3005. IEEE.

Almeida Araujo Vital, Filipe de, Petros Ioannou, and Arti Gupta. 2020. “Survey on Intelligent Truck Parking: Issues and Approaches.” *IEEE Intelligent Transportation Systems Magazine*. https://doi.org/10.1109/MITS.2019.2926259.

American Transportation Research Institute. 2018. “MAASTO Truck Parking Survey Analysis.” Arlington, VA.

———. 2019. “Critical Issues in the Trucking Industry – 2019.” Arlington, VA.

Archetti, C., and M. Savelsbergh. 2009. “The Trip Scheduling Problem.” *Transportation Science* 43 (4): 417–31. https://doi.org/10.1287/trsc.1090.0278.

Argonne National Laboratory. 2020. “GREET.Net 2020.”

Bayraktar, Mehmet Emre, Farrukh Arif, Halit Ozen, and Gorm Tuxen. 2015. “Smart Parking-Management System for Commercial Vehicle Parking at Public Rest Areas.” *Journal of Transportation Engineering* 141 (5): 04014094. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000756.

Bektaş, Tolga, and Gilbert Laporte. 2011. “The Pollution-Routing Problem.” *Transportation Research Part B: Methodological* 45 (8): 1232–50. https://doi.org/10.1016/j.trb.2011.02.004.

Bertsekas, Dimitri. 2017. *Dynamic Programming and Optimal Control, Vol. I*. 4th Editio. Athena Scientific.

Bertsekas, Dimitri P. 1993. “A Simple and Fast Label Correcting Algorithm for Shortest Paths.” *Networks* 23 (8): 703–9. https://doi.org/10.1002/net.3230230808.

Boris, Caroline, and Rebecca Brewster. 2018. “A Comparative Analysis of Truck Parking Travel Diary Data.” *Transportation Research Record: Journal of the Transportation Research Board* 2672 (9): 242–48. https://doi.org/10.1177/0361198118775869.

Çabukoglu, Emir, Gil Georges, Lukas Küng, Giacomo Pareschi, and Konstantinos Boulouchos. 2018. “Battery Electric Propulsion: An Option for Heavy-Duty Vehicles? Results from a Swiss Case-Study.” *Transportation Research Part C: Emerging Technologies*. https://doi.org/10.1016/j.trc.2018.01.013.

Cambridge Systematics. 2019. “Nevada Truck Parking Implementation Plan - Task 5: Draft Recommendations.” Los Angeles, CA.

Code of Federal Regulations. n.d. “Title 49 Part 395 (49CFR3.395) (2013), Hours of Service of Drivers.”

Costa, Luciano, Claudio Contardo, and Guy Desaulniers. 2019. “Exact Branch-Price-and-Cut Algorithms for Vehicle Routing.” *Transportation Science* 53 (4): 946–85. https://doi.org/10.1287/trsc.2018.0878.

Drexl, Michael, and Eric Prescott-Gagnon. 2010. “Labelling Algorithms for the Elementary Shortest Path Problem with Resource Constraints Considering EU Drivers’ Rules.” *Logistics Research* 2 (2): 79–96. https://doi.org/10.1007/s12159-010-0022-9.

Earl, Thomas, Lucien Mathieu, Stef Cornelis, Samuel Kenny, Carlos Calvo Ambel, and James Nix. 2018. “Analysis of Long Haul Battery Electric Trucks in EU.” In *8th Commercial Vehicle Workshop*, 17–18. Graz.

Federal Motor Carrier Safety Administration. 2021. “Title 49 Part 395, HOURS OF SERVICE OF DRIVERS.” Electronic Code of Federal Regulations. 2021. https://ecfr.federalregister.gov/current/title-49/subtitle-B/chapter-III/subchapter-B/part-395.

Gaddy, Alexis Christine, Sarah Hernandez, and Sarah Nurre. 2018. “Incorporating Truck Parking and Hours of Service (HOS) into a Truck Routing Heuristic.” In *Transportation Research Board 97th Annual Meeting*. Washington DC, United States.

Gao, Zhiming, Zhenhong Lin, and Oscar Franzese. 2017. “The Energy Consumption and Cost Savings of Truck Electrification for Heavy-Duty Vehicle Applications.” *Transportation Research Record: Journal of the Transportation Research Board* 2628 (1): 99–109. https://doi.org/10.3141/2628-11.

Goel, A. 2010. “Truck Driver Scheduling in the European Union.” *Transportation Science* 44 (4): 429–41. https://doi.org/10.1287/trsc.1100.0330.

Goel, Asvin. 2012. “The Minimum Duration Truck Driver Scheduling Problem.” *EURO Journal on Transportation and Logistics* 1 (4): 285–306. https://doi.org/10.1016/j.cor.2011.12.016.

Goel, Asvin, and Stefan Irnich. 2017. “An Exact Method for Vehicle Routing and Truck Driver Scheduling Problems.” *Transportation Science* 51 (2): 737–54. https://doi.org/10.1287/trsc.2016.0678.

Goel, Asvin, and Leendert Kok. 2012. “Truck Driver Scheduling in the United States.” *Transportation Science* 46 (3): 317–26. https://doi.org/10.1287/trsc.1110.0382.

Goel, Asvin, and Louis-Martin Rousseau. 2011. “Truck Driver Scheduling in Canada.” *Journal of Scheduling* 15 (6): 783–99. https://doi.org/10.1007/s10951-011-0249-6.

Goel, Asvin, and Thibaut Vidal. 2014. “Hours of Service Regulations in Road Freight Transport: An Optimization-Based International Assessment.” *Transportation Science* 48 (3): 391–412. https://doi.org/10.1287/trsc.2013.0477.

Hagemann, Garrett, Kent Hymel, Adam Klauber, D. Lee, George Noel, David Pace, and Catherine Taylor. 2013. “Delay and Environmental Costs of Truck Crashes,” no. March: 132.

Hernández, Salvador, and Jason Anderson. 2017. “Truck Parking: An Emerging Safety Hazard to Highway Users.” Corvallis, OR.

Horváth, Markó, and Tamás Kis. 2016. “Solving Resource Constrained Shortest Path Problems with LP-Based Methods.” *Computers and Operations Research* 73: 150–64. https://doi.org/10.1016/j.cor.2016.04.013.

Irnich, Stefan, and Guy Desaulniers. 2005. “Shortest Path Problems with Resource Constraints.” In *Column Generation*, edited by Guy Desaulniers, Jacques Desrosiers, and Marius M. Solomon, 33–65. Springer US.

Judicial Council of California. 2019. *Uniform Bail and Penalty Schedules*.

Kimley Horn. 2015. “Virginia Truck Parking Study.”

Koç, Çağrı, Tolga Bektaş, Ola Jabali, and Gilbert Laporte. 2016. “A Comparison of Three Idling Options in Long-Haul Truck Scheduling.” *Transportation Research Part B: Methodological* 93 (November): 631–47. https://doi.org/10.1016/j.trb.2016.08.006.

Koç, Çağrı, Ola Jabali, and Gilbert Laporte. 2018. “Long-Haul Vehicle Routing and Scheduling with Idling Options.” *Journal of the Operational Research Society* 69 (2): 235–46. https://doi.org/10.1057/s41274-017-0202-y.

Kok, A. L., E. W. Hans, and J. M J Schutten. 2011. “Optimizing Departure Times in Vehicle Routes.” *European Journal of Operational Research* 210 (3): 579–87. https://doi.org/10.1016/j.ejor.2010.10.017.

Kok, A. L., E. W. Hans, J. M J Schutten, and W. H M Zijm. 2010. “A Dynamic Programming Heuristic for Vehicle Routing with Time-Dependent Travel Times and Required Breaks.” *Flexible Services and Manufacturing Journal* 22 (1–2): 83–108. https://doi.org/10.1007/s10696-011-9077-4.

Kok, A. L., C. M. Meyer, H. Kopfer, and J. M. J. Schutten. 2010. “A Dynamic Programming Heuristic for the Vehicle Routing Problem with Time Windows and European Community Social Legislation.” *Transportation Science* 44 (4): 442–54. https://doi.org/10.1287/trsc.1100.0331.

Liimatainen, Heikki, Oscar van Vliet, and David Aplyn. 2019. “The Potential of Electric Trucks – An International Commodity-Level Analysis.” *Applied Energy*. https://doi.org/10.1016/j.apenergy.2018.12.017.

Lin, Jane, Wei Zhou, and Ouri Wolfson. 2016. “Electric Vehicle Routing Problem.” In *Transportation Research Procedia*. https://doi.org/10.1016/j.trpro.2016.02.007.

Lin, Trista, Herve Rivano, and Frederic Le Mouel. 2017. “A Survey of Smart Parking Solutions.” *IEEE Transactions on Intelligent Transportation Systems*, 1–25. https://doi.org/10.1109/TITS.2017.2685143.

Lozano, Leonardo, Daniel Duque, and Andrés L. Medaglia. 2016. “An Exact Algorithm for the Elementary Shortest Path Problem with Resource Constraints.” *Transportation Science* 50 (1): 348–57. https://doi.org/10.1287/trsc.2014.0582.

Lysiak, Fran Matso. 2019. “Special Report: Trucking and Insurance.” *Transport Topics*, September 2019.

Mareev, Ivan, Jan Becker, and Dirk Uwe Sauer. 2018. “Battery Dimensioning and Life Cycle Costs Analysis for a Heavy-Duty Truck Considering the Requirements of Long-Haul Transportation.” *Energies* 11 (1). https://doi.org/10.3390/en11010055.

Martin, Elliot W, and Susan A Shaheen. 2013. “Truck Parking and Traffic on I-5 in California: Analysis of a Clipboard Survey and Annual Average Daily Traffic Data.” In *TRB 92nd Annual Meeting Compendium of Papers*. Washington DC.

Mayerle, Sérgio Fernando, Daiane Maria De Genaro Chiroli, João Neiva de Figueiredo, and Hidelbrando Ferreira Rodrigues. 2020. “The Long-Haul Full-Load Vehicle Routing and Truck Driver Scheduling Problem with Intermediate Stops: An Economic Impact Evaluation of Brazilian Policy.” *Transportation Research Part A: Policy and Practice* 140 (December 2018): 36–51. https://doi.org/10.1016/j.tra.2020.07.021.

Monteiro, Fernando V., and Petros Ioannou. 2018. “On-Street Parking Prediction Using Real-Time Data.” In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018-Novem:2478–83. IEEE. https://doi.org/10.1109/ITSC.2018.8569921.

Morris, Ted, Nikos Pananikolopolous, Travis Henderson, and Vassilios Morellas. 2018. “A Real-Time Truck Availability System for the State of Wisconsin.” Minneapolis, MN.

Murray, Dan, and Seth Glidewell. 2019. “An Analysis of the Operational Costs of Trucking: 2019 Update.” Minneapolis, MN.

NCDOT. 2017. “North Carolina Statewide Multimodal Freight Plan: Truck Parking Study.”

Palaniappan, Meena, Diana Wu, and Jacki Kohleriter. 2003. “Clearing the Air: Reducing Diesel Pollution in West Oakland.” Oakland, CA.

Phelan, Thomas, Vince Mantero, Jeff Purdy, and Thomas Kearney. 2016. “National Coalition on Truck Parking Activity Report.” Washington, DC.

Pugliese, Luigi Di Puglia, and Francesca Guerriero. 2013. “A Survey of Resource Constrained Shortest Path Problems: Exact Solution Approaches.” *Networks* 62 (3): 183–200. https://doi.org/10.1002/net.21511.

Rajabioun, Tooraj, and Petros Ioannou. 2015. “On-Street and off-Street Parking Availability Prediction Using Multivariate Spatiotemporal Models.” *IEEE Transactions on Intelligent Transportation Systems* 16 (5): 2913–24. https://doi.org/10.1109/TITS.2015.2428705.

Rancourt, Marie-Eve, Jean-Francois Cordeau, and Gilbert Laporte. 2013. “Long-Haul Vehicle Routing and Scheduling with Working Hour Rules.” *Transportation Science* 47 (1): 81–107.

Rodier, Caroline J, Susan A Shaheen, Denise M Allen, and Brenda Dix. 2010. “Commercial Vehicle Parking in California : Exploratory Evaluation of the Problem and Solutions.”

Sadek, Bassel A, Elliot W Martin, and Susan Shaheen. 2018. “Truck Parking Forecasting and Error Correction Using a Fourier Method: A Case Study of I-5 in California.” In *Transportation Research Board 97th Annual Meeting*. Washington DC.

Schiffer, M, G Laporte, M Schneider, and G Walther. 2017. “The Impact of Synchronizing Driver Breaks and Recharging Operations for Electric Vehicles.”

Schoomer, B. Alva. 1964. “The Incorporation of Step Functions and Ramp Functions into a Linear Programming Model.” *Operations Research* 12 (5): 773–77.

Smith, D, B Ozpineci, R L Graves, P T Jones, J Lustbader, Ken Kelly, Kevin Walkowicz, et al. 2019. “Medium-and Heavy-Duty Vehicle Electrification: An Assessment of Technology and Knowledge Gaps,” no. December. https://www.osti.gov/servlets/purl/1615213.

Sochor, Jana, and Gideon Mbiydzenyuy. 2013. “Assessing the Benefits of Intelligent Truck Parking.” *International Journal of Intelligent Transportation Systems Research* 11 (2): 43–53. https://doi.org/10.1007/s13177-012-0055-3.

Sripad, Shashank, and Venkatasubramanian Viswanathan. 2017. “Performance Metrics Required of Next-Generation Batteries to Make a Practical Electric Semi Truck.” *ACS Energy Letters* 2 (7): 1669–73. https://doi.org/10.1021/acsenergylett.7b00432.

———. 2019. “Quantifying the Economic Case for Electric Semi-Trucks.” *ACS Energy Letters* 4 (1): 149–55. https://doi.org/10.1021/acsenergylett.8b02146.

Sun, Yichen, Tomer Toledo, Katherine Rosa, Moshe E. Ben-Akiva, Kate Flanagan, Ricardo Sanchez, and Erika Spissu. 2013. “Route Choice Characteristics for Truckers.” *Transportation Research Record*, no. 2354: 115–21. https://doi.org/10.3141/2354-12.

Tavafoghi, Hamidreza, Kameshwar Poolla, and Pravin Varaiya. 2019. “A Queuing Approach to Parking: Modeling, Verification, and Prediction,” August. http://arxiv.org/abs/1908.11479.

Ticha, Hamza Ben, Nabil Absi, Dominique Feillet, Alain Quilliot, and Tom Van Woensel. 2021. “The Time-Dependent Vehicle Routing Problem with Time Windows and Road-Network Information.” *SN Operations Research Forum* 2 (1): 4. https://doi.org/10.1007/s43069-020-00049-6.

U.S. Department of Energy. 2015. “Long-Haul Truck Idling Burns Up Profits.” https://afdc.energy.gov/files/u/publication/hdv_idling_2015.pdf.

U.S. Department of Transportation. 2015. “Jason’s Law Truck Parking Survey Results and Comparative Analysis.”

U.S. Department Of Transportation. 2013. “Freight Facts and Figures 2013.”

———. 2015. “Freight Facts and Figures 2015.”

U.S. Energy Information Administration. 2016. “Carbon Dioxide Emissions Coefficients.” 2016. https://www.eia.gov/environment/emissions/co2_vol_mass.php.

Vital, Filipe, and Petros Ioannou. 2019. “Long-Haul Truck Scheduling with Driving Hours and Parking Availability Constraints.” In *2019 IEEE Intelligent Vehicles Symposium (IV)*, June:620–25. IEEE. https://doi.org/10.1109/IVS.2019.8814011.

———. 2020. “Truck Routing under Rest Area Parking Constraints.” In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6. IEEE. https://doi.org/10.1109/ITSC45102.2020.9294253.

Wang, Jinghui, and Hesham A. Rakha. 2017. “Fuel Consumption Model for Heavy Duty Diesel Trucks: Model Development and Testing.” *Transportation Research Part D: Transport and Environment* 55: 127–41. https://doi.org/10.1016/j.trd.2017.06.011.

Weenen, Rob de Leeuw van, Sean Newton, Menno Menist, Frédéric Maas, Dirk Penasse, Michael Nielsen, Aristos Halatsis, Toni Männistö, Iraklis Stamos, and Peter Paul Ruschin. 2019. “Study on Safe and Secure Parking Places for Trucks.” https://doi.org/10.2832/067535.

Zaloshnja, Eduard, and Ted Miller. 2007. “Unit Costs of Medium/Heavy Truck Crashes.” *Report No. FMCSA-RRA-07-034*. Calverton, MD.

Data Summary

Products of Research

As part of this project, we collected real-world diesel truck' characteristics and trajectory data traveling on two urban freight corridors—Alameda St and Wilmington Ave—near the Port of Los Angeles in California. These data were used to evaluate the effectiveness of a truck EAD system, called Eco-Drive, in reducing energy consumption of the truck.

Data Format and Content

Both real-world baseline and Eco-Drive data files are in .csv format. The contents of each file include vehicle speed (in mph), fuel rate (in liters/s), and distance to the next intersection (in meters) collected every second (1Hz).

Data Access and Sharing

The data are made available publicly via DataDRYAD: https://datadryad.org/stash, which is licensed under a CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license. The DOI for the dataset is https://doi.org/10.6086/D1BT3K.

Reuse and Redistribution

The data are restricted for research use only. If the data are used, our work should be properly cited as:

> Wei, Zhensong; Brown, Dylan; Hao, Peng; Boriboonsomsin, Kanok (2022), Real-world heavy-duty truck trajectories on signalized corridors, Dryad, Dataset, https://doi.org/10.6086/D1BT3K

Appendix A: Pre-Survey of Truck EAD Experiment

ECO-Driving Technology and Behavior Research for Heavy-Duty Trucks

"Before" Survey

Thank you for your participation. This is the first of four surveys you will take in this study. This survey should take about 3-5 minutes to complete. You are not required to answer any question that you would prefer not to answer. All responses are confidential and you may withdraw from the study at any time.

Let's start with your typical day at work.

1. The majority of your driving job is within how many miles from the home base?

   ☐ Less than 25 miles ☐ 25 to 49 miles ☐ 50 to 99 miles
   ☐ 100 to 249 miles ☐ 250 miles or more

2. What portion of your driving while on the job occur on streets with traffic lights?

   ☐ Less than 20% ☐ 20% to 39% ☐ 40% to 59%
   ☐ 60% to 79% ☐ 80% or more

Now, imagine that you are driving on a street that has traffic lights.

3. When approaching a GREEN light, how useful is it for you to know how many seconds are left before the light will change from GREEN to YELLOW?

   ☐ Extremely useful ☐ Very useful ☐ Moderately useful
   ☐ Somewhat useful ☐ Not at all

4. When approaching a RED light, how useful is it for you to know how many seconds are left before the light will change from RED to GREEN?

   ☐ Extremely useful ☐ Very useful ☐ Moderately useful
   ☐ Somewhat useful ☐ Not at all

5. When stopping at a RED light, how useful is it for you to know how many seconds are left before the light will change from RED to GREEN?

   ☐ Extremely useful ☐ Very useful ☐ Moderately useful
   ☐ Somewhat useful ☐ Not at all

Let's say we put a smart device on your truck. Again, imagine that you are driving on a street that has traffic lights, and you are approaching a traffic light.

6. How useful is the device if it can recommend what speed you should be driving at in order to pass through the intersection on green?

   ☐ Extremely useful ☐ Very useful ☐ Moderately useful
   ☐ Somewhat useful ☐ Not at all

7. How useful is the device if it can recommend you to slow down ahead of time because it knows that you will not be able to pass through the intersection on green?

   ☐ Extremely useful ☐ Very useful ☐ Moderately useful
   ☐ Somewhat useful ☐ Not at all

ECO-Driving Technology and Behavior Research for Heavy-Duty Trucks

Encuesta preliminar

Gracias por su participación. Esta es la primera de cuatro encuestas que realizará en este estudio. Esta encuesta tomara alrededor de 3-5 minutos para completar. No está obligado a responder a ninguna pregunta que prefiera no responder. Todas las respuestas son confidenciales y usted puede retirarse del estudio en cualquier momento.

Comencemos con tu día típico en el trabajo.

1. ¿La mayoría de su trabajo de conducción está dentro de cuántas millas de la base de operaciones?

   ☐ Menos de 25 millas ☐ 25 a 49 millas ☐ 50 a 99 millas
   ☐ 100 a 249 millas ☐ 250 millas o más

2. ¿Qué parte de su conducción mientras está en el trabajo se produce en las calles con semáforos?

   ☐ Menos del 20% ☐ 20% a 39% ☐ 40% a 59%
   ☐ 60% a 79% ☐ 80% o más

Ahora, imagine que usted está conduciendo en una calle que tiene semáforos.

3. Al acercarse a una luz VERDE, ¿qué tan útil es para usted saber cuántos segundos quedan antes de que la luz cambie de VERDE a AMARILLO?

   ☐ Extremadamente útil ☐ Muy útil ☐ Moderadamente útil
   ☐ Algo útil ☐ Para nada

4. Al acercarse a una luz ROJA, ¿qué tan útil es para usted saber cuántos segundos quedan antes de que la luz cambie de ROJO a VERDE?

   ☐ Extremadamente útil ☐ Muy útil ☐ Moderadamente útil
   ☐ Algo útil ☐ Para nada

5. Al detenerse en una luz ROJA, ¿qué tan útil es para usted saber cuántos segundos quedan antes de que la luz cambie de ROJO a VERDE?

   ☐ Extremadamente útil ☐ Muy útil ☐ Moderadamente útil
   ☐ Algo útil ☐ Para nada

Digamos que ponemos un dispositivo inteligente en su camión. Una vez más, imagine que usted está conduciendo en una calle que tiene semáforos, y se estás acercando a un semáforo.

6. ¿Qué tan útil es el dispositivo si puede recomendar a qué velocidad debe conducir para pasar a través de la intersección en verde?

   ☐ Extremadamente útil ☐ Muy útil ☐ Moderadamente útil
   ☐ Algo útil ☐ Para nada

7. ¿Qué tan útil es el dispositivo si puede recomendarle que reduzca su velocidad con anticipación porque sabe que no será capaz de pasar a través de la intersección en verde?

   ☐ Extremadamente útil ☐ Muy útil ☐ Moderadamente útil
   ☐ Algo útil ☐ Para nada

Appendix B: Post-Survey of Truck EAD Experiment

ECO-Driving Technology and Behavior Research for Heavy-Duty Trucks

"After" Survey

Thank you for your participation. This is the last of four surveys you will take in this study. This survey should take about 3-5 minutes to complete. You are not required to answer any question that you would prefer not to answer. All responses are confidential and you may withdraw from the study at any time.

Think about your experience driving in the simulator with Eco-Drive.

1. Between the two user interface options of Eco-Drive, which one(s) would you use in your driving job?

   ☐ Either of them ☐ The one with visual & audio feedback

   ☐ The one with audio only feedback ☐ None of them

2. If the truck you drive at your job is equipped with Eco-Drive that provides visual & audio feedback, how often will you use it when driving on streets with traffic lights?

   ☐ Always ☐ Often ☐ Sometimes ☐ Seldom

   ☐ Never

3. If the truck you drive at your job is equipped with Eco-Drive that provides audio-only feedback, how often will you use it when driving on streets with traffic lights?

   ☐ Always ☐ Often ☐ Sometimes ☐ Seldom

   ☐ Never

Please tell us more about what you think of Eco-Drive.

4. What do you like most about Eco-Drive?

   _____

   _____

   _____

   _____

5. What do you dislike most about Eco-Drive?

   _____

   _____

   _____

   _____

6. How can we make Eco-Drive better for you and other truck drivers?

_____
_____
_____
_____

Finally, let's wrap up this survey.

7. What is your age?_____
8. How many years have you been a professional truck driver?_____

THANK YOU!

ECO-Driving Technology and Behavior Research for Heavy-Duty Trucks

Encuesta posterior

Gracias por su participación. Esta es la última de las cuatro encuestas que realizará en este estudio. Esta encuesta tomara alrededor de 3-5 minutos para completar. No está obligado a responder a ninguna pregunta que prefiera no responder. Todas las respuestas son confidenciales y usted puede retirarse del estudio en cualquier momento.

Piense en su experiencia conduciendo en el simulador con Eco-Drive.

1. ¿Entre las dos opciones de interfaz de usuario de Eco-Drive, cuál(es) utilizará en su trabajo de conducción?

   ☐ Cualquiera de las dos ☐ El que tiene Visual & Comentarios de audio

   ☐ El que tiene sólo comentarios de audio ☐ Ninguna de las dos

2. ¿Si el camión que conduce en su trabajo está equipado con Eco-Drive que proporciona visual & comentarios de audio, con qué frecuencia lo usará cuando se conduce en las calles con semáforos?

   ☐ Siempre ☐ frecuentemente ☐ Algunas veces ☐ Rara vez ☐ Nunca

3. ¿Si el camión que conduce en su trabajo está equipado con Eco-Drive que proporciona sólo comentarios de audio, con qué frecuencia lo utilizará cuando conduzca en las calles con semáforos?

   ☐ Siempre ☐ frecuentemente ☐ Algunas veces ☐ Rara vez ☐ Nunca

Por favor, cuéntanos más sobre lo que piensas de Eco-Drive.

4. ¿Qué es lo que más le gusta de Eco-Drive?

   _____
   _____
   _____
   _____

5. ¿Qué es lo que no le gusta de Eco-Drive?

   _____
   _____
   _____
   _____

6. ¿Cómo podemos hacer Eco-Drive mejor para usted y otros conductores de camiones?

_____
_____
_____
_____

Por último, vamos a terminar esta encuesta.

7. ¿Cuántos años tiene?_____
8. ¿Cuántos años has sido un camionero profesional?_____

¡Gracias!

Appendix C: Parts of Proof of Theorem 1.3

Case a), i.e., $I \in \Omega_1$

Consider the Lyapunov function

$$V(\rho) = \frac{\left(\rho - d/v_f\right)^2}{2},$$

then we have

$$\dot{V}(\rho) = \left(\rho - \frac{d}{v_f}\right)\dot{\rho} = -\left(\rho - \frac{d}{v_f}\right)(q_2 - q_1).$$

Using equation (41) and Figure 79, we have that when $0 \le \rho \le \frac{C_d}{v_f}$, $q_1 = d$ and $q_2 = v_f\rho$. Thus

$$q_2 - q_1 = v_f\left(\rho - \frac{d}{v_f}\right) \tag{155}$$

therefore $\dot{V} = -v_f\left(\rho - \frac{d}{v_f}\right)^2, \forall\rho \in \left[0, \frac{C_d}{v_f}\right]$. When $\frac{C_d}{v_f} < \rho \le \tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\tilde{w}}$, $\rho - \frac{d}{v_f} > 0$, $q_1 = d$ and $q_2 = (1-\epsilon_0)C_d$. Thus

$$q_2 - q_1 = (1-\epsilon_0)C_d - d \ge \frac{(1-\epsilon_0)C_d - d}{\rho^j - d/v_f}\left(\rho - d/v_f\right) \tag{156}$$

due to $(1-\epsilon_0)C_d - d > 0$ and $d/v_f < \rho \le \rho^j, \forall\rho \in (\frac{C_d}{v_f}, \rho^j]$, which implies $0 < \frac{\rho - d/v_f}{\rho^j - d/v_f} \le 1$. Therefore $\dot{V} \le -\frac{(1-\epsilon_0)C_d - d}{\rho^j - d/v_f}\left(\rho - \frac{d}{v_f}\right)^2, \forall\rho \in (\frac{C_d}{v_f}, \le \tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\tilde{w}}]$. When $\tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\tilde{w}} < \rho \le \rho^j - \frac{d}{w}$, $\rho - \frac{d}{v_f} > 0$, $q_1 = d$ and $q_2 = \tilde{w}\left(\tilde{\rho}^j - \rho\right)$. Thus

$$\begin{aligned} q_2 - q_1 &= \tilde{w}(\tilde{\rho}^j - \rho) - d \ge \tilde{w}\left[\tilde{\rho}^j - \left(\rho^j - \tfrac{d}{w}\right)\right] - d = \tilde{w}\left[\tilde{\rho}^j - \left(\rho^j - \tfrac{d}{w}\right)\right] - \\ & w\left[\rho^j - \left(\rho^j - \tfrac{d}{w}\right)\right] \\ &= \tilde{w}(\tilde{\rho}^j - \rho_c) - w(\rho^j - \rho_c) + (\tilde{w} - w)[\rho_c - \left(\rho^j - \frac{d}{w}\right)] \\ &= C - C + (\tilde{w} - w)[\rho_c - \left(p^j - \frac{d}{w}\right)] \\ &\ge \frac{(\tilde{w} - w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right), \end{aligned} \tag{157}$$

due to $0 < \frac{\rho - d/v_f}{\rho^j - d/v_f} \leq 1$, $\widetilde{w} - w < 0$ and $\rho_c - \left(\rho^j - \frac{d}{w}\right) < 0$. Therefore, $\dot{V} \leq -\frac{(\widetilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right)^2$, $\forall \rho \in (\tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\widetilde{w}}, \rho^j - \frac{d}{w}]$. When $\rho^j - \frac{d}{w} < \rho \leq \rho^j$, $\rho - \frac{d}{v_f} > 0$, $q_1 = w\left(\rho^j - \rho\right)$ and $q_2 = \widetilde{w}\left(\tilde{\rho}^j - \rho\right)$. Thus

$$\begin{aligned} q_2 - q_1 &= \widetilde{w}\left(\tilde{\rho}^j - \rho\right) - w\left(\rho^j - \rho\right) \\ &= \widetilde{w}\left(\tilde{\rho}^j - \rho_c\right) - w\left(\rho^j - \rho_c\right) + (\widetilde{w} - w)(\rho_c - \rho) \\ &\geq (\widetilde{w} - w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right] \\ &\geq \frac{(\widetilde{w} - w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right), \end{aligned} \tag{158}$$

since $0 < \frac{\rho - d/v_f}{\rho^j - d/v_f} \leq 1$, $\widetilde{w} - w < 0$ and $\rho_c - \left(\rho^j - \frac{d}{w}\right) < 0$. Therefore $\dot{V} \leq -\frac{(\widetilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right)^2$, $\forall \rho \in (\rho^j - \frac{d}{w}, \rho^j]$. From (155)-(158), we can conclude that

$$\dot{V} \leq -\alpha\left(\rho - \frac{d}{v_f}\right)^2, \ \forall \rho(0) \in \left[0, \rho^j\right] \tag{159}$$

where $\alpha = \min\{v_f, \frac{(1-\epsilon_0)C_d - d}{\rho^j - d/v_f}, \frac{(\widetilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\} > 0, \ \forall \rho(0) \in \left[0, \rho^j\right]$, which implies exponential stability of the equilibrium point $\rho^e = \frac{d}{v_f}$ and exponential convergence of $\rho(t)$ to $\frac{d}{v_f}$, $\forall \rho(0) \in \left[0, \rho^j\right]$. The rate of convergence of ρ to the equilibrium $\frac{d}{v_f}$ is greater or equal to α and can be shown by substituting for $\left(\rho - \frac{d}{v_f}\right)^2 = 2V$ in (159) and integrating both sides of the inequality.

Case b), i.e., $I \in \Omega_2$

From Figure 80, the situation where $\rho(0) \in (\rho^j - \frac{d}{w}, \rho^j]$ is divided into two cases:

- Case I: $\rho(0) \in (\rho^j - \frac{d}{w}, \tilde{\rho}^j - \frac{d}{\widetilde{w}}]$.
- Case II: $\rho(0) \in (\tilde{\rho}^j - \frac{d}{\widetilde{w}}, \rho^j]$.

In case I, $q_1 = w\left(\rho^j - \rho\right)$ and $q_2 = d$ as long as $\rho \in (\rho^j - \frac{d}{w}, \tilde{\rho}^j - \frac{d}{\widetilde{w}}], \ \forall t \geq 0$, which we need to show. We have

$$\dot{\rho} = q_1 - q_2 = -w\rho + w\rho^j - d,$$

whose solution is

$$\rho(t) = \left(\rho^j - \frac{d}{w}\right) + \left[\rho(0) - \left(\rho^j - \frac{d}{w}\right)\right] e^{-wt}.$$

Since $0 < \rho^j - \frac{d}{w} < \rho(0) \leq \tilde{\rho}^j - \frac{d}{\tilde{w}}$ and $w > 0$, it follows that $\rho \in (\rho^j - \frac{d}{w}, \tilde{\rho}^j - \frac{d}{\tilde{w}}]$, $\forall t \geq 0$ and $\rho(t)$ converges exponentially fast to $\rho^j - \frac{d}{w}$ according to the above equation.

In case II, when $\rho(0) \in (\tilde{\rho}^j - \frac{d}{\tilde{w}}, \rho^j]$, it follows from Figure 80 that $q_1 = w(\rho^j - \rho)$, $q_2 = \tilde{w}(\tilde{\rho}^j - \rho)$ and

$$\dot{\rho} = -(w - \tilde{w})\rho + (w - \tilde{w})\rho_c,$$

as long as $\rho(t) \in (\tilde{\rho}^j - \frac{d}{\tilde{w}}, \rho^j]$, whose solution is

$$\rho(t) \quad = \rho_c + (\rho(0) - \rho_c)e^{-w_0 t} \leq \left(\rho^j - \frac{d}{w}\right) + (\rho(0) - \rho_c)e^{-w_0 t},$$

where $w_0 = w - \tilde{w} > 0$. Since $\rho_c < \tilde{\rho}^j - \frac{d}{\tilde{w}}$, it follows that $\rho(t)$ will decrease exponentially to the value of $\tilde{\rho}^j - \frac{d}{\tilde{w}}$ at which instant $\dot{\rho}$ switches to case I which guarantees exponential convergence to $\rho^j - \frac{d}{w}$. The above equation implies that

$$\left|\rho(t) - \left(\rho^j - \frac{d}{w}\right)\right| \leq c_0 e^{-\alpha t},$$

where $c_0 = \rho(0) - \left(\rho^j - \frac{d}{w}\right)$ and $\alpha = \min\{w, w - \tilde{w}\}$.

Case c), i.e., $I \in \Omega_3$

Consider the Lyapunov function

$$V(\rho) = \frac{(\rho - \rho_2^e)^2}{2},$$

where $\rho_2^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$. Then $\dot{V} = -(\rho - \rho_2^e)(q_2 - q_1)$. As shown before, when $\frac{C_d}{v_f} < \rho \leq \rho^j - \frac{d}{w}$, we have $\rho - \rho_2^e < 0$ and

$$q_2 - q_1 = (1 - \epsilon_0)C_d - d \leq \frac{(1 - \epsilon_0)C_d - d}{C_d/v_f - \rho_2^e}(\rho - \rho_2^e), \tag{160}$$

due to $0 < \frac{(\rho - \rho_2^e)}{C_d/v_f - \rho_2^e} < 1, \forall \rho \in (\frac{C_d}{v_f}, \rho^j - \frac{d}{w}]$ and $(1 - \epsilon_0)C_d - d < 0$. Therefore $\dot{V} \leq -\frac{(1-\epsilon_0)C_d - d}{C_d/v_f - \rho_2^e}(\rho - \rho_2^e)^2$. When $\rho^j - \frac{d}{w} < \rho \leq \tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\tilde{w}}$,

$$q_2 - q_1 \quad = (1-\epsilon_0)C_d - w(\rho^j - \rho) = w\left[\rho - \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right)\right] = w(\rho - \rho_2^e). \quad (161)$$

Therefore $\dot{V} = -w(\rho - \rho_2^e)^2$. When $\tilde{\rho}^j - \frac{(1-\epsilon_0)C_d}{\tilde{w}} < \rho \leq \rho^j$, we have $\rho - \rho_2^e > 0$ and

$$q_2 - q_1 \quad = \tilde{w}(\tilde{\rho}^j - \rho) - w(\rho^j - \rho) = (w - \tilde{w})(\rho - \rho_c) \geq (w - \tilde{w})(\rho - \rho_2^e), \quad (162)$$

due to $w - \tilde{w} > 0$ and $\rho_2^e > \rho_c$. Therefore, $\dot{V} \leq -(w - \tilde{w})(\rho - \rho_2^e)^2$. From (160)-(162), we conclude that $\forall \rho \in (\frac{C_d}{v_f}, \rho^j]$,

$$\dot{V} \leq -\alpha(\rho - \rho_2^e)^2,$$

where $\alpha = \min\{\frac{d-(1-\epsilon_0)C_d}{\rho_2^e - C_d/v_f}, w, (w - \tilde{w})\} > 0$, $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$ which implies exponential stability of the equilibrium point $\rho_2^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$ and exponential convergence of $\rho(t)$ to ρ_2^e, $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$.

Case d), i.e., $I \in \Omega_4$

Consider the Lyapunov function

$$V(\rho) = \frac{(\rho - \rho^e)^2}{2},$$

where $\rho^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$. Thus

$$\dot{V} = -(\rho - \rho^e)(q_2 - q_1).$$

From Figure 82, it is clear that $\forall \rho \in \left[0, \frac{C_d}{v_f}\right]$, $\rho - \rho^e < 0$ and

$$q_2 - q_1 \leq C_d - d \leq \frac{d - C_d}{\rho^e}(\rho - \rho^e),$$

therefore,

$$\dot{V} \leq -\frac{d - C_d}{\rho^e}(\rho - \rho^e)^2, \forall \rho \in \left[0, \frac{C_d}{v_f}\right].$$

Similar to the case $I \in \Omega_3$, we have $\forall \rho \in (\frac{C_d}{v_f}, \rho^j]$,

$$\dot{V} \leq -\min\{\frac{d - (1-\epsilon_0)C_d}{\rho^e - C_d/v_f}, w, (w - \tilde{w})\}(\rho - \rho^e)^2.$$

Therefore, $\forall \rho \in \left[0, \rho^j\right]$, the time derivative of the Lyapunov function satisfies

$$\dot{V} \leq -\alpha(\rho - \rho^e)^2,$$

where $\alpha = \min\{\frac{d-C_d}{\rho^e}, \frac{d-(1-\epsilon_0)C_d}{\rho^e - C_d/v_f}, w, (w-\widetilde{w})\} > 0$, which implies exponential convergence to the equilibrium point $\rho^e, \forall \rho(0) \in [0, \rho^j]$.

Case e), i.e., $I \in \Omega_5$

Consider the Lyapunov function

$$V(\rho) = \frac{\left(\rho - \min\{d, C\}/v_f\right)^2}{2}.$$

Then if $d < C$, $\dot{V} = -\left(\rho - d/v_f\right)(q_2 - q_1)$. According to Figure 83, when $0 \le \rho \le \rho_c$, we have that $q_1 = d$, and $q_2 = v_f \rho$. Thus

$$q_2 - q_1 = v_f\left(\rho - d/v_f\right). \tag{163}$$

Therefore $\dot{V} = -v_f\left(\rho - d/v_f\right)^2$. When $\rho_c < \rho \le \rho^j - \frac{d}{w}$, we have $\rho - d/v_f > 0$, $q_1 = d$ and $q_2 = \widetilde{w}\left(\tilde{\rho}^j - \rho\right)$. Using equation (157), we have

$$q_2 - q_1 = \widetilde{w}\left(\tilde{\rho}^j - \rho\right) - d \ge \widetilde{w}\left[\tilde{\rho}^j - \left(\rho^j - \frac{d}{w}\right)\right] - d \ge \frac{(\widetilde{w} - w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right) \tag{164}$$

Therefore $\dot{V} \le -\frac{(\widetilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right)^2$. When $\rho^j - \frac{d}{w} < \rho \le \rho^j$, we have $\rho - d/v_f > 0$, $q_1 = w\left(\rho^j - \rho\right)$ and $q_2 = \widetilde{w}\left(\tilde{\rho}^j - \rho\right)$, which together with equation (158) gives

$$q_2 - q_1 \quad = \widetilde{w}\left(\tilde{\rho}^j - \rho\right) - w\left(\rho^j - \rho\right) \ge \frac{(\widetilde{w} - w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right),$$

Therefore $\dot{V} \le -\frac{(\widetilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\left(\rho - d/v_f\right)^2$. As a result, we conclude that $\forall \rho \in [0, \rho^j]$,

$$\dot{V} \le -\alpha\left(\rho - d/v_f\right)^2,$$

where $\alpha = \min\{v_f, \frac{(\widetilde{w}-w)\left[\rho_c - \left(\rho^j - \frac{d}{w}\right)\right]}{\rho^j - d/v_f}\} > 0$, which guarantees exponential stability of the equilibrium point $\rho^e = d/v_f$ and exponential convergence of $\rho(t)$ to $\rho^e, \forall \rho(0) \in [0, \rho^j]$.

If $d \ge C, \forall \rho \in [0, \rho_c], q_1 = C, q_2 = v_f \rho$, and $\forall \rho \in (\rho_c, \rho^j], q_1 = w\left(\rho^j - \rho\right), q_2 = \widetilde{w}\left(\tilde{\rho}^j - \rho\right)$. Therefore

$$\dot{V} \quad = \begin{cases} -v_f(\rho - \rho_c)^2, & \text{if } \rho \in [0, \rho_c] \\ -(w - \widetilde{w})(\rho - \rho_c)^2, & \text{if } \rho \in (\rho_c, \rho^j]' \end{cases}$$

which implies that $\dot{V} \leq -\min\{v_f, (w - \widetilde{w})\}(\rho - \rho_c)^2$. The properties of V and $\dot{V}$ imply exponential stability of the equilibrium point $\rho^e = \rho_c = \frac{C}{v_f}$ and exponential convergence of $\rho(t)$ to ρ^e, $\forall \rho(0) \in [0, \rho^j]$, due to $w - \widetilde{w} > 0$.

Appendix D: Proof of Theorem 1.4

For the proof of Theorem 1.4, we use the following two lemmas: Lemma 4.1 gives the region of ρ^e within the set S. For a set $A \subset \Re^N$ and a point $x_0 \in \Re^N$, the distance between x_0 and A is defined as:

$$d(x_0, A) = \inf_{x \in A} \| x - x_0 \|.$$

Then we have the following lemma.

Lemma 4.1. *Let ρ^e be an equilibrium state of system (51) then we have the following results:*

a) *If $C_d < C$, i.e., $I \in \bigcup_{i=1}^4 \Omega_i$, then $\rho^e \in S^I$, where $S^I = \{\rho | \frac{\min\{d,C\}}{v_f} \le \rho_i \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}, i = 1,2,\dots,N\} \subset S$. Furthermore, $\forall \rho(0) \in S$, $d(\rho(t), S^I)$ converges to 0 exponentially fast.*

b) *If $C_d \ge C$, i.e., $I \in \Omega_5$, then $\rho^e \in \bar{S}^I$, where $\bar{S}^I = \{\rho | \frac{\min\{d,C\}}{v_f} \le \rho_i \le \rho_c, i = 1,2,\dots,N\} \subset S$. Furthermore, $\forall \rho(0) \in S$, $d(\rho(t), \bar{S}^I)$ converges to 0 exponentially fast.*

Proof of Lemma 4.1:

a) For $I \in \bigcup_{i=1}^4 \Omega_i$, we first show that $\rho_i^e \ge \frac{\min\{d,C\}}{v_f}$, for $i = 1,2,\dots,N$. Assume that $0 \le \rho_1^e < \frac{\min\{d,C\}}{v_f}$, then $w(\rho^j - \rho_1^e) \ge C$ due to $\rho_1^e < \frac{\min\{d,C\}}{v_f} \le \rho_c$. Therefore the corresponding equilibrium flow rate

$$q_1^e = \min\{d, C, w(\rho^j - \rho_1^e)\} = \min\{d, C\},$$

$$q_2^e = \min\{v_f \rho_1^e, \tilde{w}(\tilde{\rho}^j - \rho_1^e), C, w(\rho^j - \rho_2^e)\} \le v_f \rho_1^e,$$

which implies that

$$\dot{\rho}_1 = q_1^e - q_2^e \ge \min\{d, C\} - v_f \rho_1^e > 0, \text{ as } \rho_1^e < \frac{\min\{d, C\}}{v_f},$$

which violates the equilibrium condition (43) hence $\rho_1^e \ge \frac{\min\{d,C\}}{v_f}$. For any $i = 1,2,..,N-1$, assume $\rho_i^e \ge \frac{\min\{d,C\}}{v_f}$ and check the property of ρ_{i+1}^e. If $0 \le \rho_{i+1}^e < \frac{\min\{d,C\}}{v_f}$, we have $v_f \rho_{i+1}^e < \min\{d, C\} < C < w(\rho^j - \rho_{i+1}^e)$. Thus

$$q_{i+1}^e = \min\{v_f \rho_i^e, \tilde{w}(\tilde{\rho}^j - \rho_i^e), C, w(\rho^j - \rho_{i+1}^e)\} = \min\{v_f \rho_i^e, \tilde{w}(\tilde{\rho}^j - \rho_i^e)\},$$

$$q_{i+2}^e = \min\{v_f \rho_{i+1}^e, \tilde{w}(\tilde{\rho}^j - \rho_{i+1}^e), C, w(\rho^j - \rho_{i+2}^e)\} \le v_f \rho_{i+1}^e < \min\{d, C\}.$$

If $q_{i+1}^e = \widetilde{w}(\tilde{\rho}^j - \rho_i^e)$, then $\widetilde{w}(\tilde{\rho}^j - \rho_i^e) \le v_f \rho_i^e$, which implies $\rho_i^e \ge \rho_c$. Since ρ_i^e is the equilibrium density in section i, we have $q_i^e = q_{i+1}^e$, and

$$w(\rho^j - \rho_i^e) \ge q_i^e = q_{i+1}^e = \widetilde{w}(\tilde{\rho}^j - \rho_i^e),$$

which implies $\rho_i^e \le \rho_c$. Thus $\rho_i^e = \rho_c$ and $q_i^e = q_{i+1}^e = \widetilde{w}(\tilde{\rho}^j - \rho_c) = C \ge \min\{d, C\} > q_{i+2}^e$. If $q_{i+1}^e = v_f \rho_i^e$, then $q_{i+1}^e \ge \min\{d, C\} > q_{i+2}^e$ due to $\rho_i^e \ge \frac{\min\{d,C\}}{v_f}$. Therefore, for all possible $q_{i+1} = \min\{v_f \rho_i^e, \widetilde{w}(\tilde{\rho}^j - \rho_i^e)\}$, we have $q_{i+1}^e > q_{i+2}^e$, which violates the equilibrium condition (43). Therefore, the assumption $0 \le \rho_{i+1}^e < \frac{\min\{d,C\}}{v_f}$ is invalid, which implies that $\rho_{i+1}^e \ge \frac{\min\{d,C\}}{v_f}$. By mathematical induction, we know that

$$\rho_i^e \ge \frac{\min\{d, C\}}{v_f}, i = 1,2,\dots,N. \tag{165}$$

Then we show that $\rho_i^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, for $i = 1,2,\dots,N$. Assume that $\rho^j - \frac{(1-\epsilon_0)C_d}{w} < \rho_N^e \le \rho^j$, then

$$\begin{aligned} q_{N+1}^e &= \min\{(1-\epsilon_0)C_d, \widetilde{w}(\tilde{\rho}^j - \rho_N^e)\}, \\ q_N^e &= \min\{v_f \rho_{N-1}^e, \widetilde{w}(\tilde{\rho}^j - \rho_{N-1}^e), C, w(\rho^j - \rho_N^e)\} \le w(\rho^j - \rho_N^e). \end{aligned}$$

Since $\rho_N^e > \rho^j - \frac{(1-\epsilon_0)C_d}{w} > \rho_c$, we have $w(\rho^j - \rho_N^e) < (1-\epsilon_0)C_d$ and $w(\rho^j - \rho_N^e) < \widetilde{w}(\tilde{\rho}^j - \rho_N^e)$. Therefore $q_N^e \le w(\rho^j - \rho_N^e) < q_{N+1}^e$, which contradicts the the equilibrium condition (43). Thus $\rho_N^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$.

Assume $\rho_i^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, for any $i = 2,3,\dots,N$, we check the property of ρ_{i-1}^e. If $\rho^j - \frac{(1-\epsilon_0)C_d}{w} < \rho_{i-1}^e \le \rho^j$, then $\widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e) < C < v_f \rho_{i-1}^e$ as $\rho_{i-1}^e > \rho_c$. Therefore

$$\begin{aligned} q_i^e &= \min\{v_f \rho_{i-1}^e, \widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e), C, w(\rho^j - \rho_i^e)\} = \min\{\widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e), w(\rho^j - \rho_i^e)\}, \\ q_{i-1}^e &= \min\{v_f \rho_{i-2}^e, \widetilde{w}(\tilde{\rho}^j - \rho_{i-2}^e), C, w(\rho^j - \rho_{i-1}^e)\} \le w(\rho^j - \rho_{i-1}^e). \end{aligned}$$

Since $\rho_i^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w} < \rho_{i-1}^e$, we have $w(\rho^j - \rho_{i-1}^e) < \widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e)$ and $w(\rho^j - \rho_{i-1}^e) < (1-\epsilon_0)C_d \le w(\rho^j - \rho_i^e)$. Thus $q_{i-1}^e < q_i^e$, which violates the equilibrium condition (43). Therefore $\rho_{i-1}^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$. By mathematical induction, we have

$$\rho_i^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}, i = 1,2,\dots,N. \tag{166}$$

Combining the two inequalities (165)-(166), we can conclude

$$\frac{\min\{d, C\}}{v_f} \le \rho_i^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}, i = 1,2,\dots,N.$$

To show that $d(\rho(t), S^I)$ converges to 0 exponentially fast $\forall \rho(0) \in S$, it is equivalent to show that $\forall \delta > 0, \exists T > 0$, such that $\forall t > T$

$$\frac{\min\{d, C\}}{v_f} - \delta < \rho_i(t) < \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta,\ i = 1,2,\dots,N \tag{167}$$

and $d(\rho(t), S^I)$ is bounded from above by a decaying exponential function.

First we show the left half of inequality (167). Since $q_1 = \min\{d, C, w(\rho^j - \rho_1)\}$ and $q_2 \le v_f \rho_1$, we have

$$\dot{\rho}_1 \quad = q_1 - q_2 \ge \min\{d, C, w(\rho^j - \rho_1)\} - v_f \rho_1. \tag{168}$$

If $\exists t_0 \ge 0$, such that $\rho_1(t_0) \ge \frac{\min\{d,C\}}{v_f}$, then for all $t \ge t_0$ we have the following result: since $\rho_1(t)$ is uniformly continuous, if $\rho_1(t)$ keeps decreasing and $\rho_1(t_1) = \frac{\min\{d,C\}}{v_f}$ for some $t_1 \ge t_0$, then from (168) we have $\dot{\rho}_1(t_1) \ge 0$, which implies that $\rho_1(t)$ will no longer decrease and $\rho_1(t) \ge \frac{\min\{d,C\}}{v_f}, \forall t \ge t_0$. Therefore $\forall \delta_1 > 0$ and $\forall t \ge t_0$, $\rho_1(t) \ge \frac{\min\{d,C\}}{v_f} - \delta_1$.

If $\forall t \ge 0, \rho_1(t) < \frac{\min\{d,C\}}{v_f}$, then in the region $\rho_1(t) < \frac{\min\{d,C\}}{v_f}$, we have

$$\dot{\rho}_1(t) \ge \min\{d, C\} - v_f \rho_1 = -v_f \left(\rho_1 - \frac{\min\{d, C\}}{v_f}\right). \tag{169}$$

By Lemma 3.2.4 in (P. A. Ioannou and Sun 2012), we have

$$\rho_1(t) \ge e^{-v_f t}\left[\rho_1(0) - \frac{\min\{d, C\}}{v_f}\right] + \frac{\min\{d, C\}}{v_f}, \forall t \ge 0. \tag{170}$$

The right side of (170) converges to $\frac{\min\{d,C\}}{v_f}$ exponentially fast, therefore $\forall \delta_1 > 0$, $\exists T_1 > 0$, such that $\forall t > T_1$, $\rho_1(t) \ge \frac{\min\{d,C\}}{v_f} - \delta_1$.

For $i = 1,2,\dots,N-1$, we assume $\rho_i \ge \frac{\min\{d,C\}}{v_f} - \delta_i, \forall t > 0$, where $\delta_i > 0$, then we examine the dynamics of ρ_{i+1}. We have

$$\dot{\rho}_{i+1} \quad = q_{i+1} - q_{i+2} \ge \min\{v_f \rho_i, \tilde{w}(\tilde{\rho}^j - \rho_i), C, w(\rho^j - \rho^{i+1})\} - v_f \rho_{i+1}.$$

Since $\rho_i \ge \frac{\min\{d,C\}}{v_f} - \delta_i$, we have $v_f \rho_i \ge \min\{d, C\} - v_f \delta_i$, therefore

$$\dot{\rho}_{i+1} \ge \min\{\min\{d, C\} - v_f \delta_i, \tilde{w}(\tilde{\rho}^j - \rho_i), w(\rho^j - \rho^{i+1})\} - v_f \rho_{i+1}. \tag{171}$$

Similar to (168), we can show that if $\exists t_0 \geq 0$, such that $\rho_{i+1}(t) \geq \frac{\min\{d,C\}}{v_f} - \delta_i$, then $\rho_{i+1}(t) \geq \frac{\min\{d,C\}}{v_f} - \delta_i, \forall t \geq t_0$, that is, $\forall \delta_{i+1} > \delta_i$ and $\forall t \geq t_0$, $\rho_i(t) > \frac{\min\{d,C\}}{v_f} - \delta_{i+1}$.

If $\rho_{i+1}(t) < \frac{\min\{d,C\}}{v_f} - \delta_i, \forall t \geq 0$, then in the region $\rho_{i+1}(t) < \frac{\min\{d,C\}}{v_f} - \delta_i$, we have

$$\dot{\rho}_{i+1}(t) \geq \min\{d, C\} - v_f\delta_i - v_f\rho_{i+1} = -v_f\left(\rho_{i+1} - \frac{\min\{d, C\}}{v_f} + \delta_i\right).$$

By Lemma 3.2.4 in (P. A. Ioannou and Sun 2012), we have

$$\rho_{i+1}(t) \geq e^{-v_f t}\left[\rho_{i+1}(0) - \frac{\min\{d, C\}}{v_f} + \delta_i\right] + \frac{\min\{d, C\}}{v_f} - \delta_i, \forall t \geq 0. \tag{172}$$

Similarly, the right hand side of equation (172) converges exponentially fast to $\frac{\min\{d,C\}}{v_f} - \delta_i$. Therefore, $\forall \delta_{i+1} > \delta_i$, $\exists T_{i+1} > 0$, such that $\forall t > T_{i+1}$, $\rho_i(t) > \frac{\min\{d,C\}}{v_f} - \delta_{i+1}$. By mathematical induction, we can conclude that for $i = 1,2,\ldots,N$, $\forall \delta_i > 0$, $\exists T_i > 0$, such that $\forall t > \sum_{j=1}^{i} T_j$, $\rho_i(t) \geq \frac{\min\{d,C\}}{v_f} - \delta_i$. If we take $\delta_N < \delta$, $T_1 + T_2 + \ldots + T_N < T$, then the left side of inequality (167) holds.

Next we prove the right half of the inequality (167). Since $q_{N+1} = \min\{v_f\rho_N, \tilde{w}(\tilde{\rho}^j - \rho_N), (1 - \epsilon(\rho_N))C_d\}$ and $q_N < \min\{C, w(\rho^j - \rho_N)\}$, we have

$$\dot{\rho}_N \quad = q_N - q_{N+1} \leq \min\{C, w(\rho^j - \rho_N)\} - \min\{v_f\rho_N, \tilde{w}(\tilde{\rho}^j - \rho_N), (1 - \epsilon(\rho_N))C_d\}.$$

Similar to (168), we can show that if $\exists t_0 \geq 0$, such that $\rho_N(t_0) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, then $\rho_N(t_0) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}, \forall t \geq t_0$, that is, $\forall 0 < \delta_N < \delta$, and $\forall t \geq 0$, $\rho_N(t) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_N$.

If $\rho_N(t) > \rho^j - \frac{(1-\epsilon_0)C_d}{w}, \forall t \geq 0$, then in the region $\rho_N > \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, we have

$$\dot{\rho}_N \leq (\tilde{w} - w)\left(\rho_N - \left(\rho^j - \frac{(1 - \epsilon_0)C_d}{w}\right)\right).$$

By Lemma 3.2.4 in (P. A. Ioannou and Sun 2012), we have

$$\rho_N(t) \leq e^{(\tilde{w}-w)t}\left[\rho_N(0) - \left(\rho^j - \frac{(1 - \epsilon_0)C_d}{w}\right)\right] + \left(\rho^j - \frac{(1 - \epsilon_0)C_d}{w}\right), \forall t \geq 0. \tag{173}$$

Since $\widetilde{w} - w < 0$, the right side of (173) converges to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$ exponentially fast. Therefore, $\forall 0 < \delta_N < \delta, \exists T_N > 0$, such that $\forall t > T_N, \rho_N(t) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_N$.

For $i = 1, \ldots, N-1$, we assume $\rho_{i+1} \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}, \forall t > 0$, where $\delta_{i+1} > 0$, then we check the dynamics of ρ_i. We have

$$\dot{\rho}_i \quad = q_i - q_{i+1} \leq \min\{C, w(\rho^j - \rho_i)\} - \min\{v_f \rho_i, \widetilde{w}(\tilde{\rho}^j - \rho_i), C, w(\rho^j - \rho_{i+1})\}.$$

Since $\rho_{i+1} \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}$, we have $w(\rho^j - \rho_{i+1}) \geq (1-\epsilon_0)C_d - w\delta_{i+1}$. Thus

$$\dot{\rho}_i \quad \leq \min\{C, w(\rho^j - \rho_i)\} - \min\{v_f \rho_i, \widetilde{w}(\tilde{\rho}^j - \rho_i), (1-\epsilon_0)C_d - w\delta_{i+1}\}.$$

Thus, we can show that if $\exists t_0 \geq 0$, such that $\rho_i(t_0) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}$, then $\rho_i(t_0) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}, \forall t \geq t_0$, that is, $\forall \delta_i > \delta_{i+1}$, and $\forall t \geq 0$, $\rho_i(t) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_i$.

If $\rho_i(t_0) > \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}, \forall t \geq 0$, then in the region $\rho_i > \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}$, we have

$$\dot{\rho}_i \leq (\widetilde{w} - w)\left(\rho_N - \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) - \delta_{i+1}\right).$$

By Lemma 3.2.4 in (P. A. Ioannou and Sun 2012), we have

$$\rho_i(t) \leq e^{(\widetilde{w}-w)t}\left[\rho_i(0) - \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}\right)\right] + \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}\right) \quad (174)$$

Thus, the right hand side of equation (174) converges exponentially fast to $\rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_{i+1}$. Therefore, $\forall \delta_i > \delta_{i+1}, \exists T_i > 0$, such that $\forall t > T_i, \rho_i(t) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_i$.

By mathematical induction, we can conclude that for $i = 1,2,\ldots,N, \forall \delta_i > 0, \exists T_i > 0$, such that $\forall t > \sum_{j=1}^{i} T_j, \rho_i(t) \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta_i$. If we take $\delta_1 < \delta$, $T_1 + T_2 + \ldots + T_N < T$, then the right side of inequality (167) holds. Therefore, $d(\rho(t), S^I)$ converges to 0 exponentially fast for all $\rho(0)$ in the feasible set S.

b) Part b) of Lemma 4.1 can be proved in a similar manner.

Specifically, when $I \in \Omega_2$, the equilibrium points of system (51) satisfy the properties given by the following lemma.

Lemma 4.2. *Let $I \in \Omega_2$. If ρ^e is an equilibrium state of system (51), then the corresponding equilibrium flow rate is $q_1^e = q_2^e = \ldots = q_{N+1}^e = d = (1-\epsilon_0)C_d$. Furthermore, ρ^e has the following properties:*

c) *For $i = 1,2,\ldots,N-1$, if $\frac{d}{v_f} < \rho_i^e \le \rho^j - \frac{d}{w}$, then $\rho_k^e = \rho^j - \frac{d}{w}$, for all $i < k \le N$.*

d) *For $i = 2,3,\ldots,N$, if $\frac{d}{v_f} \le \rho_i^e < \rho^j - \frac{d}{w}$, then $\rho_k^e = \frac{d}{v_f}$, for all $1 \le k < i$.*

Proof of Lemma 4.2:

Assume ρ^e is an equilibrium state of system (51) then using Lemma 4.1, we have $\frac{d}{v_f} \le \rho_i^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w} = \rho^j - \frac{d}{w}, i = 1,2,\ldots,N$, therefore

$$q_1^e = \min\{d, C, w(\rho^j - \rho_1^e)\} \le d,$$

$$q_{N+1}^e = \min\{v_f \rho_N^e, \tilde{w}(\tilde{\rho}^j - \rho_N^e), (1-\epsilon(\rho_N^e))C_d\} \ge d.$$

Using the equilibrium condition (53) we have that the equilibrium flow $q_1^e = q_{N+1}^e = d$. Therefore $q_i^e = d$, for $i = 1,2,\ldots,N+1$.

For any $i = 1,2,\ldots,N-1$, if $\frac{d}{v_f} < \rho_i^e \le \rho^j - \frac{d}{w}$, we have $v_f \rho_i^e > d$ and $\tilde{w}(\tilde{\rho}^j - \rho_i^e) \ge \tilde{w}[\tilde{\rho}^j - (\rho^j - d/w)] > w[\rho^j - (\rho^j - d/w)] = d$, therefore

$$d = q_{i+1}^e = \min\{v_f \rho_i^e, \tilde{w}(\tilde{\rho}^j - \rho_i^e), C, w(\rho^j - \rho_{i+1}^e)\} = w(\rho^j - \rho_{i+1}^e),$$

which gives that $\rho_{i+1}^e = \rho^j - d/w$. By mathematical induction, $\rho_k^e = \rho^j - \frac{d}{w}$, for all $i < k \le N$.

For any $i = 2,3,\ldots,N$, if $\frac{d}{v_f} \le \rho_i^e < \rho^j - \frac{d}{w}$, we have that $w(\rho^j - \rho_i^e) > d$, therefore

$$d = q_i^e = \min\{v_f \rho_{i-1}^e, \tilde{w}(\tilde{\rho}^j - \rho_{i-1}^e), C, w(\rho^j - \rho_i^e)\} = \min\{v_f \rho_{i-1}^e, \tilde{w}(\tilde{\rho}^j - \rho_{i-1}^e)\}.$$

If $q_i^e = \tilde{w}(\tilde{\rho}^j - \rho_{i-1}^e) = d$, then $q_{i-1}^e \le w(\rho^j - \rho_{i-1}^e) < \tilde{w}(\tilde{\rho}^j - \rho_{i-1}^e) < d$, which contradicts the fact that $q_{i-1}^e = d$, therefore $q_i^e = v_f \rho_{i-1}^e = d, \rho_{i-1}^e = \frac{d}{v_f}$. By mathematical induction, $\rho_k = \frac{d}{v_f}$, for all $1 \le k < i$. 0□

Using the above two lemmas the proof of Theorem 1.4 is completed as follows:

Proof of Theorem 1.4:

From the part a) of Lemma Lemma 4.1, we know that for $I \in \bigcup_{i=1}^{4} \Omega_i$ if ρ^e is an equilibrium state of system (51), then $\rho^e \in S^I$ and $d(\rho(t), S^I)$ converges to 0 exponentially fast, $\forall \rho(0) \in S$. Therefore, we only need to find all equilibrium states of system (51) in S^I and analyze the dynamics of $\rho(t)$ for all $\rho(0) \in S_\delta^I$, where

$$S_\delta^I = \{\rho | \frac{\min\{d, C\}}{v_f} - \delta \leq \rho_i \quad \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta, i = 1,2,\dots,N\}$$

and $\delta > 0$ can be arbitrarily small.

From the part b) of Lemma 4.1, we know that when $I \in \Omega_5$, we only need to find all equilibrium states of system (51) in the set $\bar{S}^I$ and analyze the dynamics of $\rho(t)$ for all $\rho(0)$ in the set $\bar{S}_\delta^I$, where

$$\bar{S}_\delta^I = \{\rho | \frac{\min\{d, C\}}{v_f} - \delta \leq \rho_i \leq \rho_c + \delta, i = 1,2,\dots,N\}.$$

Now we prove the statements of Theorem 1.4 from a) to e) respectively.

a) When $I \in \Omega_1, d < (1-\epsilon_0)C_d$. By Lemma 4.1, we have that $\frac{\min\{d,C\}}{v_f} \leq \rho_i^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, therefore $w(\rho^j - \rho_i^e) \geq (1-\epsilon_0)C_d > d$, $i = 1,2,\dots,N$. Thus $q_1^e = \min\{d, C, w(\rho^j - \rho_1^e)\} = d$. Using the equilibrium condition (53) we have

$$q_i^e = d, \text{ for } i = 1,2,\dots,N+1$$

Now we show that $\rho_i^e = d/v_f$, for $i = 1,2,\dots,N$. For $i = 1,2,\dots,N-1$, $q_{i+1}^e = \min\{v_f\rho_i^e, \tilde{w}(\tilde{\rho}^j - \rho_i^e), C, w(\rho^j - \rho_{i+1}^e)\}$. If $\rho_c \leq \rho_i^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, we have

$$w(\rho^j - \rho_{i+1}^e) \geq (1-\epsilon_0)C_d > d,$$

$$v_f\rho_i^e \geq C \geq \tilde{w}(\tilde{\rho}^j - \rho_i^e) \geq w(\rho^j - \rho_i^e) > (1-\epsilon_0)C_d > d,$$

which implies that $q_{i+1}^e > d$, therefore the assumption $\rho_c \leq \rho_i^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}$ is invalid. Hence $d/v_f \leq \rho_i^e < \rho_c$, which gives that $\tilde{w}(\tilde{\rho}^j - \rho_i^e) > C > v_f\rho_i^e$, thus

$$q_{i+1}^e = \min\{v_f\rho_i^e, w(\rho^j - \rho_{i+1}^e)\}.$$

By Lemma 4.1, we have $\rho_{i+1}^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, thus $w(\rho^j - \rho_{i+1}^e) \geq (1-\epsilon_0)C_d > d$. Solving the equation $q_{i+1}^e = d$ gives the unique equilibrium density $\rho_i^e = d/v_f$. Therefore, we have $\rho_i^e = d/v_f, i = 1,\dots,N-1$. For $i = N$, we have $q_{N+1}^e = \min\{v_f\rho_N^e, \tilde{w}(\tilde{\rho}^j - \rho_N^e), (1-\epsilon(\rho_N^e))C_d\}$. If $C_d/v_f < \rho_N^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, we have $q_{N+1}^e = (1-\epsilon_0)C_d > d$, therefore the

assumption $C_d/v_f < \rho_N^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$ is invalid, which together with Lemma 4.1 implies that $d/v_f \le \rho_N^e \le C_d/v_f$. Therefore $q_{N+1}^e = v_f \rho_N^e$. Solving the equation $q_{N+1}^e = d$ gives a unique solution $\rho_N^e = d/v_f$. Therefore, the point $\frac{d}{v_f} \times \mathbf{1}$ is the unique equilibrium state of system (51) when $I \in \Omega_1$.

Using Lemma 4.1, we have that for all $\rho(0) \in S_\delta^I$, $d/v_f - \delta < \rho_i(t) < \rho^j - (1-\epsilon_0)C_d/w + \delta, \forall t \ge 0, \delta > 0$. Therefore

$$w(\rho^j - \rho_i) \quad > w\left[\rho^j - \left(\rho^j - (1-\epsilon_0)C_d/w + \delta\right)\right] = (1-\epsilon_0)C_d - w\delta$$

and

$$\begin{aligned}\widetilde{w}(\tilde{\rho}^j - \rho_i) \quad &> \widetilde{w}\left[\tilde{\rho}^j - \left(\rho^j - (1-\epsilon_0)C_d/w + \delta\right)\right] > w\left[\rho^j - \left(\rho^j - (1-\epsilon_0)C_d/w + \delta\right)\right] \\ &= (1-\epsilon_0)C_d - w\delta,\end{aligned}$$

for $i = 1,2,\dots,N$. Since $(1-\epsilon_0)C_d > d$, taking δ to be sufficiently small, we have $(1-\epsilon_0)C_d - w\delta > d$. Therefore $q_1 = \min\{d, C, w(\rho^j - \rho_1)\} = d$ and

$$\dot{\rho}_1 \quad = q_1 - q_2 = d - \min\{v_f\rho_1, \widetilde{w}(\tilde{\rho}^j - \rho_1), C, w(\rho^j - \rho_2)\}. \tag{175}$$

Combine (51) and (175), we can show that

$$\dot{\rho}_1 \begin{cases} = -v_f\left(\rho_1 - \dfrac{d}{v_f}\right) & \text{if } \dfrac{d}{v_f} - \delta < \rho_1 < \dfrac{d}{v_f} \\ = 0 & \text{if } \rho_1 = \dfrac{d}{v_f} \\ \le -\alpha\left(\rho_1 - \dfrac{d}{v_f}\right) & \text{if } \dfrac{d}{v_f} < \rho_1 < \rho^j - (1-\epsilon_0)C_d/w + \delta \end{cases},$$

where $\alpha = \min\{v_f, \frac{(1-\epsilon_0)C_d - w\delta - d}{\rho^j - d/v_f}\} > 0$, which implies that for all $\rho(0) \in S_\delta^I$, $\rho_1(t)$ converges to d/v_f exponentially fast.

Based on the convergence of ρ_1, we can show that ρ_2 also converges to d/v_f exponentially fast, followed by ρ_3 through ρ_N. Therefore, $\forall \rho(0) \in S$, $\rho(t)$ converges to $\frac{d}{v_f} \times \mathbf{1}$ exponentially fast.

b) When $I \in \Omega_2$, $d = (1-\epsilon_0)C_d$. Using Lemma 4.2, we have that the equilibrium flow rate $q_i^e = d = (1-\epsilon_0)C_d$, for $i = 1,2,\dots,N+1$. If $0 \le \rho_N^e \le \frac{C_d}{v_f}$, then $q_{N+1}^e = d$ gives $\rho_N^e = \frac{d}{v_f}$. By part b) of Lemma 4.2, we have $\rho_i^e = \frac{d}{v_f}$, $i = 1,2,\dots,N-1$. Therefore $\rho^e = d/v_f \times \mathbf{1}$ is a potential equilibrium point of system (51). Substituting $\rho^e = d/v_f \times \mathbf{1}$ into equation (51), we

have $q_i^e = d$, for $i = 1,2,\ldots,N+1$. Therefore, $\rho^e = d/v_f \times \mathbf{1}$ is the only equilibrium state in the region $0 \le \rho_N^e \le \frac{C_d}{v_f}$.

If $\rho_N^e \in \left(\frac{C_d}{v_f}, \rho^j - \frac{d}{w}\right)$, according to Lemma 4.2, we have that $\rho_1^e = \ldots = \rho_{N-1}^e = \frac{d}{v_f}$. Substituting any $\rho^e \in \{\rho | \rho_1 = \ldots = \rho_{N-1} = \frac{d}{v_f}, \frac{C_d}{v_f} < \rho_N < \rho^j - \frac{d}{w}\}$ into equation (51) we have $q_i^e = d$, for $i = 1,2,\ldots,N+1$. Therefore all $\rho^e \in \{\rho | \rho_1 = \ldots = \rho_{N-1} = \frac{d}{v_f}, \frac{C_d}{v_f} < \rho_N < \rho^j - \frac{d}{w}\}$ are equilibrium states of system (51).

If $\rho_N^e = \rho^j - \frac{d}{w}$, we find all the equilibrium states of the system (51) by considering the following two cases:

Case I: for all $i = 1,2,..,N-1$, $\rho_i^e = \rho^j - \frac{d}{w}$;

Case II: there exists $i \in \{1,2,\ldots,N-1\}$, $d/v_f \le \rho_i^e < \rho^j - \frac{d}{w}$ and $\rho_{i+1}^e = \rho^j - \frac{d}{w}$.

Case I contains only one point, that is, $\rho^e = \left(\rho^j - \frac{d}{w}\right) \times \mathbf{1}$. Substituting this density state into equation (51), we have $q_i^e = d$, for $i = 1,2,\ldots,N+1$. Therefore $\left(\rho^j - \frac{d}{w}\right) \times \mathbf{1}$ is an equilibrium state of system (51).

For case II, it is clear from Lemma 4.2 that $\rho_1^e = \ldots \rho_{i-1}^e = d/v_f, \rho_{i+1}^e = \ldots = \rho_N^e = \rho^j - \frac{d}{w}$. Taking $i = 1,2,\ldots,N-1$, we have that all potential equilibrium points of system (51) in case II are in the set $\bigcup_{i=1}^{N-1}\{\rho | \frac{d}{v_f} \le \rho_i < \rho^j - \frac{d}{w}, \rho_k = \frac{d}{v_f}, 1 \le k < i, \rho_r = \rho^j - \frac{d}{w}, i < r \le N\}$. Substituting any point in this set into equation (51), we have $q_i^e = d$, for $i = 1,2,\ldots,N+1$. Therefore all $\rho^e \in \bigcup_{i=1}^{N-1}\{\rho | \frac{d}{v_f} \le \rho_i < \rho^j - \frac{d}{w}, \rho_k = \frac{d}{v_f}, 1 \le k < i, \rho_r = \rho^j - \frac{d}{w}, i < r \le N\}$ are equilibrium states of system (51).

To summarize, when $I \in \Omega_2$, system (51) has an isolated equilibrium state $\frac{d}{v_f} \times \mathbf{1}$ and an equilibrium manifold

$$S^e = \{\left(\rho^j - \frac{d}{w}\right) \times \mathbf{1}\} \cup \{\rho | \rho_i = \frac{d}{v_f}, i = 1,2,\ldots,N-1, \frac{C_d}{v_f} < \rho_N < \rho^j - \frac{d}{w}\}$$
$$\cup \left[\bigcup_{i=1}^{N-1}\{\rho | \frac{d}{v_f} \le \rho_i < \rho^j - \frac{d}{w}, \rho_k = \frac{d}{v_f}, 1 \le k < i, \rho_r = \rho^j - \frac{d}{w}, i < r \le N\}\right].$$

We now prove the rest of part b) as follows: first we show that for all $\rho(0)$ in the feasible space S, $\rho(t)$ converges to one equilibrium state in $\bar{S}^e$, where $\bar{S}^e = S^e \cup \{\frac{d}{v_f} \times \mathbf{1}\}$. Then we show that $\frac{d}{v_f} \times \mathbf{1}$ is locally exponentially stable, and that every $\rho^e \in S^e$ is stable in the sense of Lyapunov,

i.e., $\forall \mu > 0, \exists \eta > 0$, such that $\forall \rho(0)$ that satisfy $\| \rho(0) - \rho^e \| < \eta$, we have $\| \rho(t) - \rho^e \| \leq \mu, \forall t > 0$. Furthermore, $\rho(t)$ converges to some $\bar{\rho}^e \in S^e$ that satisfies $\| \bar{\rho}^e - \rho^e \| < \mu$.

For all $\rho(0) \in S^I$, by letting $\delta = 0$ in the proof of Lemma 4.1, we can show $\rho(t) \in S^I, \forall t \geq 0$. From equation (51), we know that $\forall \rho \in S^I$, $q_1 = d$ and $q_i \geq d, i = 2,3,\dots,N$. Therefore,

$$\sum_{i=1}^{k} \dot{\rho}_i = q_1 - q_k \leq 0, k = 1,2,\dots,N.$$

Thus $\sum_{i=1}^{k} \rho_i$ is monotonically decreasing but bounded from below which implies that it converges to a limit. Therefore, we have $\rho = [\rho_1, \rho_2, \dots, \rho_N]^T$ converges to a constant vector ρ^e. From equation (51) we know that $\dot{\rho}$ is a piecewise uniformly continuous function of ρ, therefore, as ρ converges to a constant ρ^e, $\dot{\rho}$ also converges to a constant, which has to be 0 (otherwise $\| \rho \|$ goes to infinity). Therefore ρ^e is a equilibrium point of system (51) by definition. Thus $\rho^e \in \bar{S}^e$. From part a) of Lemma 4.1, for all $\rho(0) \in S$, $d(\rho(t), S^I)$ converges to 0 exponentially fast. Therefore $\forall \rho(0) \in S$, $\rho(t)$ converges to a equilibrium point $\rho^e \in \bar{S}^e$.

Next we show that the equilibrium state $\rho^e = \frac{d}{v_f} \times \mathbf{1}$ is exponentially stable, and that every $\rho^e \in S^e$ is stable in the sense of Lyapunov.

(1) When $\rho^e = \frac{d}{v_f} \times \mathbf{1}$, then for all $\rho(0) \in \{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\dots,N\}$, $q_i \leq C_d$ and $q_{i+1} = v_f \rho_i$, for $i = 1,2,\dots,N$. Thus $\dot{\rho}_\iota = q_i - q_{i+1} \leq C_d - v_f \rho_i$, which implies that $\rho_i(t) \leq C_d/v_f, \forall t \geq 0$. Therefore,

$$\dot{\rho}_1 = d - v_f \rho_1,$$

$$\dot{\rho}_\iota = v_f \rho_{i-1} - v_f \rho_i, i = 2,3,\dots,N,$$

which can be written in the compact form as

$$\dot{\rho} = A\left(\rho - \frac{d}{v_f} \times \mathbf{1}\right)$$

where

$$A = \begin{bmatrix} -v_f & & & \\ v_f & -v_f & & \\ & \ddots & \ddots & \\ & & v_f & -v_f \end{bmatrix}.$$

Since $v_f > 0$, we have that A is Hurwitz. Therefore $\rho(t)$ converges to $\frac{d}{v_f} \times \mathbf{1}$ exponentially fast. ρ^e is in the interior of the set $\{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\dots,N\}$, thus we can always find a $\eta > 0$, such that $\{\rho | \| \rho - \rho^e \| < \eta\} \subset \{\rho | 0 \leq \rho_i \leq C_d/v_f, i = 1,2,\dots,N\}$. Therefore, $\forall \rho(0) \in$

$\{\rho | \| \rho - \rho^e \| < \eta\}$, $\rho(t)$ converges to ρ^e exponentially fast, which implies that $\rho^e = \frac{d}{v_f} \times \mathbf{1}$ is exponentially stable.

(2) When $\rho^e = \left(\rho^j - \frac{d}{w}\right) \times \mathbf{1}$, then with $\eta > 0$ sufficiently small, equation (51) gives the flow rates as follows: $q_1 = \min\{d, w(\rho^j - \rho_1)\}$, $q_i = w(\rho^j - \rho_i)$, for $i = 2, \ldots, N$, and $q_{N+1} = (1 - \epsilon_0)C_d = d$. Therefore,

$$\begin{aligned} \dot{\rho}_1 &= \min\{d, w(\rho^j - \rho_1)\} - w(\rho^j - \rho_2), \\ \dot{\rho}_i &= w(\rho^j - \rho_i) - w(\rho^j - \rho_{i+1}), i = 2, \ldots, N-1, \\ \dot{\rho}_N &= w(\rho^j - \rho_N) - d. \end{aligned}$$

Let $e_i = \rho_i - \rho_i^e$ and $\bar{e} = [e_2, e_3, \ldots, e_N]^T$, then we have $e = [e_1, \bar{e}^T]^T$, where

$$\dot{e}_1 = \begin{cases} we_2, & e_1 \leq 0 \\ -we_1 + we_2, & e_1 > 0 \end{cases} \tag{176}$$

and

$$\dot{\bar{e}} = \begin{bmatrix} -w & w & & \\ & -w & w & \\ & & \ddots & \ddots \\ & & & -w \end{bmatrix} \bar{e}.$$

Since $w > 0$, it follows that $\bar{e}$ converges to 0 exponentially fast, i.e., there exists constants $\alpha, \beta > 0$, such that

$$|e_i(t)| \leq |e_i(0)| \alpha \exp(-\beta t), i = 2,3, \ldots, N. \tag{177}$$

From (176), we have that $\dot{e}_1 \leq -we_1 + we_2$, which together with the continuity of $\dot{e}_1$ implies that

$$e_1(t) \leq |e_1(0)| \exp(-wt) + |e_2(0)| \frac{\alpha w}{w - \beta_0} [\exp(-\beta_0 t) - \exp(-wt)], \tag{178}$$

where $0 < \beta_0 < \min\{w, \beta\}$. Therefore, e_1 is bounded from above by a function that decays exponentially fast to 0 with time. If $\forall t \geq 0, e_1(t) > 0$, then $e_1(t)$ converges exponentially fast to 0. Otherwise, if $\exists t_0 \geq 0$, such that $e_1(t_0) \leq 0$, then we have the following cases:

Case I: If $e_1(0) \leq 0$, then as long as $e_1(t) \leq 0$, we have

$$e_1(t) = e_1(0) + \int_0^t w\, e_2(\tau) d\tau \geq -|e_1(0)| - \int_0^t w\, |e_2(\tau)| d\tau \geq -|e_1(0)| - \frac{\alpha w}{\beta} |e_2(0)| \exp(-\beta t),$$

which implies that for any given $\epsilon > 0$, there exists a finite time T such that

$$-|e_1(0)| - \epsilon \leq e_1(t) \leq \epsilon, \forall t \geq T,$$

which implies that the equilibrium $e_1 = 0$ is stable in the sense of Lyapunov.

Case II: If $e_1(0) > 0$, note that $e_1(t_0) \leq 0$, then due to the uniform continuity of e_1, we have that $\exists t_1 \in (0, t_0]$, such that $e_1(t_1) = 0$. Then $\forall t \geq t_1$, as long as $e_1(t) \leq 0$, we have

$$\begin{aligned} e_1(t) &= \int_{t_1}^{t} w\, e_2(\tau) d\tau \geq -\int_{t_1}^{t} w\, |e_2(\tau)| d\tau \geq -\frac{\alpha w}{\beta} |e_2(0)| \exp(-\beta t_1)\left[1 - \exp\left(-\beta(t - t_1)\right)\right] \\ &\geq -\frac{\alpha w}{\beta} |e_2(0)| \exp(-\beta t_1), \end{aligned}$$

which implies that for any given $\epsilon > 0$, there exists a finite time $T \geq t_1$ such that

$$-\frac{\alpha w}{\beta} |e_2(0)| \exp(-\beta t_1) \leq e_1(t) \leq \epsilon, \forall t \geq T,$$

which implies that the equilibrium $e_1 = 0$ is stable in the sense of Lyapunov.

To summarize the above analysis, we have that the equilibrium state $\rho^e = \left(\rho^j - \frac{d}{w}\right) \times \mathbf{1}$ is stable in the sense of Lyapunov.

(3) When $\rho^e \in \{\rho | \frac{d}{v_f} \leq \rho_i < \rho^j - \frac{d}{w}, \rho_k = \frac{d}{v_f}, 1 \leq k < i, \rho_r = \rho^j - \frac{d}{w}, i < r \leq N\}$, $i = 1,2,\ldots,N-1$, $\forall \rho(0)$ that satisfy $\| \rho(0) - \rho^e \| < \eta$, if η is sufficiently small, we can get the flow rates from equation (51) as follows: $q_1 = d$, $q_k = v_f \rho_{k-1}, k = 2,\ldots,i$, $q_r = w\left(\rho^j - \rho_r\right), r = i+1,\ldots,N$, and $q_{N+1} = (1 - \epsilon_0) C_d = d$. Therefore, we have

$$\begin{aligned} \dot{\rho}_1 &= d - v_f \rho_1, \\ \dot{\rho}_k &= v_f \rho_{k-1} - v_f \rho_k, k = 2,\ldots,i-1, \\ \dot{\rho}_i &= v_f \rho_{i-1} - w\left(\rho^j - \rho_{i+1}\right), \\ \dot{\rho}_r &= w\left(\rho^j - \rho_r\right) - w\left(\rho^j - \rho_{r+1}\right), r = i+1,\ldots,N-1, \\ \dot{\rho}_N &= w\left(\rho^j - \rho_N\right) - d. \end{aligned}$$

Let $e_i = \rho_i - \rho_i^e$ and $e = [e_1, e_2, \ldots, e_N]^T$, the dynamics of e can be presented in the compact form as follows:

$$\dot{e} = Ae,$$

where

$$A=\begin{bmatrix} -v_f & & & & & & 0 \\ v_f & -v_f & & & & & 0 \\ & \ddots & \ddots & & & & \ddots \\ & & v_f & 0 & & & w \\ & & & 0 & -w & & w \\ & & & & \ddots & \ddots & \ddots \\ & & & & & 0 & -w \end{bmatrix},$$

Let $\bar{e}=[e_1,e_2,\ldots,e_{i-1}]^T$ and $\underset{-}{e}=[e_{i+1},\ldots,e_N]^T$, then

$$\dot{\bar{e}}=\begin{bmatrix} v_f & & & -v_f \\ & \ddots & & -v_f \\ & & v_f & \ddots \\ & & & -v_f \end{bmatrix}\bar{e}$$

and

$$\dot{\underset{-}{e}}=\begin{bmatrix} -w & & & w \\ & \ddots & & \ddots \\ & & -w & w \\ & & & -w \end{bmatrix}\underset{-}{e}.$$

The above two subsystems are both linear and exponentially stable. Thus $\bar{e}$ and $\underset{-}{e}$ both converge to 0 exponentially fast. $\dot{e}_i=v_f e_{i-1}+we_{i+1}$, therefore

$$e_i(t)=e_i(0)+v_f\int_0^t e_{i-1}(\tau)d\tau+w\int_0^t e_{i+1}(\tau)d\tau.$$

Since

$$\begin{aligned} \| e_{i-1} \| &\le \| \bar{e} \| \le \| \bar{e}(0) \| \alpha_1 \exp(-\beta_1 t), \\ \| e_{i+1} \| &\le \| \underset{-}{e} \| \le \| \underset{-}{e}(0) \| \alpha_2 \exp(-\beta_2 t), \end{aligned}$$

where $\alpha_1,\alpha_2,\beta_1,\beta_2>0$, thus

$$\begin{aligned} e_i &= e_i(0)+\int_0^t \left(v_f e_{i-1}+we_{i+1}\right)d\tau, \\ \| e_i \| &\le \| e_i(0) \| +\int_0^t \left(v_f \| e_{i-1} \| +w \| e_{i+1} \|\right)d\tau \\ &\le \| e_i(0) \| +v_f\int_0^t \| \bar{e}(0) \| \alpha_1\exp(-\beta_1 t)d\tau+w\int_0^t \| \underset{-}{e}(0) \| \alpha_2\exp(-\beta_2 t)d\tau \\ &= \| e_i(0) \| +\| \bar{e}(0) \| \frac{v_f\alpha_1}{\beta_1}\left(1-\exp(-\beta_1 t)\right)+\| \underset{-}{e}(0) \| \frac{w\alpha_2}{\beta_2}\left(1-\exp(-\beta_2 t)\right). \end{aligned}$$

Since we have shown before that $\forall\rho(0) \in S, \rho(t)$ converges to a constant, i.e., the limit $\lim_{t\to\infty}\rho(t)$ exists, which implies that the limit $\lim_{t\to\infty} \| e_i \|$ also exists. Therefore,

$$\lim_{t\to\infty} \| e_i \| \leq \| e_i(0) \| + \| \bar{e}(0) \| \frac{v_f\alpha_1}{\beta_1} + \| e(0) \| \frac{w\alpha_2}{\beta_2}.$$

For all $\mu > 0$, by selecting $\rho(0)$ sufficiently close to ρ^e, we have $\| \rho(t) - \rho^e \| \leq \mu, \forall t \geq 0$, i.e., all equilibrium points $\rho^e \in \{\rho | \frac{d}{v_f} \leq \rho_i < \rho^j - \frac{d}{w}, \rho_k = \frac{d}{v_f}, 1 \leq k < i, \rho_r = \rho^j - \frac{d}{w}, i < r \leq N\}$, $i = 1,2,\dots,N-1$ are stable in the sense of Lyapunov.

(4) When $\rho^e \in \{\rho | \rho_i = d/v_f, i = 1,2,\dots,N-1, \frac{C_d}{v_f} < \rho_N < \rho^j - \frac{d}{w}\}$, $\forall\rho(0)$ that satisfy $\| \rho(0) - \rho^e \| < \eta$, if η is sufficiently small, we get the flow rates from equation (51) as: $q_1 = d$, $q_i = v_f\rho_{i-1}, i = 2,3,\dots,N$, and $q_{N+1} = (1-\epsilon_0)C_d = d$. Therefore we have that

$$\begin{aligned} \dot{\rho}_1 &= d - v_f\rho_1, \\ \dot{\rho}_i &= v_f\rho_{i-1} - v_f\rho_i, i = 2,\dots,N-1, \\ \dot{\rho}_N &= v_f\rho_{N-1} - d. \end{aligned}$$

Let $e_i = \rho_i - \rho_i^e$ and $e = [e_1, e_2, \dots, e_N]^T$, the dynamics of e can be expressed in the compact form as follows:

$$\dot{e} = Ae,$$

where

$$A = \begin{bmatrix} -v_f & & & & \\ v_f & & & & -v_f \\ & \ddots & & & \ddots \\ & & v_f & & -v_f \\ & & & v_f & 0 \end{bmatrix}.$$

The stability of ρ^e can be shown by following a similar analysis as in previous case.

Therefore, all $\rho^e \in S^e$ are stable in the sense of Lyapunov. Recall that $\forall\rho(0) \in S, \rho(t)$ converges to an equilibrium state in S^e, thus $\forall\mu > 0, \exists\eta > 0$, such that $\forall\rho(0)$ that satisfy $\| \rho(0) - \rho^e \| < \eta$, $\rho(t)$ converges to some $\bar{\rho}^e \in S^e$ that satisfies $\| \bar{\rho}^e - \rho^e \| < \mu$.

c) For the case $I \in \Omega_3$, from part a) of Lemma 4.1, we know $\frac{d}{v_f} \leq \rho_i^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}, i = 1,2,\dots,N$. If $\frac{d}{v_f} \leq \rho_1^e \leq \rho^j - \frac{d}{w}$, then $w(\rho^j - \rho_1^e) \geq d$, thus $q_1^e = \min\{d, C, w(\rho^j - \rho_1^e)\} = d$ and $q_i^e = q_1^e = d, i = 2,3,\dots,N+1$, according to the equilibrium condition. Solving the

equation $q_{N+1}^e = d$ gives only one solution $\rho_N^e = d/v_f$. For $i = 1,2,\dots,N-1$, given $\rho_{i+1}^e = d/v_f, q_i^e = q_{i+1}^e = d$, we have $d = q_i^e \le w(\rho^j - \rho_i^e)$, thus $\rho_i^e \le \rho^j - d/w$. Since

$$q_{i+1}^e = \min\{v_f\rho_i^e, \widetilde{w}(\tilde{\rho}^j - \rho_i^e), C, w(\rho^j - \rho_{i+1}^e)\}$$

and

$$w(\rho^j - \rho_{i+1}^e) > C > d, \text{ as } \rho_{i+1}^e = d/v_f < \rho_c,$$

$$\widetilde{w}(\tilde{\rho}^j - \rho_i^e) \ge \widetilde{w}[\tilde{\rho}^j - (\rho^j - d/w)] > w[\rho^j - (\rho^j - d/w)] = d.$$

The equation $q_{i+1}^e = d$ gives only one solution, that is $v_f\rho_i^e = d$,

$$\rho_i^e = d/v_f.$$

By mathematical induction, we have that $\rho_i^e = d/v_f, i = 1,2,\dots,N$. Therefore in the region $\frac{d}{v_f} \le \rho_1^e \le \rho^j - \frac{d}{w}$, system (51) has only one equilibrium state $\frac{d}{v_f} \times \mathbf{1}$.

If $\rho^j - \frac{d}{w} < \rho_1^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, we have $w(\rho^j - \rho_1^e) < d < C$, thus

$$q_1^e = \min\{d, C, w(\rho^j - \rho_1^e)\} = w(\rho^j - \rho_1^e) < d.$$

For $i = 2,3,\dots,N$, given $\rho^j - \frac{d}{w} < \rho_{i-1}^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$ and $q_{i-1}^e = q_i^e = q_1^e = w(\rho^j - \rho_1^e)$, then we have

$$q_i^e = \min\{v_f\rho_{i-1}^e, \widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e), C, w(\rho^j - \rho_i^e)\} = w(\rho^j - \rho_1^e).$$

Since $v_f\rho_{i-1}^e > C > d$ as $\rho_{i-1}^e > \rho_c$, $v_f\rho_{i-1}^e \ne w(\rho^j - \rho_1^e)$. If $\widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e) = w(\rho^j - \rho_1^e)$, then

$$q_{i-1}^e \le w(\rho^j - \rho_{i-1}^e) < \widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e) = w(\rho^j - \rho_1^e),$$

which contradicts the fact that $q_{i-1}^e = w(\rho^j - \rho_1^e)$, therefore $\widetilde{w}(\tilde{\rho}^j - \rho_{i-1}^e) \ne w(\rho^j - \rho_1^e)$. Thus we have

$$q_i^e = w(\rho^j - \rho_i^e) = w(\rho^j - \rho_1^e) \text{ and } \rho_i^e = \rho_1^e,\ i = 2,3,\dots,N.$$

Therefore $\rho^j - \frac{d}{w} < \rho_N^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, equation (51) gives that $q_{N+1}^e = (1-\epsilon_0)C_d$. Using the equilibrium condition (53) we have $q_i^e = w(\rho^j - \rho_i^e) = (1-\epsilon_0)C_d$ for $i = 1,2,\dots,N$, which gives only one solution, that is, $\left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$ is the only equilibrium state of system (51) in the region $\rho^j - \frac{d}{w} < \rho_1^e \le \rho^j - \frac{(1-\epsilon_0)C_d}{w}$. To summarize, $\rho^{e1} = \frac{d}{v_f} \times \mathbf{1}$ and $\rho^{e2} = \left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$ are 2 isolated equilibrium states of system (51) when $I \in \Omega_3$.

Now we are going to show that for all $0 \le \rho(0) \le \rho^j$, $\rho(t)$ converges to either ρ^{e1} or ρ^{e2}. According to part a) of Lemma 4.1, we have that $\forall \delta > 0, \exists T > 0$, such that $d/v_f - \delta < \rho_i < \rho^j - (1-\epsilon_0)C_d + \delta, i = 1,2,\dots,N$. Without loss of generality, let $T = 0$.

If $\forall t \ge 0$, $\rho_N(t) \le \frac{C_d}{v_f}$, then $w(\rho^j - \rho_N) > C$, due to $\frac{C_d}{v_f} < \rho_c$. We have

$$q_N = \min\{v_f \rho_{N-1}, \widetilde{w}(\tilde{\rho}^j - \rho_{N-1})\},$$

$$q_{N-1} = \min\{v_f \rho_{N-2}, \widetilde{w}(\tilde{\rho}^j - \rho_{N-2}), C, w(\rho^j - \rho_{N-1})\} \le w(\rho^j - \rho_{N-1}).$$

Therefore, $\forall \rho_{N-1} > \rho_c$, $\dot{\rho}_{N-1} = q_{N-1} - q_N \le w(\rho^j - \rho_{N-1}) - \widetilde{w}(\tilde{\rho}^j - \rho_{N-1}) < 0$, thus $\limsup_{t\to\infty} \rho_{N-1} \le \rho_c$. Consequently, we have that $\limsup_{t\to\infty} \rho_i \le \rho_c, i - 1,2,\dots,N-1$. When $\rho_i^e \le \rho_c, i = 1,2,\dots,N-1$, $\rho_N \le C_d/v_f$, equation (51) gives that $q_1 = d, q_i = v_f \rho_{i-1}, i = 2,3,\dots,N+1$, therefore,

$$\begin{aligned} \dot{\rho}_1 &= d - v_f \rho_1, \\ \dot{\rho}_i &= v_f \rho_{i-1} - v_f \rho_i, i = 2,3,\dots,N. \end{aligned}$$

which can be written in the compact form as

$$\dot{\rho} = A\left(\rho - \frac{d}{v_f} \times \mathbf{1}\right)$$

where

$$A = \begin{bmatrix} -v_f & & & \\ v_f & -v_f & & \\ & \ddots & \ddots & \\ & & v_f & -v_f \end{bmatrix}.$$

Since $v_f > 0$, we have that A is Hurwitz. Therefore $\rho(t)$ converges to $\rho^{e1} = \frac{d}{v_f} \times \mathbf{1}$ exponentially fast.

If there exists $t_0 \ge 0$, $\rho_N(t_0) > \frac{C_d}{v_f}$, then $q_{N+1}(t_0) = (1-\epsilon_0)C_d$, due to $\frac{C_d}{v_f} < \rho_N(t_0) < \rho^j - (1-\epsilon_0)C_d + \delta$. Recall that

$$q_N = \min\{v_f \rho_{N-1}, \widetilde{w}(\tilde{\rho}^j - \rho_{N-1}), C, w(\rho^j - \rho_N)\}.$$

and $d/v_f - \delta < \rho_{N-1} < \rho^j - (1-\epsilon_0)C_d + \delta$, we have that

$$\begin{aligned} v_f\rho_{N-1} &> d - v_f\delta, \\ \widetilde{w}\left(\tilde{\rho}^j - \rho_{N-1}\right) &> \widetilde{w}\left(\tilde{\rho}^j - \left(\rho^j - (1-\epsilon_0)C_d + \delta\right)\right) \\ &= (w-\widetilde{w})\left(\rho^j - (1-\epsilon_0)C_d - \rho_c\right) + (1-\epsilon_0)C_d - \widetilde{w}\delta. \end{aligned}$$

Therefore

$$\dot{\rho}_N = q_N - q_{N+1}\begin{cases} > -\alpha_1\left[\rho_N - \left(\rho^j - \dfrac{(1-\epsilon_0)C_d}{w}\right)\right] > 0, & \text{if } \rho_N < \rho^j - \dfrac{(1-\epsilon_0)C_d}{w} \\ = 0, & \text{if } \rho_N = \rho^j - \dfrac{(1-\epsilon_0)C_d}{w}, \\ < -\alpha_2\left[\rho_N - \left(\rho^j - \dfrac{(1-\epsilon_0)C_d}{w}\right)\right] < 0, & \text{if } \rho_N > \rho^j - \dfrac{(1-\epsilon_0)C_d}{w} \end{cases} \tag{179}$$

where

$$\alpha_1 = \min\{\frac{d - v_f\delta - (1-\epsilon_0)C_d}{\rho^j - (1-\epsilon_0)C_d/w}, \frac{(w-\widetilde{w})\left(\rho^j - (1-\epsilon_0)C_d - \rho_c\right) - \widetilde{w}\delta}{\rho^j - (1-\epsilon_0)C_d/w}, w\}$$

and

$$\alpha_2 = \min\{\frac{d - v_f\delta - (1-\epsilon_0)C_d}{(1-\epsilon_0)C_d/w}, \frac{(w-\widetilde{w})\left(\rho^j - (1-\epsilon_0)C_d - \rho_c\right) - \widetilde{w}\delta}{(1-\epsilon_0)C_d/w}, w\}.$$

When δ is sufficiently small, α_1 and α_2 are both positive. Therefore $\rho_N(t) > \frac{C_d}{v_f}, \forall t \geq t_0$ and $\rho_N(t)$ converges to $\rho^j - \frac{(1-\epsilon_0)}{w}$ exponentially fast. Consequently, $\rho_i(t)$ also converges exponentially fast to $\rho^j - \frac{(1-\epsilon_0)}{w}$ for $i = 1,2,\ldots,N-1$, that is, $\rho(t)$ converges exponentially fast to ρ^{e2}. Therefore, for all initial condition $\rho(0) \in S$, $\rho(t)$ converges to one of the two equilibrium states exponentially fast.

From the analysis above, we have that $\forall\rho(0) \in \{\rho|\frac{d}{v_f} - \delta \leq \rho_i \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta, i = 1,2,\ldots,N-1, \frac{C_d}{v_f} < \rho_N \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w} + \delta\}$, $\rho(t)$ converges to the equilibrium state ρ^{e2} exponentially fast. Therefore this equilibrium state is locally exponentially stable.

Similar to the case $I \in \Omega_2$, we can show that for all $\rho(0) \in \{\rho|0 \leq \rho_i \leq C_d/v_f, i = 1,2,\ldots,N\}$, $\rho(t)$ converges to the point ρ^{e1} exponentially fast. Therefore this equilibrium state is exponentially stable.

d) When $I \in \Omega_4$, $d > C_b$. If ρ^e is an equilibrium state of system (51) then we have $\frac{\min\{d,C\}}{v_f} \leq \rho_i^e \leq \rho^j - \frac{(1-\epsilon_0)C_d}{w}, i = 1,2,\ldots,N$ by using Lemma 4.1, in this region

$$q_{N+1}^e = \min\{v_f\rho_N^e, \widetilde{w}\left(\tilde{\rho}^j - \rho_N^e\right), \left(1-\epsilon(\rho_N^e)\right)C_d\} = (1-\epsilon_0)C_d,$$

From the equilibrium condition, we have that $q_i^e = (1-\epsilon_0)C_d, i = 1,2,\dots,N$. Recall that

$$q_1^e = \min\{d, C, w\left(\rho^j - \rho_1^e\right)\}.$$

Since $d > (1-\epsilon_0)C_d$ and $C > (1-\epsilon_0)C_d$, $q_1^e = (1-\epsilon_0)C_d$ gives only one solution $\rho_1^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$. For $i = 2,3,\dots,N$, given $\rho_{i-1}^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$, we check the value of ρ_i^e. Recall that

$$q_i^e = \min\{v_f\rho_{i-1}^e, \widetilde{w}\left(\tilde{\rho}^j - \rho_{i-1}^e\right), C, w\left(\rho^j - \rho_i^e\right)\}.$$

Since $v_f\rho_{i-1}^e > C > (1-\epsilon_0)C_d$ and $\widetilde{w}\left(\tilde{\rho}^j - \rho_{i-1}^e\right) > w\left(\rho^j - \rho_{i-1}^e\right) = (1-\epsilon_0)C_d$ as $\rho_{i-1}^e > \rho_c$, thus $q_i^e = (1-\epsilon_0)C_d$ gives $w\left(\rho^j - \rho_i^e\right) = (1-\epsilon_0)C_d$, i.e., $\rho_i^e = \rho^j - \frac{(1-\epsilon_0)C_d}{w}$. Therefore the point $\left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$ is the unique equilibrium state of system (51) when $I \in \Omega_4$.

For all $\rho(0) \in S_\delta^I$, we have $v_f\rho_N > d - v_f\delta$ and $\widetilde{w}\left(\tilde{\rho}^j - \rho_N\right) > \widetilde{w}\left(\tilde{\rho}^j - \rho^j + \frac{(1-\epsilon_0)C_d}{w} - \delta\right)$ by using Lemma 4.1. Take δ to be sufficiently small, we have $v_f\rho_N > C_d$ and $\widetilde{w}\left(\tilde{\rho}^j - \rho_N\right) > (1-\epsilon_0)C_d$. Thus

$$q_{N+1} = \min\{v_f\rho_N, \widetilde{w}\left(\tilde{\rho}^j - \rho_N\right), \left(1-\epsilon(\rho_N)\right)C_d\} = (1-\epsilon_0)C_d,$$

Then

$$\dot{\rho_N} \quad = q_N - q_{N+1} = \min\{v_f\rho_{N-1}, \widetilde{w}\left(\tilde{\rho}^j - \rho_{N-1}\right), C, w\left(\rho^j - \rho_N\right)\} - (1-\epsilon_0)C_d. \quad (180)$$

Similar to equation (179), we have

$$\dot{\rho}_N \begin{cases} > -\alpha_1\left[\rho_N - \left(\rho^j - \dfrac{(1-\epsilon_0)C_d}{w}\right)\right] > 0, & \text{if } \rho_N < \rho^j - \dfrac{(1-\epsilon_0)C_d}{w} \\ = 0, & \text{if } \rho_N = \rho^j - \dfrac{(1-\epsilon_0)C_d}{w}, \\ < -\alpha_2\left[\rho_N - \left(\rho^j - \dfrac{(1-\epsilon_0)C_d}{w}\right)\right] < 0, & \text{if } \rho_N > \rho^j - \dfrac{(1-\epsilon_0)C_d}{w} \end{cases}$$

where

$$\alpha_1 = \min\{\frac{d - v_f\delta - (1-\epsilon_0)C_d}{\rho^j - (1-\epsilon_0)C_d/w}, \frac{(w-\widetilde{w})\left(\rho^j - (1-\epsilon_0)C_d - \rho_c\right) - \widetilde{w}\delta}{\rho^j - (1-\epsilon_0)C_d/w}, w\}$$

and

$$\alpha_2 = \min\{\frac{d - v_f\delta - (1-\epsilon_0)C_d}{(1-\epsilon_0)C_d/w}, \frac{(w-\widetilde{w})\left(\rho^j - (1-\epsilon_0)C_d - \rho_c\right) - \widetilde{w}\delta}{(1-\epsilon_0)C_d/w}, w\}.$$

When δ is sufficiently small, α_1 and α_2 are both positive. Therefore ρ_N converges to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$ exponentially fast.

Based on the converges of ρ_N, we can show that ρ_{N-1} also convergences to $\rho^j - \frac{(1-\epsilon_0)C_d}{w}$, followed by ρ_{N-2} through ρ_1. Therefore, $\forall \rho(0) \in S$, $\rho(t)$ converges to $\left(\rho^j - \frac{(1-\epsilon_0)C_d}{w}\right) \times \mathbf{1}$ exponentially fast.

e) For the case $I \in \Omega_5$ The proof of this part can be demonstrated by following the same routine of the case of $I \in \Omega_1$ based on part b) of Lemma 4.1. For the sake of briefness, we omit the detailed proof here.

Appendix E: Proof of Theorem 1.5

a) If $I \in \bigcup_{i=1}^{4} \Omega_i$, the VSL controller (49) is applied. First we show that the controller v is well-defined $\forall \rho \in [0, \rho^j]$. According to (49), $\bar{v}_1$ is defined in the region $\frac{C_d}{v_f} - \delta_2 \le \rho \le \rho^j$, in which $q_2 \le C_d$ and $x + \delta_1 > 0$. Therefore the denominator of $\bar{v}_1$

$$w\rho^j - q_2 + \lambda(x + \delta_1) \quad \ge w\rho^j - C_d > w(\rho^j - \rho_c) - C_d = C - C_d > 0.$$

Hence $v_1 = \text{med}\{0, \bar{v}_1, v_f\}$ is well-defined in the region $\frac{C_d}{v_f} - \delta_2 \le \rho \le \rho^j$. $\bar{v}_2$ is defined in the region $0 \le \rho \le \frac{C_d}{v_f}$, in which $q_2 = v_f \rho = C_d + v_f x$. Since $0 < \lambda < \frac{v_f w \rho^j}{C_d}$ and $-\frac{C_d}{v_f} \le x \le 0$, we have that

$$q_2 - \lambda x = C_d + v_f x - \lambda x > C_d + v_f x \ge 0$$

and

$$w\rho^j - (q_2 - \lambda x) > w\rho^j - C_d - \left(v_f - \frac{v_f w \rho^j}{C_d}\right) x \ge w\rho^j - C_d - \left(v_f - \frac{v_f w \rho^j}{C_d}\right)\left(-\frac{C_d}{v_f}\right) = 0$$

due to $v_f - \frac{v_f w \rho^j}{C_d} < 0$. Therefore, the denominator of $\bar{v}_2$ is greater than 0, $v_2 = \text{med}\{0, \bar{v}_2, v_f\}$ is well-defined, and $\bar{v}_2 = \frac{w(q_2 - \lambda x)}{w\rho^j - (q_2 - \lambda x)} > 0$.

Now we find the equilibrium point of system (44)-(49) and analyze its stability properties. We have that $\forall \rho(0) \in (C_d/v_f, \rho^j], v = v_1$. If $v_1 = 0$, i.e., $\bar{v}_1 \le 0$, we have $q_1 = \frac{v_1 w \rho^j}{v_1 + w} = 0$. In the region $\frac{C_d}{v_f} - \delta_2 \le \rho \le \rho^j$, we have

$$\begin{aligned}
&v_f \rho \ge C_d - v_f \delta_2 > 0, \text{ as } \delta_2 < \frac{C_d}{v_f}, \\
&\tilde{w}(\tilde{\rho}^j - \rho) \ge \tilde{w}(\tilde{\rho}^j - \rho^j) > 0, \text{ as } \tilde{\rho}^j > \rho^j \ge \rho, \\
&(1 - \epsilon(\rho)) C_d \ge (1 - \epsilon_0) C_d > 0, \text{ as } \epsilon_0 < 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
q_2 \quad &= \min\{v_f \rho, \tilde{w}(\tilde{\rho}^j - \rho), (1 - \epsilon(\rho)) C_d\} \ge \min\{C_d - v_f \delta_2, \tilde{w}(\tilde{\rho}^j - \rho^j), (1 - \epsilon_0) C_d\} \\
&\ge \frac{\min\{C_d - v_f \delta_2, \tilde{w}(\tilde{\rho}^j - \rho^j), (1 - \epsilon_0) C_d\}}{\rho^j - C_d/v_f + \delta_1} \left(\rho - \frac{C_d}{v_f} + \delta_1\right)
\end{aligned}$$

due to $\frac{C_d}{v_f} - \delta_2 \le \rho \le \rho^j$, which implies $0 < \frac{\rho - C_d/v_f + \delta_1}{\rho^j - C_d/v_f + \delta_1} \le 1$, since $\delta_2 < \delta_1$. Thus we have

$$\dot{\rho} \quad = q_1 - q_2 \le -\frac{\min\{C_d - v_f\delta_2, \tilde{w}(\tilde{\rho}^j - \rho^j), (1-\epsilon_0)C_d\}}{\rho^j - C_d/v_f + \delta_1}\left(\rho - \frac{C_d}{v_f} + \delta_1\right) \tag{181}$$

If $v_1 > 0$, i.e., $\bar{v}_1 > 0$, $v_1 = \min\{\bar{v}_1, v_f\} \le \bar{v}_1$ and

$$\frac{v_1 w\rho^j}{v_1 + w} - \frac{\bar{v}_1 w\rho^j}{\bar{v}_1 + w} = w\rho^j \frac{(v_1 - \bar{v}_1)w}{(v_1 + w)(\bar{v}_1 + w)} \le 0,$$

which implies $\frac{v_1 w\rho^j}{v_1+w} \le \frac{\bar{v}_1 w\rho^j}{\bar{v}_1+w}$. Hence,

$$q_1 \quad = \min\{d, \frac{v_1 w\rho^j}{v_1 + w}, C, w(\rho^j - \rho)\} \le \frac{v_1 w\rho^j}{v_1 + w} \le \frac{\bar{v}_1 w\rho^j}{\bar{v}_1 + w} = q_2 - \lambda(x + \delta_1) \tag{182}$$

and

$$\dot{\rho} = q_1 - q_2 \le -\lambda\left(\rho - \frac{C_d}{v_f} + \delta_1\right) < 0.$$

According to equation (181)-(182)

$$\dot{\rho} \le -\alpha\left(\rho - \frac{C_d}{v_f} + \delta_1\right),$$

where $\alpha = \min\{\lambda, \frac{\min\{C_d - v_f\delta_2, \tilde{w}(\tilde{\rho}^j - \rho^j), (1-\epsilon_0)C_d\}}{\rho^j - C_d/v_f + \delta_1}\} > 0$. Using Lemma 3.2.4 in (P. A. Ioannou and Sun 2012), we have

$$\rho(t) \le \frac{C_d}{v_f} - \delta_1 + \left[\rho(0) - \frac{C_d}{v_f} + \delta_1\right] e^{-\alpha t}.$$

Since $C_d/v_f - \delta_1 < C_d/v_f - \delta_2 < C_d/v_f < \rho(0)$, $\rho(t)$ will decrease exponentially to the value $\rho(t_0) = C_d/v_f - \delta_2$ at some finite time t_0, at which v switches to v_2, in which case the dynamics of $\rho(t)$ are analyzed below.

Either the initial condition $0 \le \rho(0) \le \frac{C_d}{v_f}$ or v switches to v_2 from v_1, there exists a $t_0 \ge 0$, at which $0 \le \rho(t_0) \le \frac{C_d}{v_f}$ and $v = v_2$. Since $\bar{v}_2 > 0$, we have $v_2 = \min\{\bar{v}_2, v_f\} \le v_f$ and

$$\frac{v_2 w\rho^j}{v_2 + w} \le \frac{v_f w\rho^j}{v_f + w} = C < w(\rho^j - \rho) \text{ as } \rho \le C_d/v_f < \rho_c.$$

Therefore,

$$q_1 = \min\{d, \frac{v_2 w \rho^j}{v_2 + w}, C, w(\rho^j - \rho)\} = \min\{d, \frac{v_2 w \rho^j}{v_2 + w}\} = \min\{d, \frac{v_f w \rho^j}{v_f + w}, \frac{\bar{v}_2 w \rho^j}{\bar{v}_2 + w}\}$$
$$= \min\{d, \frac{\bar{v}_2 w \rho^j}{\bar{v}_2 + w}\} = \min\{d, q_2 - \lambda x\}$$

and $q_2 = C_d + v_f x$. Consequently,

$$\dot{\rho} = q_1 - q_2 = \min\{d - v_f \rho, -\lambda\left(\rho - \frac{C_d}{v_f}\right)\}. \quad (183)$$

In the case $d > C_d$, we have $\dot{\rho} \geq -\min\{v_f, \lambda\}\left(\rho - \frac{C_d}{v_f}\right)$, $\forall \rho \in [0, C_d/v_f]$ and $\dot{\rho} = 0$ at $\rho = C_d/v_f$, which implies that $\rho(t)$ converges exponentially fast to $\rho = \frac{C_d}{v_f}$, and $\forall t \geq t_0, \rho \leq \frac{C_d}{v_f}$, therefore the flow at the exit of the section $q_2 = v_f \rho$ converges to C_d. In the case $d \leq C_d$,

$$\dot{\rho} = \min\left\{d - v_f \rho, -\lambda\left(\rho - \frac{C_d}{v_f}\right)\right\} \begin{cases} \geq -\min\{v_f, \lambda\}\left(\rho - \frac{d}{v_f}\right) & \text{if } \rho \in \left[0, \frac{d}{v_f}\right) \\ = 0 & \text{if } \rho = \frac{d}{v_f} \\ \leq -v_f\left(\rho - \frac{d}{v_f}\right) & \text{if } \rho \in (\frac{d}{v_f}, \frac{C_d}{v_f}] \end{cases} \quad (184)$$

Therefore, $\rho(t)$ converges to $\frac{d}{v_f}$ exponentially fast, and $q_2 = v_f \rho$ converges to d with the same rate. In summary, the closed-loop system (44)-(49) has a unique equilibrium point $\rho^e = \frac{\min\{d, C_d\}}{v_f}$. In addition, $\forall \rho(0) \in \left[0, \frac{C_d}{v_f}\right]$, $\rho(t)$ converges to ρ^e exponentially fast and $\forall \rho(0) \in (\frac{C_d}{v_f}, \rho^j]$, $\rho(t)$ decreases to $\frac{C_d}{v_f} - \delta_2$ exponentially fast and then converges to ρ^e exponentially fast. The flow rate at the exit of the section converges to the maximum possible value $\min\{d, C_d\}$ exponentially fast while the speed of flow converges with the same rate to v_f.

b) Part b) of Theorem 1.5 can be derived directly from part e) of Theorem 1.3.

Appendix F: Proof of Theorem 1.6

a) In controller (56), v_1 through v_{N-1} is well-defined by letting $v_i = v_f$ when $\rho_i = 0$, for $i = 1,2,\dots,N-1$. Since $0 < \lambda_0 \le \frac{v_f w \rho^j}{C_d}$, we can show that v_0 is also well-defined in a similar manner to the single section case in Theorem 1.5.

For all $\rho_N(0) \in (\frac{C_d}{v_f}, \rho^j]$, $\bar{v}_{N-1} = \bar{v}_{N-1,1}$. If $v_{N-1} = 0$, i.e., $\bar{v}_{N-1,1} \le 0$, we have $q_N = 0$, thus in the region $\frac{C_d}{v_f} - \delta_2 \le \rho_N(t) \le \rho^j$, we have

$$\begin{aligned}\dot{\rho}_N &= -q_{N+1} = -\min\{v_f\rho_N, (1-\epsilon(\rho_N))C_d, \widetilde{w}(\tilde{\rho}^j - \rho_N)\} \\ &\le -\frac{\min\{C_d - v_f\delta_2, \widetilde{w}(\tilde{\rho}^j - \rho^j), (1-\epsilon_0)C_d\}}{\rho^j - C_d/v_f + \delta_1}\left(\rho_N - \frac{C_d}{v_f} + \delta_1\right).\end{aligned}$$

If $v_{N-1} > 0$, i.e., $\bar{v}_{N-1} > 0$, then $q_N \le v_{N-1}\rho_{N-1} \le \bar{v}_{N-1}\rho_{N-1}$,

$$\dot{\rho}_N \le \bar{v}_{N-1}\rho_{N-1} - q_{N+1} = -\lambda_{N-1}\left(\rho_N - \frac{C_d}{v_f} + \delta_1\right).$$

Therefore, $\forall \rho_N(0) \in (\frac{C_d}{v_f}, \rho^j]$, we have

$$\dot{\rho}_N \le -\alpha\left(\rho_N - \frac{C_d}{v_f} + \delta_1\right),$$

where $\alpha = \min\{\lambda_{N-1}, \frac{\min\{C_d - v_f\delta_2, \widetilde{w}(\tilde{\rho}^j - \rho^j), (1-\epsilon_0)C_d\}}{\rho^j - C_d/v_f + \delta_1}\} > 0$. Since $\alpha > 0$ and $\frac{C_d}{v_f} - \delta_1 < \frac{C_d}{v_f} - \delta_2 < \frac{C_d}{v_f} < \rho_N(0)$, and $\frac{C_d}{v_f} - \delta_2 > 0$, $\rho_N(t)$ will decrease exponentially fast to the value $\frac{C_d}{v_f} - \delta_2$ at some finite time t_0, at which $\bar{v}_{N-1}$ switches to $\bar{v}_{N-1,2}$, in which case $\rho(t)$ evolves as analyzed below.

Either the initial condition $\rho_N(0) \in \left[0, \frac{C_d}{v_f}\right]$ or $\bar{v}_{N-1}$ switches to $\bar{v}_{N-1,2}$ from $\bar{v}_{N-1,1}$ at $\rho_N = \frac{C_d}{v_f} - \delta_2$, there exists $t_0 \ge 0$, at which time instant $\rho_N(t_0) \le \frac{C_d}{v_f}$ and $\bar{v}_{N-1} = \bar{v}_{N-1,2}$. Since $\rho_N \le \frac{C_d}{v_f}$, $q_{N+1} \ge 0$, and $\rho_{N-1} \ge 0$, we have $\bar{v}_{N-1,2} \ge 0$ from equation (56), thus $v_{N-1} = \min\{v_f, \bar{v}_{N-1,2}\}$. Therefore

$$\dot{\rho}_N \le \bar{v}_{N-1,2}\rho_{N-1} - q_2 = -\lambda_{N-1}\left(\rho_N - \frac{C_d}{v_f}\right).$$

Without loss of generality, let $t_0 = 0$, then we have

$$\rho_N(t) \le \frac{C_d}{v_f} + \left(\rho_N(0) - \frac{C_d}{v_f}\right)e^{-\lambda_{N-1}t},$$

which implies that $\forall t \geq 0, \rho_N(t) \leq \frac{C_d}{v_f}$, $\bar{v}_{N-1} = \bar{v}_{N-1,2}$. Then we examine the dynamics of ρ_{N-1}. If $v_{N-2} = 0$, i.e., $\bar{v}_{N-2} \leq 0$, $q_{N-1} = 0$, we have

$$\dot{\rho}_{N-1} = -q_N = -\min\{v_{N-1}\rho_{N-1}, \frac{v_{N-1}w\rho^j}{v_{N-1}+w}, C, w(\rho^j - \rho_N)\}.$$

Since $\rho_N \leq \frac{C_d}{v_f} < \rho_c$, $w(\rho^j - \rho_N) > C$ and $v_{N-1} = \min\{v_f, \bar{v}_{N-1,2}\}$, we have

$$\dot{\rho}_{N-1} = -\min\{\bar{v}_{N-1,2}\rho_{N-1}, v_f\rho_{N-1}, \frac{\bar{v}_{N-1,2}w\rho^j}{\bar{v}_{N-1,2}+w}, \frac{v_f w\rho^j}{v_f+w}, C\}.$$

Since $0 \leq \rho_N \leq \frac{C_d}{v_f}, \lambda_{N-1} > v_f$ and $\rho_{N-1} \leq \rho^j$, we have

$$\begin{aligned}
\bar{v}_{N-1,2}\rho_{N-1} &= C_d + (v_f - \lambda_{N-1})\left(\rho_N - \frac{C_d}{v_f}\right) \geq C_d \geq \frac{C_d}{\rho^j - \frac{C_d}{v_f}}\left(\rho_{N-1} - \frac{C_d}{v_f}\right), \\
v_f\rho_{N-1} &\geq v_f\left(\rho_{N-1} - \frac{C_d}{v_f}\right), \\
\frac{\bar{v}_{N-1,2}w\rho^j}{\bar{v}_{N-1,2}+w} &= w\rho^j \frac{C_d + (v_f - \lambda_{N-1})x_N}{C_d + (v_f - \lambda_{N-1})x_N + w\rho_{N-1}} \geq \frac{w\rho^j v_f C_d}{v_f w\rho^j + \lambda_{N-1}C_d} \\
&\geq \frac{w\rho^j v_f C_d}{(v_f w\rho^j + \lambda_{N-1}C_d)\left(\rho^j - \frac{C_d}{v_f}\right)}\left(\rho_{N-1} - \frac{C_d}{v_f}\right), \\
\frac{v_f w\rho^j}{v_f+w} &= C \geq \frac{C}{\rho^j - \frac{C_d}{v_f}}\left(\rho_{N-1} - \frac{C_d}{v_f}\right).
\end{aligned}$$

Thus $\dot{\rho}_{N-1} \leq -\min\{\frac{C_d}{\rho^j - \frac{C_d}{v_f}}, v_f, \frac{w\rho^j v_f C_d}{(v_f w\rho^j + \lambda_{N-1}C_d)\left(\rho^j - \frac{C_d}{v_f}\right)}, \frac{C}{\rho^j - \frac{C_d}{v_f}}\}\left(\rho_{N-1} - \frac{C_d}{v_f}\right)$. If $v_{N-2} > 0$, i.e., $\bar{v}_{N-2} > 0$,

$$\dot{\rho}_{N-1} \leq \bar{v}_{N-2}\rho_{N-2} - q_N = -\lambda_{N-2}\left(\rho_{N-1} - \frac{C_d}{v_f}\right).$$

To conclude,

$$\dot{\rho}_{N-1} \leq -\alpha\left(\rho_{N-1} - \frac{C_d}{v_f}\right),$$

where

$$\alpha = \min\{\frac{C_d}{\rho^j - \frac{C_d}{v_f}}, v_f, \frac{w\rho^j v_f C_d}{\left(v_f w\rho^j + \lambda_{N-1} C_d\right)\left(\rho^j - \frac{C_d}{v_f}\right)}, \frac{C}{\rho^j - \frac{C_d}{v_f}}, \lambda_{N-2}\} > 0.$$

Therefore $\limsup_{t\to\infty} \rho_{N-1}(t) \le C_d/v_f$ and $\liminf_{t\to\infty} \bar{v}_{N-1,2} \ge v_f$ due to $\bar{v}_{N-1,2} = \frac{C_d + (v_f - \lambda_{N-1})x_N}{C_d/v_f + x_{N-1}} \ge \frac{C_d}{\rho_{N-1}}$, which implies $\lim_{t\to\infty} v_{N-1} = v_f$.

Similarly, we can show that $\limsup_{t\to\infty} \rho_i(t) \le C_d/v_f$ and $\lim_{t\to\infty} v_i = v_f$ for $i = 1,2,\dots,N-1$. Then the dynamics of $\rho(t)$ become

$$\begin{aligned}
\dot{\rho}_1 &= \min\{d, \frac{\bar{v}_0 w\rho^j}{\bar{v}_0 + w}, \frac{v_f w\rho^j}{v_f + w}\} - v_f\rho_1 = \min\{d - v_f\rho_1, C - v_f\rho_1, -\lambda_0(\rho_1 - C_d/v_f)\}, \\
\dot{\rho}_i &= v_f\rho_{i-1} - v_f\rho_i, i = 2,\dots,N.
\end{aligned}$$

Note that the first differential equation is the same as equation (183) in the single-section case. Therefore we can directly take the analysis result of equation (183), which shows that ρ_1 converges to $\rho_1 = \frac{\min\{d, C_d\}}{v_f}$ exponentially fast. Consequently, ρ_i converges exponentially fast to $\rho_i = \frac{\min\{d, C_d\}}{v_f}$, for $i = 1,2,\dots,N$. Recall that $\rho_N(t) \le \frac{C_d}{v_f}$, $\forall t \ge t_0$, thus q_{N+1} converges to C_d exponentially fast. Consequently, q_i converge to C_d exponentially fast for $i = 1,2,\dots,N$.

b) This part can be shown directly with part e) of Theorem 1.4.

Appendix G: The Proof of Lemma 1.2

Proof. Let us use the transformation $x_1(t) = \rho(t) - \rho^\star$ and $x_2(t) = \Phi(t) - \frac{\mu}{\lambda_2}$ to shift the equilibrium point of system (69) to the origin (0,0). Let $x = [x_1, x_2]^T$ be the shifted state vector of system (69). Based on the value of q_1, we have the following three cases:

Case 1. When $0 \le \bar{q}_{1v} \le d$, i.e., $x \in S_1 = \{x | v_f\rho^\star - \mu - d \le (\lambda_1 - v_f)x_1 + \lambda_2 x_2 \le v_f\rho^\star - \mu\}$, we have $q_1 = \bar{q}_{1v}$, and the dynamics of system (69) can be rewritten as:

$$\begin{aligned} \dot{x}_1 &= -\lambda_1 x_1 - \lambda_2 x_2 \\ \dot{x}_2 &= x_1 \end{aligned} \tag{185}$$

Case 2. When $\bar{q}_{1v} < 0$, i.e., $x \in S_2 = \{x | (\lambda_1 - v_f)x_1 + \lambda_2 x_2 > v_f\rho^\star - \mu\}$, we have $q_1 = 0$, and the dynamics of system (69) can be rewritten as:

$$\begin{aligned} \dot{x}_1 &= -v_f x_1 - v_f\rho^\star + \mu \\ \dot{x}_2 &= x_1 \end{aligned} \tag{186}$$

$$\begin{aligned} \dot{x}_1 &= -v_f x_1 - v_f\rho^\star + \mu \\ \dot{x}_2 &= x_1 \end{aligned}$$

Case 3. When $\bar{q}_{1v} > d$, i.e., $x \in S_3 = \{x | (\lambda_1 - v_f)x_1 + \lambda_2 x_2 < v_f\rho^\star - \mu - d\}$, we have $q_1 = d$, and the dynamics of system (69) can be rewritten as:

$$\begin{aligned} \dot{x}_1 &= -v_f x_1 - v_f\rho^\star + \mu + d \\ \dot{x}_2 &= x_1 \end{aligned} \tag{187}$$

Therefore, the state space is divided as shown in Figure 146

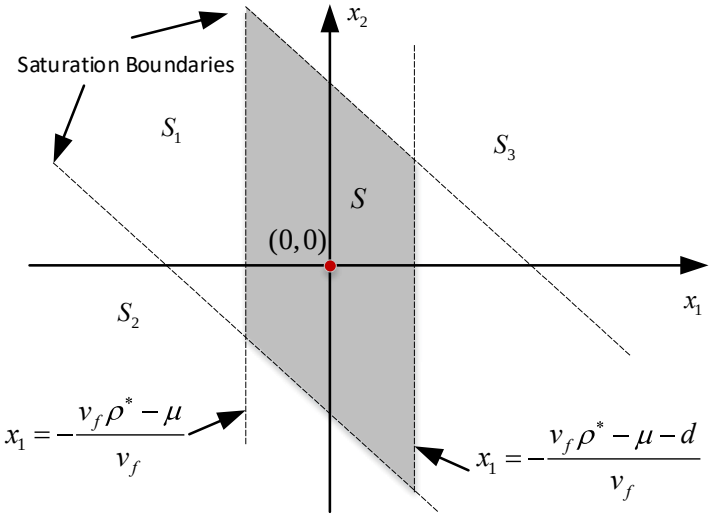


Figure 146. State Space

It is clear that system (185) has a unique equilibrium point $x^e = [0,0]^T \in S_1$, whereas system (186) and (187) have no equilibrium points.

Consider the following Lyapunov function

$$V(x) = x^T P x, \tag{188}$$

where

$$P = \begin{bmatrix} 2 & \lambda_1 \\ \lambda_1 & \lambda_1^2 + 2\lambda_2 \end{bmatrix}$$

$V(x)$ is positive definite since the matrix P is symmetric and positive definite.

For all $x \in S_1$, we have

$$\dot{x} = Ax$$

where

$$A = \begin{bmatrix} -\lambda_1 & -\lambda_2 \\ 1 & 0 \end{bmatrix}$$

The derivative of the above Lyapunov function (188) is

$$\dot{V}(x) = x^T(A^T P + PA)x = -x^T Q x,$$

where

$$Q = \begin{bmatrix} 2\lambda_1 & 0 \\ 0 & 2\lambda_1\lambda_2 \end{bmatrix}.$$

Thus, $\dot{V}(x) < 0$ for all $x \in S_1 \setminus \{0\}$.

Now, we are going to show that if $\exists t_0 \geq 0$, such that $x(t_0) \in S_3$, then $\exists t_1 > t_0$, such that $x(t_1) \in S_1$. Let $\alpha = v_f \rho^\star - \mu - d$. If $\exists t_0 \geq 0$, such that $x(t_0) \in S_3$, then according to (187) we have that

$$\begin{aligned} \frac{dx_2}{dx_1} &= \frac{x_1}{-v_f x_1 - \alpha} = -\frac{1}{v_f} + \frac{\alpha}{v_f^2 x_1 + v_f \alpha} \\ dx_2 &= -\frac{1}{v_f} dx_1 + \frac{\alpha}{v_f^2 x_1 + v_f \alpha} dx_1 \end{aligned}$$

taking the integral of both sides, we get

$$x_2(t) - x_2(t_0) = -\frac{1}{v_f}\left(x_1(t) - x_1(t_0)\right) + \frac{\alpha}{v_f^2}\left[\ln\left(v_f^2 x_1(t) + v_f\alpha\right) - \ln\left(v_f^2 x_1(t_0) + v_f\alpha\right)\right] \quad (189)$$

$$x_2(t) = -\frac{1}{v_f}x_1(t) + \frac{\alpha}{v_f^2}\ln(v_f^2 x_1(t) + v_f\alpha) + x_2(t_0) + \frac{1}{v_f}x_1(t_0) - \frac{\alpha}{v_f^2}\ln(v_f^2 x_1(t_0) + v_f\alpha)$$

From (187), we know that $x_1(t)$ approaches $-\frac{\alpha}{v_f}$ when $x \in S_3$. According to the above equation (189), x_2 approaches infinity as $x_1(t)$ approaches $-\frac{\alpha}{v_f}$. Thus, at some finite time instant t_1, $x_2(t_1)$ is large enough and $\left(\lambda_1 - v_f\right)x_1 + \lambda_2 x_2 = v_f\rho^\star - \mu - d$, i.e., $x \in S_1$.

Then, we show that if $\exists t_0 \geq 0$, at which $x(t_0)$ lies on the boundary between S_1 and S_3, and $x(t)$ moves towards S_3, then $\exists t_1 > t_0$, at which $x(t_1)$ lies on the boundary between S_1 and S_3, and $x(t)$ moves towards S_1. Furthermore, $V\left(x(t_0)\right) > V\left(x(t_1)\right)$. The normal vector of the boundary line which points to S_3 is $V_n = \left[v_f - \lambda_1, -\lambda_2\right]^T$. If $\exists t_0 \geq 0$, at which $x(t_0)$ lies on the boundary between S_1 and S_3, and $x(t)$ moves towards S_3, we have the following:

$$x(t_0) = [x_1(t_0), x_2(t_0)]^T$$

and

$$\dot{x}(t_0) = \left[-v_f x_1(t_0) - \alpha, x_1(t_0)\right]^T$$

Since $\dot{x}(t_0)$ points to S_3, $V_n^T\dot{x}(t_0) > 0$, that is

$$\left[v_f\left(\lambda_1 - v_f\right) - \lambda_2\right]x_1(t_0) - \left(v_f - \lambda_1\right)\alpha > 0$$

$$x_1(t_0) > \frac{\left(v_f - \lambda_1\right)\alpha}{v_f\left(\lambda_1 - v_f\right) - \lambda_2} > -\frac{\alpha}{v_f}$$

due to $\lambda_1 > v_f + \frac{\lambda_2}{v_f}$. According to (187), if $x(t_0) \in S_3$ and $x_1(t_0) > -\frac{\alpha}{v_f}$, then $x_1(t) > -\frac{\alpha}{v_f}$ and $\dot{x}_1(t) < 0, \forall t > t_0$, as long as $x(t)$ stays in S_3. Together with (189), we know that $\exists t_1 > t_0$, at which the trajectory of $x(t)$ crosses the boundary line and moves towards S_1, $x_1(t_1) > -\frac{\alpha}{v_f}$ and $x_1(t_1) < x_1(t_0)$. For all points on the boundary line between S_1 and S_3, the Lyapunov function is evaluated as

$$V(x) = 2x_1^2 + 2\lambda_1 x_1 x_2 + (\lambda_1^2 + 2\lambda_2)x_2^2,$$

whose partial derivative with respect to x_1 along the boundary line is

$$\frac{\partial V}{\partial x_1} = 4x_1 + 2\lambda_1 x_2 + 2\lambda_1 x_1\frac{\partial x_2}{\partial x_1} + 2(\lambda_1^2 + 2\lambda_2)\frac{\partial x_2}{\partial x_1} \quad (190)$$

Since $x(t_0) = \left[x_1(t_0), -\frac{\lambda_1 - v_f}{\lambda_2}x_1(t_0) + \frac{\alpha}{\lambda_2}\right]^T$ on the boundary line, we have that

$$\frac{\partial x_2}{\partial x_1} = -\frac{\lambda_1 - v_f}{\lambda_2}.$$

Thus, (190) can be rewritten as

$$\frac{\partial V}{\partial x_1} = ax_1 + b$$

where

$$a = 4 - 4\lambda_1 \frac{\lambda_1 - v_f}{\lambda_2} + 2(\lambda_1^2 + 2\lambda_2)\left(\frac{\lambda_1 - v_f}{\lambda_2}\right)^2$$

and

$$b = 2[\lambda_1 - (\lambda_1^2 + 2\lambda_2)]\frac{\alpha}{\lambda_2}.$$

Note that

$$a = 2\left[1, \frac{\lambda_1 - v_f}{\lambda_2}\right]^T P\left[1, \frac{\lambda_1 - v_f}{\lambda_2}\right] > 0$$

due to P been positive definite, and

$$a\left(-\frac{\alpha}{\lambda_2}\right) + b = -\left[2\lambda_1^2(\lambda_1 - v_f) + 2\lambda_1\lambda_2\right]\alpha > 0$$

due to $\alpha < 0$ and $\lambda_1 > v_f + \frac{\lambda_2}{v_f}$. Thus, $\forall x_1 > -\frac{\alpha}{\lambda_2}, \frac{\partial V}{\partial x_1} = ax_1 + b > 0$. Therefore, $-\frac{\alpha}{\lambda_2} < x_1(t_1) < x_1(t_0)$, indicating that $V\big(x(t_0)\big) > V\big(x(t_1)\big)$.

Similarly, we can show that if $\exists t_0 \geq 0$, such that $x(t_0) \in S_2$, then $\exists t_1 > t_0$, such that $x(t_1) \in S_1$. Moreover, if $\exists t_0 \geq 0$, at which $x(t_0)$ lies on the boundary line between S_1 and S_2 and $x(t)$ moves towards S_2, then $\exists t_1 > t_0$, at which $x(t_0)$ lies on the boundary between S_1 and S_2, and $x(t)$ moves towards S_1. Furthermore, $V\big(x(t_0)\big) > V\big(x(t_1)\big)$.

Summarizing the behavior of the Lyapunov function $V(x)$, we conclude that $\forall x(t_0) \in \Re^2$, $x(t)$ converges to $x^e = [0,0]^T$ asymptotically, which implies that $\forall\big(\rho(t_0), \Phi(t_0)\big) \in \Re^2$, $[\rho(t), \Phi(t)]^T$ asymptotically converges to $\left[\rho^\star, \frac{\mu}{\lambda_2}\right]^T$.

It follows from (69) that

$$-v_f x_1 - \big(v_f\rho^\star - \mu\big) \leq \dot{x}_1 \leq -v_f x_1 - \big(v_f\rho^\star - \mu - d\big),$$

thus for all $-\frac{v_f\rho^\star-\mu}{v_f} < x_1(t_0) < -\frac{v_f\rho^\star-\mu-d}{v_f}$, $-\frac{v_f\rho^\star-\mu}{v_f} < x_1(t) < -\frac{v_f\rho^\star-\mu-d}{v_f}$, $\forall t \geq t_0$. We have also shown that if $x(t) \in S_1 \cap \{x| -\frac{v_f\rho^\star-\mu}{v_f} < x_1 < -\frac{v_f\rho^\star-\mu-d}{v_f}\}$, $x(t)$ will not leave S_1. Therefore, $\forall x(t_0) \in S = \{x\big|v_f\rho^\star - \mu - d \leq (\lambda_1 - v_f)x_1 + \lambda_2 x_2 \leq v_f\rho^\star - \mu\} \cap \{x\big| -\frac{v_f\rho^\star-\mu}{v_f} < x_1 < -\frac{v_f\rho^\star-\mu-d}{v_f}\}$, $x_1(t) \in S, \forall t \geq t_0$, implying that $\forall\big(\rho(t_0), \Phi(t_0)\big) \in S, (\rho(t) - \rho^\star) \in S, \forall t \geq t_0$.

Appendix H: TDSP Extensions

Multiple Clients with Service Time

We consider the problem of scheduling the rest stops for a long-haul truck trip with a known route and multiple clients while taking into account the USA HOS regulations and estimated parking availability windows for all rest areas along the route. It is assumed that the rest areas are located on the route and require no detours to be accessed. The parking availability time-windows are assumed known. The route has $n+1$nodes, 2 of which are the origin, node 0, and destination of the truck, node n . The other nodes are either clients, rest areas, or dummy locations used to model special cases. Let $N = \{0,1, \dots, n\}$ be the set of all locations, $N_r \subset N$ the set of rest areas, $N_c \subset N$ the set of client locations, and $N_{cr} \subset N$ is the set of dummy locations used to model resting at a client location immediately after completing the required service. Every node in N_{cr} is positioned right after a node in N_c, and the travel time between them is zero. The dummy locations used for resting at a client location before starting the service are treated as regular rest areas. The travel time, d_i , between locations i and $i+1$ is assumed constant, $x_{i,a}$ and $x_{i,d}$ represent the arrival and departure time of the truck at location i , respectively. Each location i has T_i time-windows $\left[t_{i,\tau}^{min}, t_{i,\tau}^{max}\right]$, where $\tau \in \{1,2, \dots, T_i\}$ indicates the time-window's index. The time-windows restrict the arrival time at that node and are only in effect when the truck has to stop at that specific node, driving by it is not constrained by the time windows. For each location and time-window, a binary variable $y_{i,\tau}$ represents if that specific time window is being used (yes:1, no:0). Driving by without stopping is represented by the variable $y_{i,0}$ (drive by:1, stop:0). In the case of rest areas, the time-windows represent the area's parking availability, whereas for client locations the time-windows represent delivery time constraints imposed by the client. The driver must stop at all clients, so $y_{i,0} = 0$ for all client locations. Each client has a service time s_i which is assumed constant.

The schedule must comply with the USA HOS regulations. Some HOS rules regulate driving time or elapsed time, so they are not affected by the inclusion of service time in the model. However, regulations that consider on-duty time must account for the service time, as the driver is working during that time. R is defined as the set of different types of rest period described in the regulation. For each $r \in R$, t_r defines the minimum duration of that type of rest period. C is the set of constraints imposed by the regulation. $C_1 \subseteq C$ is the set of constraints controlling the maximum elapsed time between off-duty periods. $C_2 \subseteq C$ is the set of constraints controlling the maximum accumulated driving time between off-duty periods. $C_3 \subseteq C$ is the set of constraints controlling the maximum accumulated on-duty time during a rolling time-window; the width of the time-window for a constraint $c \in C_3$ is represented by δ_c. In the USA regulation δ_c is 7 days, so these rolling time-window constraints will be referred to as weekly constraints. For each constraint $c \in C$, t_c is the time limit imposed by the regulation and $R_c \subseteq R$ is the set of rest types that can reset this counter. The binary variable $z_{i,r}$ indicates whether a rest of type r is taken at location i (yes:1, no:0). The driver cannot take more than one type of rest at the same location. If no type of rest is schedule for a rest area, the driver cannot stop there. The departure time from the origin must be within the interval $\left[t_0, t_{dep}\right]$. It is

assumed that the driver has been off-duty for long enough before the departure time, so that all constraints' counters are reset before departure.

It is important to note that these rules state that the driver cannot drive after the time limit is reached, not that he/she cannot work. Therefore, the limit can be exceeded during service time as long as the driver can rest at the client location immediately after, without having to drive. Because of this the dummy nodes used to model resting at client locations after working need to be treated differently. The extended model without the weekly constraints is the following:

$$\min x_{n,a} - x_{0,d} \tag{191}$$

s.t.:

$$x_{i,d} + d_i = x_{i+1,a}, \forall 0 \leq i \leq n-1 \tag{192}$$

$$x_{i,a} + \sum_{r \in R} t_r z_{i,r} \leq x_{i,d}, i \in N \backslash N_c \tag{193}$$

$$x_{i,a} + s_i = x_{i,d}, i \in N_c \tag{194}$$

$$x_{i,d} \leq x_{i,a} + (1 - y_{i,0}) t_{hor}, \forall 1 \leq i \leq n \tag{195}$$

$$y_{i,0} + \sum_{\tau=1}^{T_i} y_{i,\tau} = 1, \forall 1 \leq i \leq n \tag{196}$$

$$y_{i,0} = 0, i \in N_c \tag{197}$$

$$\sum_{\tau=1}^{T_i} y_{i,\tau} = \sum_{r \in R} z_{i,r}, \forall i / \in N_c \cup \{0\} \tag{198}$$

$$\sum_{r \in R} z_{i,r} = 0, \forall i \in N_c \cup \{0\} \tag{199}$$

$$\sum_{\tau=1}^{T_i} y_{i,\tau} \, t_{i,\tau}^{min} \leq x_{i,a}, \forall 1 \leq i \leq n \tag{200}$$

$$x_{i,a} \leq t_{hor} - \sum_{\tau=1}^{T_i} \left[y_{i,\tau} \left(t_{hor} - t_{i,\tau}^{max} \right) \right], \forall 1 \leq i \leq n \tag{201}$$

$$x_{k,a} - x_{i,d} \leq t_c + t_{hor} \sum_{j=i+1}^{k-1} \sum_{r \in R_c} z_{j,r}, \forall 0 \leq i < k \leq n, k / \in N_{cr}, c \in C_1 \tag{202}$$

$$x_{k,a} - x_{i,d} \leq t_c + t_{hor} \sum_{j=i+1}^{k} \sum_{r \in R_c} z_{j,r}, \forall 0 \leq i < k \leq n, k \in N_{cr}, c \in C_1 \tag{203}$$

$$\sum_{j=i}^{k-1} d_j \leq t_c + t_{hor} \sum_{j=i+1}^{k-1} \sum_{r \in R_c} z_{j,r}, \forall 0 \leq i < k \leq n, c \in C_2 \quad (204)$$

$$x_i \in [0, t_{hor}]^2, y_i \in \{0,1\}^{T_i+1}, z_i \in \{0,1\}^{|R|}, \forall 0 \leq i \leq n+1 \quad (205)$$

$$x_{0,d} \in [0, t_{dep}] \quad (206)$$

The objective function (191) is set to minimize the total trip duration. Constraint (192) guarantees that the arrival time equals the departure time of the previous location plus the driving time. Constraint (193) states that the vehicle must stay at a location (clients excluded) at least for the minimum rest time decided for that location. Constraint (194) states that the vehicle must depart a client location as soon as the service time is over. Constraint (195) controls what happens when the driver does not stop at a certain location. If the vehicle does not stop at location i, the arrival time equals the departure time. This constraint works with constraints (193), (196) and (198) to assure this. Equality will hold when $y_{i,0} = 1$. If $y_{i,0} = 0$, then constraint (195) is always true as t_{hor} is large. Constraint (196) states that at any location, either exactly one time-window is used or the vehicle does not stop. Constraint (197) states that the driver must stop at all client locations. Constraint (198) states that the driver only stops if an off-duty period is scheduled. Constraints (200) and (201) check the time-windows. Arrival must happen after the beginning and before the end of the chosen time window. Constraints (202) and (203) check that the time elapsed since the last rest in $R_c, c \in C_1$ is less than t_c. Constraint (204) checks if the accumulated driving time between rest periods in $R_c, c \in C_2$ is less than t_c. Constraint (205) sets the variables' domains, and (206) guarantees that the departure time from the origin is within the required period and that the vehicle will stop at the destination.

Weekly Constraints

The constraints described here are from the set C_3, for simplicity we will consider that $|C_3| = 1$ and will omit the index c, which indicates to which rule the variable refers to, from the variables. In the USA regulation only one rolling time-window constraint is used at a time, so this assumption is reasonable. Nevertheless, if more constraints are needed for a specific country, they can be added separately. The constraints C_3 can be simplified to limit the on-duty time between two weekly rests, instead of limiting the on-duty time during a rolling time-windows. This formulation will give suboptimal results in certain scenarios, but also guarantees regulation compliant schedules. These two formulations will be presented separately to avoid confusion.

Simplified Constraint

As mentioned previously, this simplified formulation limits the on-duty time between weekly rests, accounting for the possibility of resting immediately after finishing work at a client.

$$\sum_{j=i}^{k-1} d_j + \sum_{j=i+1}^{k-1} s_j \leq t_c + t_{hor} \sum_{j=i+1}^{k-1} \sum_{r \in R_c} z_{j,r}\,, \forall 0 \leq i \leq k \leq n,\ k \in N \setminus N_{cr},\ c \in C_3 \quad (207)$$

$$\sum_{j=i}^{k-1} d_j + \sum_{j=i+1}^{k-1} s_j \leq t_c + t_{hor} \sum_{j=i+1}^{k} \sum_{r \in R_c} z_{j,r}\,, \forall 0 \leq i \leq k \leq n, k \in N_{cr}, c \in C_3 \quad (208)$$

Constraint (207) checks if the accumulated driving and service time between departure from location i and arrival at location k is below the limit; the constraint is automatically satisfied if an appropriate rest is taken at any location in between, not including i and k. Constraint (208) treats the particular case of resting after working at a client. Similarly, it checks if the accumulated driving and service time between departure from location i and arrival at location k is below the limit, but now, an appropriate rest at k can also satisfy the constraint. As $k \in N_{cr}$, resting at k is equivalent to resting at the client location right after the work is completed. The driver does not need to drive between $k-1$ and k, so the regulation is not violated even if the on-duty time limit is exceeded during the service at k.

Rolling Time-Window Constraint

This constraint makes it so the driver cannot drive after having accumulated more than t_c on-duty hours over the last δ_c hours, with $c \in C_3$. Let $\lambda_i(t)$ represent the accumulated driving time generated by the service at location i and the displacement between locations i and $i+1$ measured at time t. $\lambda_i(t)$ is described by:

$$\lambda_i(t) = R(t - x_{i,d}) - R(t - x_{i+1,a}) - R(t - x_{i,d} - \delta_c) + R(t - x_{i+1,a} - \delta_c), \quad \forall 0 \leq i \leq n-1,\ i \in N \setminus N_c,\ c \in C_3 \quad (209)$$

$$\lambda_i(t) = R(t - x_{i,a}) - R(t - x_{i+1,a}) - R(t - x_{i,a} - \delta_c) + R(t - x_{i+1,a} - \delta_c), \quad \forall 0 \leq i \leq n-1, i \in N_c, c \in C_3 \quad (210)$$

where $R(t)$ is the unit ramp function. Equation (209) represents the accumulated on-duty time generated when i is not a client location. As any time spent at i is used for resting, the on-duty time only starts being counted after departure ($t = x_{i,d}$), the count stops upon arrival at the next location ($t = x_{i+1,a}$), and stays constant until $t = x_{i,d} + \delta_c$, when that on-duty time starts leaving the rolling time-window until reaching zero again at $t = x_{i+1,a} + \delta_c$. Equation (210) is used when i is a client location, so the service time at i also needs to be counted. Therefore, the on-duty time starts being counted when the driver arrives at i, $t = x_{i,a}$. The functions $\lambda_i(t)$ are included in the model as piecewise linear functions. The domains of the functions are divided in sections according to when the slope of the functions change, and auxiliary variables are used to write t according to where it is located relative to the sections' boundaries. For this problem, it suffices to evaluate $\lambda_i(t)$ at the arrival times $x_{j,a}, j > i$, so we define the variables $\lambda_{i,j} = \lambda_i(x_{j,a})$. The sets of variables $\{\alpha_{i,j,p}\}$, $\{\beta_{i,j,q}\}$ are used to define $\{\lambda_{i,j}\}$ as follows:

$$\alpha_{i,j,p} \in \{0,1\}, \beta_{i,j,q} \in [0,1], \qquad \forall 0 \le i < j \le n,\ 0 \le p \le 4,\ 1 \le q \le 5 \tag{211}$$

$$1 \ge \alpha_{i,j,0} \ge \beta_{i,j,1} \ge \alpha_{i,j,1} \ge \cdots \ge \alpha_{i,j,4} \ge \beta_{i,j,5}, \qquad \forall 0 \le i < j \le n \tag{212}$$

$$\alpha_{i,j,p} < \beta_{i,j,p+1} + 1, \qquad \forall 0 \le i < j \le n \tag{213}$$

$$x_{j,a} = x_{i,a}\beta_{i,j,1} + d_i\beta_{i,j,2} + (\delta_c - d_i)\beta_{i,j,3} + d_i\beta_{i,j,4} + t_{hor}\beta_{i,j,5}, \quad \forall 0 \le i < j \le n,\ i \in N \setminus N_c,\ c \in C_3 \tag{214}$$

$$x_{j,a} = x_{i,a}\beta_{i,j,1} + (d_i + s_i)\beta_{i,j,2} + (\delta_c - d_i - s_i)\beta_{i,j,3} + (d_i + s_i)\beta_{i,j,4} + t_{hor}\beta_{i,j,5}, \quad \forall 0 \le i < j \le n,\ i \in N_c,\ c \in C_3 \tag{215}$$

$$\lambda_{i,j} = (d_i + s_i)\left(\beta_{i,j,2} - \beta_{i,j,4}\right), \forall 0 \le i < j \le n \tag{216}$$

where the α's and β's are auxiliary variables used to model the piecewise definition of $\lambda_{i,j}$. The α's determine in which section of the function domain t is, and the β's define its exact position within the section, the indexes p and q represent the sections. Constraints (211), (212) and (213) imply that, for a section q , whenever $0 < \beta_{*,q} < 1$, then $\beta_{*,q^-} = 1, \alpha_{*,q^-} = 1,\ \forall q^- < q$, and $\beta_{*,q^+} = 0, \alpha_{*,p} = 0,\ \forall q^+ > q, p \ge q$. Constraints (214) and (215) write the time instant to be evaluated, $x_{j,a}$, as a function of the α's and β's. Constraint (216) uses the β's to calculate $\lambda_i\left(x_{j,a}\right)$. Due to (214) this problem would be a quadratically constrained problem. However, as (214) only considers $j > i$, the variables $\alpha_{i,j,p}$ for $p < 2$ and $\beta_{i,j,q}$ for $q < 3$ will be always 1 and can be defined as constants. The accumulated driving time generated by all trips or service starting at locations $\{0, \dots, i\}$ measured at time $x_{j,a}$ is represented by $\psi_{i,j}$.

$$\psi_{i,j} = \begin{cases} \psi i - 1, j + \lambda_{i,j}, if \sum_{r \in R_c} z_{i,r} = 0 \\ \lambda_{i,j}, if \sum_{r \in R_c} z_{i,r} = 1 \end{cases} \qquad \forall 1 \le i < j \le n,\ c \in C_3 \tag{217}$$

$$\psi_{0,j} = \lambda_{0,j}, \quad \forall 1 \le i < j \le n \tag{218}$$

$$\psi_{j-1,j} \le t_c, \quad \forall 1 \le j \le n, j \notin N_{cr}, c \in C_3 \tag{219}$$

where constraint (217) defines $\psi_{i,j}$ and sets to zero all contributions from nodes before location i when an appropriate rest is taken at location i. This model assumes that all constraints' counters are reset before departure, so (218) sets the initial accumulated driving time to zero. Constraint (219) limits the accumulated on-duty time to the regulation limit t_c, for all nodes except if they are used for after-work rest at a client location. Similarly to what was explained for constraint (208), the limit can be exceed during a service time, as long as the driver rests immediately after, without driving.

Electric Vehicles

Consider the previous problem, but now the truck used is electric, more specifically, battery electric. The truck has a battery with total capacity B in kWh. When driving along a link $(i, i+1)$ the battery is consumed at a rate of $\gamma_i^a \ kWh/h$, this rate is assumed constant for each link. The route has a set $N_s \subset N \setminus N_c$ of charging stations that allow the driver to both recharge the vehicle's battery and rest. It is assumed that all charging stations can be used for resting, but not all rest areas are charging stations. During a stay at a charging station the battery will be recharged at a rate of $\gamma_i^v \ kWh/h$, whereas during a stay at a regular rest area the battery charge is consumed at a rate $\gamma_i^v \ kWh/h$ due to the trucks auxiliary systems, like AC. As rest areas/charging stations have varying levels of infrastructure available, the rates of consumption/recharge can vary between locations. It is assumed that the battery cannot be charged during service at a client location. The battery charge cannot exceed the capacity B and it cannot be negative either. The time that a vehicle stays at a charging station is not limited, but it is assumed that the battery will automatically stop charging when full. The effective charging time at location $i \in N_s$ is represented by η_i . At locations without charging stations the vehicle will be consuming energy during the whole stay, so η_i will be equal to the whole duration of stay. For simplicity, the model assumes that all parking spaces of a charging station are electrified, so the parking availability time-windows are also the charging station availability time-windows. If necessary, electrified and regular parking spaces can be separated by including a dummy location specifically for the charging station with a different time-window adjacent to the regular rest area. This can also be used to separate spaces with different charging rates at the same location. Let $(b_{i,a}, b_{i,d})$ represent the battery charge at the time of arrival and departure at location i. The battery charge constraints can be described as follows:

$$\eta_i \leq \quad x_{i,d} - x_{i,a}, \qquad \forall i \in N_s \tag{220}$$

$$\eta_i = \quad x_{i,d} - x_{i,a}, \qquad \forall i \in N \setminus N_s \tag{221}$$

$$b_{i,d} = \quad b_{i,a} + \eta_i \gamma_i^v, \qquad \forall 1 \leq i \leq n \tag{222}$$

$$b_{i,a} = \quad b_{i-1,d} + d_i \gamma_i^a, \qquad \forall 1 \leq i \leq n \tag{223}$$

$$\eta_i \geq \quad 0, \ (b_{i,a}, b_{i,d}) \in [0, B]^2, \qquad \forall 0 \leq i \leq n \tag{224}$$

Constraint (220) defines the effective charging time at charging stations, whereas constraint (221) fixes the effective consumption time to the whole stay everywhere else. Constraint (222) updates the battery level after a stay at any location. Constraint (223) updates the battery level after a displacement. Note that the γ's will be negative when consuming energy and positive when charging the battery. Constraint (224) defines the domains of the b's and η's.

This model allows the optimization of the schedules of electric trucks with heterogeneous charging stations and rest areas. This extension can be directly added to the model presented previously, so the HOS and parking availability constraints are still considered, as well as the possibility of servicing multiple clients. Although this model allows for different consumption

rates to be used at each trip section, the rates are fixed. Energy consumption models found in the literature represent the energy consumption as a non-linear function of the vehicle speed, but this kind of consumption cannot be directly included in the current model. This model generates energy-feasible schedules, but does not optimize the energy consumption.

Appendix I: Trip Duration Lower-Bound

The schedule is affected by the truck stops' locations, by their availability windows, and by the HOS regulation. In order to calculate lower bounds for this problem, we can consider an ideal scenario where trucks have no restriction on parking location and time, only considering the HOS regulation. As the bounds are dependent only on the regulation being used, they can be calculated beforehand and used to evaluate the partial solutions obtained during the optimization. Here we consider the same structure of the USA regulation and assume that all resources are set to zero. An extension for the non-zero resources is presented later. Also, as the regulation treats service time and driving time differently, this formulation calculates the minimum trip duration for a trip with zero service time. When this is not the case, the service time can be added to the calculated lower-bound afterward. The bound will be looser in this case, but it still is a valid lower-bound for the trip duration. We calculated the lower bounds by optimizing the total trip duration for trip lengths (in terms of driving time) that do not use *daily rests*, then used this as a building block to optimize trips that do not need *weekly rests*, then for trips of any length. The parameters used are defined as follows:

t_b: minimum *break* duration;

t_r: minimum *daily rest* duration;

t_w: minimum *weekly rest* duration;

t_{eb} :limit for elapsed time between *breaks*;

t_{er} :limit for elapsed time between *daily rests*;

t_{ar}: limit for accumulated driving time between *daily rests*;

t_{aw}: limit for accumulated on-duty time between *weekly rests*;

ϵ: arbitrarily small positive constant.

First we calculate the minimum trip duration for a trip with less than a day's worth of driving time, $\mathcal{L}_d(\cdot)$.

$$\mathcal{L}_d(x) = x + t_b \left\lfloor \frac{x - \epsilon}{t_{eb}} \right\rfloor, \quad 0 \le x \le t_{ar}$$

where x represents the trip length in hours, i.e., the total driving time of the trip. As x was limited to less than the daily driving limit t_{ar}, we just need to calculate the number of *breaks* that will be necessary during the day. The ϵ is used to avoid including a *break* when the driving time is exactly on the limit.

Then $\mathcal{L}_d(\cdot)$ is used to describe the trip duration for trips with less than a week's worth of driving time, $\mathcal{L}_w(\cdot)$, as follows:

$$u(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases} \tag{225}$$

$$g_w(\overline{y}) = \mathcal{L}_d(y_0) + \sum_{i=1}^{n} [t_r u(y_i) + \mathcal{L}_d(y_i)], \quad \overline{y} \in \mathbb{R}^{n+1}, 0 \leq \|\overline{y}\|_\infty \leq t_{ar} \tag{226}$$

$$A(x) = \left\{\overline{y} \in \mathbb{R}^{n+1} | 0 \leq \|\overline{y}\|_\infty \leq t_{ar}, \|\overline{y}\|_1 = x, n = \left\lfloor \frac{x-\epsilon}{t_{eb}} \right\rfloor\right\} \tag{227}$$

$$\mathcal{L}_w(x) = \begin{cases} \min_{\overline{y} \in A(x)} g_w(\overline{y}), & t_{ar} < x \leq t_{aw} \\ \mathcal{L}_d(x), & 0 \leq x \leq t_{ar} \end{cases} \tag{228}$$

The 'shorter than a week' trips described here can be divided in multiple 'less than a day' trips separated by *daily rests*. Function $g_w(\cdot)$ optimizes these smaller sections using $\mathcal{L}_d(\cdot)$ and adds a *daily rests* for each non-zero section, calculating the minimum trip duration given a vector $\overline{y} = (y_0, y_1, \dots, y_n)$ composed of the $n+1$ section lengths y_i. Equation (228) optimizes the trip duration over all valid combinations of section lengths, with up to n *daily rests*. Equation (227) defines the valid section length vectors, choosing n such that the optimization considers enough *daily rest* to account for the case of taking a *daily rest* every time a *break* is needed.

The same approach is used to calculate the trip duration for longer trips, $\mathcal{L}_l(\cdot)$, as follows:

$$g_l(\overline{y}) = \mathcal{L}_w(y_0) + \sum_{1}^{n} [t_w u(y_i) + \mathcal{L}_w(y_i)], \quad \overline{y} \in \mathbb{R}^{n+1}, 0 \leq \|\overline{y}\|_\infty \leq t_{aw} \tag{229}$$

$$B(x) = \left\{\overline{y} \in \mathbb{R}^{n+1} | 0 \leq \|\overline{y}\|_\infty \leq t_{aw}, \|\overline{y}\|_1 = x, n = \left\lfloor \frac{x-\epsilon}{L} \right\rfloor\right\} \tag{230}$$

$$\mathcal{L}_l(x) = \left\{\min_{\overline{y} \in B(x)} g_l(\overline{y}), \quad t_{aw} < x\mathcal{L}_w(x), \quad 0 \leq x \leq t_{aw}\right. \tag{231}$$

where $t_{eb} \leq L \leq t_{aw}$ should be chosen in a way that creates a vector long enough to test the different possibilities of moving driving hours between weeks to avoid *breaks* and *daily rests* when possible. This value affects the maximum number of *weekly rests* that the problem will consider in the optimization. Taking $L = t_{eb}$ certainly works as it is the maximum driving time allowed without any kind of rest. In the case of USA regulations, the optimization would consider even the case of taking a *weekly rest* every 8h of driving. For long trips this resolution might be excessive and will unnecessarily slow down the calculations due to the large number of possible vectors. An optimal schedule will have a high driving time to trip duration ratio, so we suggest taking L as the length of the trip when the intermediate weekly cycles (the ones that are followed by a *weekly rest*) would be the most efficient, i.e., $L = \underset{x \leq t_{aw}}{\operatorname{argmax}} \frac{x}{\mathcal{L}_w(x) + t_w}$. In the case of the regulations considered in this project, L is equal to 55 hours.

This formulation considers that all resources are set to zero, so it provides a very loose bound in most cases. At the cost of computational complexity and/or memory space, this can be

improved by including the current resources in the calculation. Consider the variables γ_b, γ_r and γ_w defined as follows:

$$\begin{aligned} \gamma_b &= t_{eb} - \eta^b \\ \gamma_e &= t_{er} - \eta^r \\ \gamma_r &= \min(t_{er} - \eta^r, t_{ar} - \psi^r) \\ \gamma_w &= t_{aw} - \psi^w \end{aligned} \tag{232}$$

where η^b, η^r, ψ^r and ψ^w are the current values of the relevant resources. The resources were defined in the section System Equations. The lower bounds with non-zero resources are defined as follows:

$$\mathcal{L}_d{}'(x, \gamma_b) = x + t_b + t_b \left\lfloor \frac{x - \gamma_b - \epsilon}{t_{eb}} \right\rfloor, 0 \le x \le t_{ar}$$

$$\mathcal{L}_w{}'(x, \gamma_b, \gamma_r, \gamma_e) = \min_{\substack{z \le \min(\gamma_r, x) \\ \mathcal{L}_d{}'(z, \gamma_b) \le \gamma_e}} (\mathcal{L}_d{}'(z, \gamma_b) + t_r u(x - z) + \mathcal{L}_w(x - z)), \quad 0 \le x \le t_{aw}$$

$$\mathcal{L}_l{}'(x, \gamma_b, \gamma_r, \gamma_e, \gamma_w) = \min_{z \le \min(\gamma_w, x)} (\mathcal{L}_w{}'(z, \gamma_b, \gamma_r, \gamma_e) + t_w u(x - z) + \mathcal{L}_l(x - z)), \quad 0 \le x$$

The $D_{HOS}(x, \theta)$ used in A^* is given by $\mathcal{L}_l{}'(x, \gamma_b, \gamma_r, \gamma_w)$, where equations (232) are used to calculate γ_b, γ_r and γ_w from θ.

Appendix J: Dominance Rules Derivation

For convenience, the notation used in this appendix is slightly different from the rest of the paper. Consider two labels $\theta_i = \left(\eta_i^0, c_i, \eta_i^b, \eta_i^r, \psi_i^r, \psi_i^w\right), i \in \{1,2\}$, with associated slack variables $\left(\lambda_{0,i}, \sigma_{0,i}, \lambda_{r,i}, \sigma_{r,i}\right), i \in \{1,2\}$, assigned to the same node v. For the θ's and resources, the index (1 or 2) is used to differentiate between the two labels being compared. For the λ's and σ's, the first index indicates if the slack is relative to the origin (0) or the last *daily/weekly rest* (r), the second index indicates if the slack is part of label θ_1 or θ_2. We say that θ_2 is dominated by θ_1, if θ_2 cannot generate solutions better than the ones generated by θ_1.

The intuitive dominance condition is to have each resource in θ_1 be smaller or equal to the corresponding resource in θ_2, with at least one being strictly smaller. However, as the partial solutions can be improved by removing the slacks in the schedule, these slacks must also be considered. Besides that, as we impose time-window constraints on the arrival time instead of on the activity starting time, the time resource has to be treated differently from other papers that use similar dominance rules, such as (Asvin Goel 2012). Without allowing early arrival, if $\eta_1^0 < \eta_2^0$, then the solutions generated from θ_1 may be unable to satisfy the same time-windows as solutions generated from θ_2 due to arriving too early.

Equal time, no slack

The base case for the dominance check is when the complexities of time-window constraints and slacks need not be considered. Let $\lambda_{0,i} = \sigma_{0,i} = \lambda_{r,i} = \sigma_{r,i} = 0, i \in \{1,2\}$ and $\eta_1^0 = \eta_2^0$. In this case, the labels cannot be improved by removing slacks and there is no difference between the time-windows that can be satisfied by paths generated from either label. θ_2 is dominated by θ_1 if all the following conditions hold, with at least one inequality being strict:

$$\begin{aligned} \eta_1^0 &= \eta_2^0 \\ c_1 &\leq c_2 \\ \eta_1^b &\leq \eta_2^b \\ \eta_1^r &\leq \eta_2^r \\ \psi_1^r &\leq \psi_2^r \\ \psi_1^w &\leq \psi_2^w \end{aligned} \tag{233}$$

Equal time, non-zero slack, after improvement

First, consider how label improvement affects the resources. The new cost resource c_i^* after removing slacks relative to the origin is given by:

$$c_i^* = c_i + \alpha_0 \min\left(\lambda_{0,i}, \sigma_{0,i}\right) - \alpha_r \min\left(\lambda_{0,i}, \sigma_{0,i}\right) \tag{234}$$

The slack removal shifts time from activities with α_r hourly cost to one with α_0. When the slack regarding *daily/weekly rests*, $\left(\lambda_{r,i}, \sigma_{r,i}\right)$, is removed, the cost does not change as these activities

have the same hourly cost as *breaks*. In this case, to define dominance rules we pay attention to how these slacks can affect the resource η_i^r. The improved resource η_i^{r*} is given by:

$$\eta_i^{r*} = \eta_i^r - \min(\lambda_{r,i}, \sigma_{r,i}) \tag{235}$$

After label improvement we have that the new slack variables have the following property $\min(\lambda_{0,i}^*, \sigma_{0,i}^*) = \min(\lambda_{r,i}^*, \sigma_{r,i}^*) = 0, i \in \{1,2\}$.

Assume that θ_2 and θ_1 are improved labels, so $\min(\lambda_{0,i}, \sigma_{0,i}) = \min(\lambda_{r,i}, \sigma_{r,i}) = 0, i \in \{1,2\}$. If all λ's are equal to zero, it falls on the previous case where no label improvement is possible by updating upstream decisions, now or in the future. However, if that is not the case, it is still possible for the remaining slacks to be used in the future if downstream decisions make the σ's increase, e.g., if θ_2 was generated by a path with very loose time-windows, it may have large λ's; this slack can be used to reduce the amount of unnecessary off-duty time at downstream nodes. It is important to note that the usage of this remaining slack is not guaranteed. The slacks may be reset or the vehicle may reach the final destination before the σ's are increased. Therefore, the dominance criteria must consider both the case when the slack remains unused and when it is fully used. The case when the slacks remain unused is covered by (233). To account for the slack usage, the following constraints are included:

$$c_1 - c_2 \leq \quad (\alpha_0 - \alpha_r)[\lambda_{0,2} - \lambda_{0,1}] \tag{236}$$

$$\eta_1^r - \eta_2^r \leq \quad \lambda_{r,1} - \lambda_{r,2} \tag{237}$$

Note that, until the node of reference is changed, the σ's can only increase, and the λ's can only decrease. So the λ's give an upper-bound on how much future labels can be improved by updating the decision taken at the current reference node. (236) and (237) check if a path generated from θ_1 will be superior to one from θ_2 even if the slacks are fully used. A given path will cause the same resource expenditures regardless of the resource vector of the label from which it was generated, so the resource variations caused by downstream decisions do not show up in (236) and (237). θ_1 dominates θ_2 if conditions (233), (236) and (237) hold, with at least one inequality being strict. These dominance rules are valid for any node, but only for improved labels with matching η^0. As label improvement affects path feasibility and it is only performed during decisions at rest nodes, the dominance rules for labels that have not been improved yet will be treated in the next section, along with labels with different η^0.

Different time, non-zero slack, before improvement

Due to the existence of time-window constraints θ_1 cannot dominate θ_2 if $\eta_1^0 > \eta_2^0$ as some time-windows may be impossible to satisfy due to late arrival at downstream nodes. Therefore, we assume that $\gamma = \eta_2^0 - \eta_1^0 >= 0$. However, as we assume early arrivals are not allowed, having $\eta_1^0 < \eta_2^0$ will also cause problems if not treated properly. Therefore, we only allow $\gamma > 0$ at nodes where this arrival time difference can be corrected by the next decision, i.e., rest nodes. The option to extend the duration of rests makes it possible to match the arrival time of labels generated by θ_1 to the ones generated by θ_2. The dominance conditions are derived by

studying when every label generated from θ_2 through a decision $u_2 = (j, \delta)$ is dominated (after label improvements) by the label generated from θ_1 through the decision $u_1 = (j, \delta + \gamma)$. Note that θ_1 and θ_2 are not improved labels, we simply calculate the effects that label improvement could have on their descendants. Let V_b, V_r and V_w represent the set of all *break*, *daily rest* and *weekly rest* nodes, respectively. Table 32 summarizes how a decision u with duration δ_u taken at a rest node affects the resources and slack variables, the second column refers to decisions taken at *break* nodes, whereas the third column refers to decisions taken at *daily rest* or *weekly rest* nodes. This table differs slightly from Table 22 because it refers specifically to decisions taken at rest nodes, and it also includes the slack variables. Note that decisions taken at a rest node always lead to the TPL's exit (node v_i^{out} in Figure 113a), which does not have time-window constraints, and that resources requiring resetting are already zero.

Table 32. Effect of decision with duration δ_u at rest nodes

Next label info	$v \in V_b$	$v \in V_r \cup V_w$
$\eta_i^{0\prime} =$	$\eta_i^0 + \delta_u$	$\eta_i^0 + \delta_u$
$c_i{}' =$	$c_i + \alpha_r \delta_u$	$c_i + \alpha_r \delta_u$
$\eta_i^{b\prime} =$	η_i^b	η_i^b
$\eta_i^{r\prime} =$	$\eta_i^r + \delta_u$	η_i^r
$\psi_i^{r\prime} =$	ψ_i^r	ψ_i^r
$\psi_i^{w\prime} =$	ψ_i^w	ψ_i^w
$\lambda_i^{0\prime} =$	$\lambda_{0,i}$	$\lambda_{0,i}$
$\sigma_i^{0\prime} =$	$\sigma_{0,i} + \delta_u$	$\sigma_{0,i} + \delta_u$
$\lambda_i^{r\prime} =$	$\lambda_{r,i}$	$\lambda_{r,i}$
$\sigma_i^{r\prime} =$	$\sigma_{r,i} + \delta_u$	$\sigma_{r,i}$

Let the resources with superscript $*$ represent the resources generated after label expansion and label improvement, in this order. The label expansion uses the decisions listed previously ($\delta_{u_1} = \delta + \gamma$ for θ_1 and $\delta_{u_2} = \delta$ for θ_2). Label improvement effects are described in (234) and (235). If the following conditions are satisfied (with at least one inequality being a strict inequality) for all values of δ, we can say that θ_1 dominates θ_2.

$$v \in V_b \cup V_r \cup V_w \tag{238}$$

$$\eta_1^0 + \gamma + \delta = \eta_1^{0*} = \eta_2^{0*} = \eta_2^0 + \delta \tag{239}$$

$$\eta_1^b = \eta_1^{b*} \leq \eta_2^{b*} = \eta_2^b \tag{240}$$

$$\psi_1^r = \psi_1^{r*} \leq \psi_2^{r*} = \psi_2^r \tag{241}$$

$$\psi_1^w = \psi_1^{w*} \leq \psi_2^{w*} = \psi_2^w \tag{242}$$

$$c_1^* = c_1 + (\alpha_0 - \alpha_r)\min\left(\lambda_{0,1}, \sigma_{0,1} + \delta + \gamma\right) + \alpha_r(\delta + \gamma) \tag{243}$$

$$c_2^* = c_2 + (\alpha_0 - \alpha_r)\min\left(\lambda_{0,2}, \sigma_{0,2} + \delta\right) + \alpha_r\delta \tag{244}$$

$$c_1^* \leq c_2^* \tag{245}$$

$$\eta_1^{r*} = \begin{cases} \eta_1^r - \min\left(\lambda_{r,1}, \sigma_{r,1} + \delta + \gamma\right) + \delta + \gamma, \text{ if } v \in V_b \\ 0, \quad \text{if } v \in V_r \cup V_w \end{cases} \tag{246}$$

$$\eta_2^{r*} = \begin{cases} \eta_2^r - \min\left(\lambda_{r,2}, \sigma_{r,2} + \delta\right) + \delta, \text{ if } v \in V_b \\ 0, \quad \text{if } v \in V_r \cup V_w \end{cases} \tag{247}$$

$$\eta_1^{r*} \leq \eta_2^{r*} \tag{248}$$

$$c_1 - c_2 \leq (\alpha_0 - \alpha_r)\left(\lambda_{0,2} - \lambda_{0,1}\right) - \alpha_r\gamma \tag{249}$$

$$\eta_1^r - \lambda_{r,1} + \gamma \leq \eta_2^r - \lambda_{r,2}, \quad \text{if } v \in V_b \tag{250}$$

(238) defines that these rules are applicable only to rest nodes. (239) is always satisfied due to how γ is defined. (240) is always satisfied because η^b is always zero at rest nodes and does not change when the rest is extended. (241) and (242) check the conditions on the accumulated driving time since the last *daily rest* and on-duty time since the last *weekly rest*, respectively. These first four conditions are independent of the value of δ, so they can be easily checked from θ_1 and θ_2. Equations (243)-(248) check the conditions on elapsed time since the last *daily rest* (η^r) and accumulated cost (c) assuming label improvement will be performed when that label is expanded. Conditions (249) and (250) are like (236) and (237) in that they guarantee that dominance will still be valid even if remaining slacks are fully used downstream. Note that (249) and (250) are generated from (243)-(248) by using the λ's to bound the effects of future label improvements, equivalent to taking a large δ. Conditions (243)-(248) depend on δ, so we determine sufficient conditions to guarantee that they hold for every δ.

Note that, for $v \in V_b$, the structure of the conditions on η^r is the same as the ones on c. Therefore, we only describe how to find sufficient conditions to satisfy (243)-(245). Analogous conditions for (246)-(248) can be found by taking $\alpha_0 = 0$ and $\alpha_r = 1$, and by replacing the resource and slack variables by the ones relative to η^r. Eqs. (243)-(245) can be rewritten as:

$$\begin{aligned} c_1 \leq \quad & c_2 + (\alpha_0 - \alpha_r)[\min\left(\lambda_{0,2}, \sigma_{0,2} + \delta\right) \\ & -\min\left(\lambda_{0,1}, \sigma_{0,1} + \delta + \gamma\right)] - \alpha_r\gamma \end{aligned} \tag{251}$$

The subscripts '0' of the λ's and σ's will be omitted for convenience. Furthermore, it is assumed that $\alpha_0 < \alpha_r$. The sufficient conditions are determined by minimizing the right side of (251) over all values of δ.

Case 1: $\boldsymbol{\lambda_1 \le \sigma_1 + \gamma \quad \wedge \quad \lambda_2 \le \sigma_2}$

$$\begin{aligned} c_1 &\le c_2 + (\alpha_0 - \alpha_r)(\lambda_2 - \lambda_1) - \alpha_r\gamma \\ &= c_2 + \alpha_0(\lambda_2 - \lambda_1) + \alpha_r(-\lambda_2 + \lambda_1 - \gamma) \end{aligned}$$

Case 2: $\boldsymbol{\lambda_1 \le \sigma_1 + \gamma \quad \wedge \quad \lambda_2 > \sigma_2}$

$$c_1 \le c_2 + (\alpha_0 - \alpha_r)(\min(\lambda_2, \sigma_2 + \delta) - \lambda_1) - \alpha_r\gamma \tag{252}$$

We determine the minimum of the right side of (252) and use it to determine a sufficient condition independent of δ. Note that $(\alpha_0 - \alpha_r) < 0$.

$$\max_{\delta \ge 0}(\min(\lambda_2, \sigma_2 + \delta) - \lambda_1) = \lambda_2 - \lambda_1$$

$$\begin{aligned} &\therefore \\ c_1 &\le c_2 + (\alpha_0 - \alpha_r)(\lambda_2 - \lambda_1) - \alpha_r\gamma \\ &= c_2 + \alpha_0(\lambda_2 - \lambda_1) + \alpha_r(-\lambda_2 + \lambda_1 - \gamma) \\ &\le c_2 + (\alpha_0 - \alpha_r)(\min(\lambda_2, \sigma_2 + \delta) - \lambda_1) - \alpha_r\gamma \end{aligned}$$

Case 3: $\boldsymbol{\lambda_1 > \sigma_1 + \gamma \quad \wedge \quad \lambda_2 \le \sigma_2}$

$$c_1 \le c_2 + (\alpha_0 - \alpha_r)(\lambda_2 - \min(\lambda_1, \sigma_1 + \delta + \gamma)) - \alpha_r\gamma \tag{253}$$

Determine the minimum of the right side of (253).

$$\max_{\delta \ge 0}(\lambda_2 - \min(\lambda_1, \sigma_1 + \gamma + \delta)) = \lambda_2 - \sigma_1 - \gamma$$

$$\begin{aligned} &\therefore \\ c_1 &\le c_2 + (\alpha_0 - \alpha_r)(\lambda_2 - \sigma_1 - \gamma) - \alpha_r\gamma \\ &= c_2 + \alpha_0(\lambda_2 - \sigma_1 - \gamma) + \alpha_r(-\lambda_2 + \sigma_1) \\ &\le c_2 + (\alpha_0 - \alpha_r)(\lambda_2 - \min(\lambda_1, \sigma_1 + \delta + \gamma)) - \alpha_r\gamma \end{aligned}$$

Case 4: $\boldsymbol{\lambda_1 > \sigma_1 + \gamma \quad \wedge \quad \lambda_2 > \sigma_2}$

$$\begin{aligned} c_1 \le c_2 \quad &+(\alpha_0 - \alpha_r)(\min(\lambda_2, \sigma_2 + \delta) \\ &-\min(\lambda_1, \sigma_1 + \delta + \gamma)) - \alpha_r\gamma \end{aligned}$$

If $\lambda_1 - \sigma_1 - \gamma < \lambda_2 - \sigma_2$, the right side is minimized by a large δ:

$$\max_{\delta \ge 0}(\min(\lambda_2, \sigma_2 + \delta) - \min(\lambda_1, \sigma_1 + \delta + \gamma)) = \lambda_2 - \lambda_1$$

$$\begin{aligned} &\therefore \\ c_1 &\le c_2 + (\alpha_0 - \alpha_r)(\lambda_2 - \lambda_1) - \alpha_r\gamma \\ &= c_2 + \alpha_0(\lambda_2 - \lambda_1) + \alpha_r(-\lambda_2 + \lambda_1 - \gamma) \\ \le c_2 \quad &+(\alpha_0 - \alpha_r)(\min(\lambda_2, \sigma_2 + \delta) \\ &-\min(\lambda_1, \sigma_1 + \delta + \gamma)) - \alpha_r\gamma \end{aligned}$$

If $\lambda_1 - \sigma_1 - \gamma \geq \lambda_2 - \sigma_2$, the right side is minimized by $\delta = 0$:

$$\max_{\delta \geq 0}(\min(\lambda_2, \sigma_2 + \delta) - \min(\lambda_1, \sigma_1 + \delta + \gamma)) = \sigma_2 - \sigma_1 - \gamma$$

$$\begin{aligned} & \therefore \\ c_1 & \leq c_2 + (\alpha_0 - \alpha_r)(\sigma_2 - \sigma_1 - \gamma) - \alpha_r\gamma \\ & = c_2 + \alpha_0(\sigma_2 - \sigma_1 - \gamma) + \alpha_r(-\sigma_2 + \sigma_1) \\ & \leq c_2 + (\alpha_0 - \alpha_r)(\min(\lambda_2, \sigma_2 + \delta) \\ & \quad -\min(\lambda_1, \sigma_1 + \delta + \gamma)) - \alpha_r\gamma \end{aligned}$$

Summary

Conditions (243)-(248) should be replaced by the following conditions:

Cost constraint: Given $\alpha_0 < \alpha_r$.

If $\lambda_{0,1} \leq \sigma_{0,1} + \gamma$:

$$c_1 \leq c_2 + \alpha_0(\lambda_{0,2} - \lambda_{0,1}) + \alpha_r(-\lambda_{0,2} + \lambda_{0,1} - \gamma)$$

If $\lambda_{0,1} > \sigma_{0,1} + \gamma \quad \wedge \quad \lambda_{0,2} \leq \sigma_{0,2}$:

$$c_1 \leq c_2 + \alpha_0(\lambda_{0,2} - \sigma_{0,1} - \gamma) + \alpha_r(-\lambda_{0,2} + \sigma_{0,1})$$

If $\lambda_{0,1} > \sigma_{0,1} + \gamma \quad \wedge \quad \lambda_{0,2} > \sigma_{0,2}$:

$$\begin{aligned} c_1 \leq c_2 & \quad +\alpha_0(\sigma_{0,2} - \sigma_{0,1} - \gamma) + \alpha_r(-\sigma_{0,2} + \sigma_{0,1}) \\ & \wedge \\ c_1 \leq c_2 & \quad +\alpha_0(\lambda_{0,2} - \lambda_{0,1}) + \alpha_r(-\lambda_{0,2} + \lambda_{0,1} - \gamma) \end{aligned}$$

η^r constraint, for $v \in V_b$:

If $\lambda_{r,1} \leq \sigma_{r,1} + \gamma$:

$$\eta_1^r \leq \eta_2^r - \lambda_{r,2} + \lambda_{r,1} - \gamma$$

If $\lambda_{r,1} > \sigma_{r,1} + \gamma \quad \wedge \quad \lambda_{r,2} \leq \sigma_{r,2}$:

$$\eta_1^r \leq \eta_2^r - \lambda_{r,2} + \sigma_{r,1}$$

If $\lambda_{r,1} > \sigma_{r,1} + \gamma \quad \wedge \quad \lambda_{r,2} > \sigma_{r,2}$:

$$\begin{aligned} \eta_1^r \leq \eta_2^r & \quad -\sigma_{r,2} + \sigma_{r,1} \\ & \wedge \\ \eta_1^r \leq \eta_2^r & \quad -\lambda_{r,2} + \lambda_{r,1} - \gamma \end{aligned}$$

Appendix K: Optimality Proof

Definition 1: We use the term 'Restricted Problem' to refer to a sub-problem that considers a single path over the extended network $\boldsymbol{G'} = (\boldsymbol{V'}, \boldsymbol{A'})$ and a single time-window at each location, without modifying the HOS constraints. Let $\boldsymbol{v_0}$ and $\boldsymbol{v_n}$ represent, respectively, the origin and destination nodes. Let $\cup_{\tau=1}^{T_{v_j}} \left[\boldsymbol{t_{v_j\tau}^{min}}, \boldsymbol{t_{v_j\tau}^{max}}\right]$ represent the set of time-windows for node $\boldsymbol{v_j}$, where $\boldsymbol{T_{v_j}}$ is the number of disjoint time-windows at node $\boldsymbol{v_j}$.

An instance of the restricted problem is defined by a path $p = (v_0, v_1, \cdots, v_n)$ such that $(v_i, v_{i+1}) \in A'$ for all $i = 0, \cdots, n-1$, and, for each location v_j in p, a time-window $[a_j, b_j] = \left[t_{v_j\tau}^{min}, t_{v_j\tau}^{max}\right], \tau \in \{1, \cdots, T_{v_j}\}$. For nodes without time-windows, consider $[a_j, b_j] = (-\infty, +\infty)$. The HOS constraints remain the same as in the original problem.

Definition 2: *Let $\boldsymbol{\theta_k} = \left(\boldsymbol{\eta_k^0}, \boldsymbol{c_k}, \boldsymbol{\eta_k^b}, \boldsymbol{\eta_k^r}, \boldsymbol{\psi_k^r}, \boldsymbol{\psi_k^w}\right)$ be the resource vector defined in the section System Equations. Let $\left(\boldsymbol{\sigma_{i,k}}, \boldsymbol{\lambda_{i,k}}\right), \boldsymbol{i} \in \{\boldsymbol{0}, \boldsymbol{r}\}$ be the slack vectors defined in the section Label Improvement, where $\boldsymbol{r}$ is the index of the last* daily or *weekly rest* node visited. The extended state vector is described by the current location, the resource vector and the slack vector, i.e., $\boldsymbol{x_k} = \left(\boldsymbol{v_k}, \boldsymbol{\eta_k^0}, \boldsymbol{c_k}, \boldsymbol{\eta_k^b}, \boldsymbol{\eta_k^r}, \boldsymbol{\psi_k^r}, \boldsymbol{\psi_k^w}, \boldsymbol{\sigma_{0,k}}, \boldsymbol{\lambda_{0,k}}, \boldsymbol{\sigma_{r,k}}, \boldsymbol{\lambda_{r,k}}\right)$.

Definition 3: *A solution to the original problem is composed of a path and a schedule. It can be uniquely described by a vector of extended states $\mathbf{x} = (\boldsymbol{x_0}, \boldsymbol{x_1}, \cdots, \boldsymbol{x_i})$, or by a triplet of initial state $\boldsymbol{x_0}$, path and decisions vector $(\boldsymbol{x_0}, \mathbf{p_i}, \boldsymbol{\delta})$, where $\mathbf{p_i} = (\boldsymbol{v_0}, \boldsymbol{v_1}, \cdots, \boldsymbol{v_i})$ and $\boldsymbol{\delta} = (\boldsymbol{\delta_0}, \boldsymbol{\delta_1}, \cdots, \boldsymbol{\delta_{i-1}})$. When considering a restricted problem, as $\mathbf{p_i}$ is given, $(\boldsymbol{x_0}, \boldsymbol{\delta})$ is enough to represent a schedule, and, consequently, a solution.*

Definition 4: *Consider a restricted problem with path $\boldsymbol{p} = (\boldsymbol{v_0}, \cdots, \boldsymbol{v_e}, \cdots, \boldsymbol{v_n})$ and time-windows $\{[\boldsymbol{a_i}, \boldsymbol{b_i}]\}, \boldsymbol{i} = \boldsymbol{1}, \cdots, \boldsymbol{n}$. Let $\boldsymbol{D_l}(\boldsymbol{i}, \boldsymbol{e})$, $\boldsymbol{D_d}(\boldsymbol{j}, \boldsymbol{e})$, $\boldsymbol{D_s}(\boldsymbol{j}, \boldsymbol{e})$ and $\boldsymbol{D_m}(\boldsymbol{j}, \boldsymbol{e})$ be, respectively, the length, driving time, service time and the mandatory off-duty time between nodes $\boldsymbol{v_j}$ and $\boldsymbol{v_e}$. Let $\boldsymbol{V_b}$, $\boldsymbol{V_r}$ and $\boldsymbol{V_w}$ represent the set of all* break*, daily rest* and *weekly rest* nodes, respectively. Let $\boldsymbol{N_d} = \boldsymbol{V_b} \cup \boldsymbol{V_r} \cup \boldsymbol{V_w} \cup \{\boldsymbol{v_0}\}$ be the set of locations with controllable duration decisions, i.e., origin and rest nodes. Let $\boldsymbol{br} = \mathbf{max}\{\boldsymbol{i} | \boldsymbol{i} \leq \boldsymbol{e}, \boldsymbol{i} \in \boldsymbol{N_d}\}$ be the index of the last rest node, $\boldsymbol{r} = \mathbf{max}\{\boldsymbol{i} | \boldsymbol{i} \leq \boldsymbol{e}, \boldsymbol{i} \in \boldsymbol{N_d} \setminus \boldsymbol{V_b}\}$ the index of the last *daily* or *weekly rest* node, and $\boldsymbol{w} = \mathbf{max}\{\boldsymbol{i} | \boldsymbol{i} \leq \boldsymbol{e}, \boldsymbol{i} \in \boldsymbol{V_w}\}$ the index of the last *weekly rest* node.

Assuming zero initial conditions, given the partial decision vector $(\delta_0, \delta_1, \cdots, \delta_{e-1})$ for a partial schedule ending at node v_e, the resource vector θ_e and its constraints can be written as:

$$a_e \leq \eta_e^0 = \delta_0 + D_d(1,e) + D_s(1,e) + D_m(1,e) + \sum_{k \in [1,e-1],\ v_k \in N_d} \delta_k \leq b_e \qquad (254)$$

$$c_e = \alpha_0\delta_0 + \beta_d D_l(1,e) + \alpha_d D_d(1,e) + \alpha_s D_s(1,e) + \alpha_r\left(D_m(1,e) + \sum_{k\in[1,e-1],\ v_k\in N_d} \delta_k\right) \tag{255}$$

$$\eta_e^r = D_d(r,e) + D_s(r,e) + D_m(r,e) + \sum_{k\in[r+1,e-1],\ v_k\in N_d} \delta_k \le t^{er} \tag{256}$$

$$\eta_e^b = D_d(br,e) + D_s(br,e) \le t^{eb} \tag{257}$$

$$\psi_e^{r\prime} = D_d(r,e) \le t^{ar} \tag{258}$$

$$\psi_e^{w\prime} = D_d(w,e) + D_s(w,e) \le t^{aw} \tag{259}$$

Let $C(i,j)$ and $D(i,j)$ represent, respectively, the fixed part of the cost and trip duration between nodes v_i and v_j. We can rewrite (254)*,*(255) *and* (256) *as:*

$$\begin{aligned} a_e \le \eta_e^0 &= D(0,e) + \sum_{\substack{k\in[0,e-1]\\ v_k\in N_d}} \delta_k \le b_e \\ c_e &= C(0,e) + \alpha_0\delta_0 + \alpha_r \sum_{\substack{k\in[1,e-1]\\ v_k\in N_d}} \delta_k \\ \eta_e^r &= D(r,e) + \sum_{\substack{k\in[0,e-1]\\ v_k\in N_d}} \delta_k \le t^{er} \end{aligned}$$

Lemma 1: *Consider the two representations of a partial solution: $\mathbf{x} = (\boldsymbol{x_0}, \boldsymbol{x_1}, \cdots, \boldsymbol{x_e})$, and $(\boldsymbol{x_0}, \mathbf{p_e}, \boldsymbol{\delta})$, where $\mathbf{p_e} = (\boldsymbol{v_0}, \boldsymbol{v_1}, \cdots, \boldsymbol{v_e})$ and $\boldsymbol{\delta} = (\boldsymbol{\delta_0}, \boldsymbol{\delta_1}, \cdots, \boldsymbol{\delta_{e-1}})$. We have that, for every $\mathbf{0} \le \boldsymbol{i} \le \boldsymbol{j} \le \boldsymbol{e}$: $\boldsymbol{\sigma_{i,j}} = \mathbf{0} \leftrightarrow \sum_{\substack{\boldsymbol{k}\in[\boldsymbol{i+1},\boldsymbol{j-1}]\\ \boldsymbol{v_k}\in\boldsymbol{N_d}}} \boldsymbol{\delta_k} = \mathbf{0}$*

Proof. Follows from the definition of $\sigma_{i,j}$, (106) . □

Corollary 1: *Let $\mathbf{x} = (\boldsymbol{x_0}, \boldsymbol{x_1}, \cdots, \boldsymbol{x_n})$ be a solution for an instance of a restricted problem. If $\boldsymbol{\sigma_{0,n}} = \mathbf{0}$, then the solution $\mathbf{x}$ has minimum trip duration for that restricted problem.*

Lemma 2: *Consider a restricted problem with path $\mathbf{p}$ and time-windows $\{[\boldsymbol{a_i}, \boldsymbol{b_i}]\}, \boldsymbol{i} = \mathbf{1}, \cdots, \boldsymbol{n}$. Consider the two representations of a partial solution: $\mathbf{x} = (\boldsymbol{x_0}, \boldsymbol{x_1}, \cdots, \boldsymbol{x_e})$, and $(\boldsymbol{x_0}, \boldsymbol{\delta})$, where $\boldsymbol{\delta} = (\boldsymbol{\delta_0}, \boldsymbol{\delta_1}, \cdots, \boldsymbol{\delta_{e-1}})$. We have that, for every $\mathbf{0} \le \boldsymbol{i} \le \boldsymbol{j} \le \boldsymbol{e} \le \boldsymbol{n}$: $\boldsymbol{\lambda_{i,j}} = \mathbf{0} \leftrightarrow \exists \boldsymbol{k} \in [\boldsymbol{i} + \mathbf{1}, \boldsymbol{j}], \boldsymbol{\eta_k^0} = \boldsymbol{b_k}$* and *$\boldsymbol{\sigma_{i,k}} = \mathbf{0}$*

Proof. Follows from the definition of $\lambda_{i,j}$, (107). $\lambda_{i,i+1}$ is set to ∞ or $b_{i+1} - \eta_{i+1}^0 + \sigma_{i,i+1}$ at node v_{i+1}, and, at every node v_k, $i < k \le j$, $\lambda_{i,k}$ is set to $\min\left(\lambda_{i,k-1}, b_k - \eta_k^0 + \sigma_{i,k}\right)$. So, if $\lambda_{i,j} = 0$,

then $b_k - \eta_k^0 + \sigma_{i,k}$ must be zero for at least one node v_k, $i < k \leq j$. This can only happen if $\eta_k^0 = b_k$ and $\sigma_{i,k} = 0$. □

Proposition 1: *Consider a restricted problem with path* $\mathbf{p}$ *and time-windows* $\{[\boldsymbol{a_i}, \boldsymbol{b_i}]\}, \boldsymbol{i} = \boldsymbol{1}, \cdots, \boldsymbol{n}$. *The state vector* $(\boldsymbol{x_0}, \boldsymbol{x_1}, \cdots, \boldsymbol{x_e})$ *and decision vector* $(\boldsymbol{\delta_0}, \boldsymbol{\delta_1}, \cdots, \boldsymbol{\delta_{e-1}})$ *representing the partial schedule found at the end of each iteration have the following properties:*

h) $\min(\sigma_{0,i}, \lambda_{0,i}) = 0, \ \forall 1 \leq i \leq e$

i) $\min(\sigma_{r,i}, \lambda_{r,i}) = 0, \ \forall 1 \leq i \leq e, r = \max\{k | k \leq i, v_k \in V_r \cup V_w \cup \{v_0\}\}.$

j) $\left(\delta_i > 0 \rightarrow \exists j > i, \eta_j^0 = a_j \text{ and } \exists l < i, \eta_l^0 = b_l\right), \ \forall 1 \leq i \leq e, v_i \in N_d.$

k) $\delta_0 > 0 \rightarrow \exists j > 0, \eta_j^0 = a_j.$

[Note that in (c) and (d), if $j > e$, η_j^0 *is the earliest possible arrival time at* v_j *given the current partial schedule, i.e.,* $\eta_j^0 = \eta_e^0 + D(e, j)$.*]*

Proof. Proposition 1a follows from the label improvement performed when labels for TPLs' exits are generated. Only rest extension decisions can increase the σ's, whereas the λ's can only decrease. So, if the slack is removed at a TPL's exit node, it cannot increase until the next TPL exit is visited. As the label improvement is performed at every TPL exit, the slack is always zero. The same reasoning applies to Proposition 1b. Proposition 1c follows from the heuristic used to choose the decisions to test and the effects of label improvement. The chosen rest extension is the smallest duration required to satisfy downstream time-windows, however, if this value would generate a slack in the schedule, the label improvement procedure updates the schedule to remove this slack. Therefore, if the rest extension is greater than zero even after the label improvement, it means that both λ's are zero, and, by Lemma 2, there is an upstream node v_l with $\eta_l^0 = b_l$. Furthermore, as δ_i is the smallest value that satisfies downstream time-windows, then at least one downstream location v_j must have $\eta_j^0 = a_j$, otherwise δ_i could be reduced. The same reasoning applies to Proposition 1d. □

Proposition 2: *If a feasible schedule exists, the algorithm finds a schedule with minimum trip duration for the restricted problem (fixed path, single time-windows).*

Proof. At the last iteration we have that, by Proposition 1a, $\min(\sigma_{0,n}, \lambda_{0,n}) = 0$. If $\sigma_{0,n} = 0$ then the schedule has no nonmandatory off-duty time, and the schedule has minimum trip duration. If $\sigma_{0,n} > 0$, then at least one nonmandatory off-duty time is nonzero. Let $i = \max\{k | \delta_k > 0\}$. By , we have that $\exists j > i$, $\eta_j^0 = a_j$, and that $\min(\sigma_{0,j}, \lambda_{0,j}) = 0$. As $\delta_i > 0$, we have that $\sigma_{0,j} > 0$ and $\lambda_{0,j} = 0$. This implies that δ_0 and η_1^0 cannot be increased due to time-window constraints between nodes v_0 and v_j. As η_1^0 cannot be increased and η_j^0 cannot be decreased, the trip duration at v_j, given by $\eta_j^0 - \eta_1^0$, is optimal. If $j = n$, then the trip duration is optimal. If $j < n$, we have that, by the definition of i, $\delta_k = 0$ for all $k \in [i + 1, n - 1]$. As no nonmandatory off-duty time is included between v_j and v_n, the trip duration at v_n is also optimal. □

Proposition 3: *If a feasible schedule exists, the algorithm finds a minimum cost schedule for the restricted problem (fixed path, single time-windows).*

Proof. From proposition Proposition 2 we have that $\eta_n^0 - \eta_1^0$ is minimum, and, consequently, $\sum_{k\in[1,n-1]} \delta_k$ is also minimum. From Proposition 1, we have that, if $\delta_0 > 0$ then $\exists j > 0$, $\eta_j^0 = \alpha_j$, and δ_0 cannot be decreased without increasing $\sum_{k\in[1,j-1]} \delta_k$. However, as $\alpha_0 < \alpha_r$, decreasing δ_0 by increasing the other δ's would only increase costs. Therefore, the solution has minimum cost. □

Proposition 4: *If a solution exists, the algorithm finds a minimum cost solution (path and schedule) for the unrestricted problem.*

Proof. As described in the section Expansion Criteria, when expanding a label, the label correcting algorithm explores every outgoing edge that can generate a feasible solution. Therefore, all paths are considered by the algorithm. In addition, at every decision with a controllable time duration, the algorithm tests the minimum values able to satisfy each time-window of each reachable downstream location, such that all possible time-window usages are accounted for. Each combination of path and time-windows used represents an instance of the restricted problem. The algorithm discards partial paths and schedules that are infeasible, dominated or unable to improve the current upper-bound (A*). Therefore, not all instances of the restricted problem are explicitly generated. However, if the dominance rules and the estimate used by the A* method are correctly defined and do not discard good solutions, then if a solution exists the restricted problem instance containing the optimal solution (or a solution within the chosen tolerance) is guaranteed to be generated.

Proposition 3 states that the algorithm finds a feasible solution for the restricted problem if one exists. As the algorithm explores all restricted problem instances that might contain the optimal solution, then the optimal solution to the original problem is found by choosing the best solution among restricted problems' optimal solutions. □

Appendix L: Label Improvement Pseudocode

Function *updatePath* takes a partial solution given by $\left((x_j)_{0\leq j\leq k};\ (\delta_j)_{0\leq j<k}\right)$, increases the decision at node v_i by h, then updates the rest of the partial solution, returning a new one. As described in the section Label Improvement, the new partial solution follows the same path as the original one, and each decision with controllable duration takes the smallest value that will not decrease the arrival time (η^0) at the following node. We assume that $v_j \in N_d$ can be easily verified.

Function updatePath

input : $(x_j)_{0\leq j\leq k};\ (\delta_j)_{0\leq j<k};\ h;\ i;$
output: $(\hat{x}_j)_{0\leq j\leq k};\ (\hat{\delta}_j)_{0\leq j<k};$
/* $x_j = (v_j, (\eta_j^0, c_j, \eta_j^b, \eta_j^r, \psi_j^r, \psi_j^w))$, and $\hat{x}_j$ follows the same format. */
1 **for** $j \leftarrow 0$ **to** $i-1$ **do**
2 $\quad \hat{x}_j \leftarrow x_j;$
3 $\quad \hat{\delta}_j \leftarrow \delta_j;$
4 **end**
5 $\hat{x}_i \leftarrow x_i;$
6 $\hat{\delta}_i \leftarrow \delta_i + h;$
7 $\hat{x}_{i+1} \leftarrow f(\hat{x}_i, (v_{i+1}, \hat{\delta}_i));$
8 **for** $j \leftarrow i+1$ **to** $k-1$ **do**
9 $\quad$ **if** $v_j \in N_d$ **then**
10 $\quad\quad p \leftarrow \eta_{j+1}^0 - \hat{\eta}_j^0;$
11 $\quad\quad \hat{\delta}_j \leftarrow \max(p, 0);$
12 $\quad$ **else**
13 $\quad\quad \hat{\delta}_j \leftarrow \delta_j;$
14 $\quad$ **end**
15 $\quad \hat{x}_{j+1} \leftarrow f(\hat{x}_j, (v_{j+1}, \hat{\delta}_j));$
16 **end**

Function *expand* takes a label's information (partial solution, slack variables and last *daily/weekly rest* node), along with the set of decisions to be tested ($\widehat{U}(\tilde{x}_i)$ as defined in the section Expansion Criteria) and generates new partial solutions/labels, performing label improvement when necessary. Note that, as described in the section Expansion Criteria, if the current label does not require a time duration decision, all decisions are tested, i.e., $\widehat{U}(\tilde{x}_i) = U(x_i)$. Lines 1-6 expand labels that do not require time duration decisions. We assume that $v_i \notin N_d$ can be easily verified. Line 9 calculates by how much the departure time from the origin should be delayed to arrive at the next node at the desired time without slacks relative to the origin. Lines 10-16 generate an updated partial solution with the new departure time. Lines 17 and 23 recalculate the duration of the decision required to reach the next node at the desired time accounting for the updates performed. Lines 18-22 do the same as lines 9-16 but with regards to the last *daily/weekly rest* node instead of the origin node. Lines 24-26 use the

updated path and decision to calculate the next state and generate the label for the expanded partial solution, which is then inserted into *OPEN*. For simplicity, we consider that any necessary checks, such as A* bound conditions and dominance rules, are included in the process of creating a label and inserting it into *OPEN*.

Function expand

input : $(x_j)_{0\le j\le i}; (\delta_j)_{0\le j<i};\ (\lambda_{0i}, \sigma_{0i}, \lambda_{ri}, \sigma_{ri});\ r;\ \hat{U}(\tilde{x}_i)$
output: None
/* $x_j = (v_j, (\eta_j^0, c_j, \eta_j^b, \eta_j^r, \psi_j^r, \psi_j^w))$, and $\hat{x}_j$ follows the same format. */
if $v_i \notin N_d$ **then**
  for $(v_{i+1}, \delta_i) \in \hat{U}(\tilde{x}_i)$ **do**
    $x_{i+1} \leftarrow f(x_i, (v_{i+1}, \delta_i))$;
    Use $(x_j)_{0\le j\le i+1}$, $(\delta_j)_{0\le j<i+1}$ to create a new label;
    Insert new label into *OPEN*;
  end
else
  for $(v_{i+1}, \delta_i) \in \hat{U}(\tilde{x}_i)$ **do**
    $h_0 \leftarrow \min(\lambda_{0i}, \sigma_{0i} + \delta_i)$;
    if $h_0 > 0$ **then**
      $(\hat{x}_j)_{0\le j\le i}, (\hat{\delta}_j)_{0\le j<i} \leftarrow$ updatePath ($(x_j)_{0\le j\le i}$, $(\delta_j)_{0\le j<i}$, h_0, 0) ;
      Calculate new slacks $(\hat{\lambda}_{0i}, \hat{\sigma}_{0i}, \hat{\lambda}_{ri}, \hat{\sigma}_{ri})$;
    else
      $(\hat{x}_j)_{0\le j\le i}, (\hat{\delta}_j)_{0\le j<i} \leftarrow (x_j)_{0\le j\le i}, (\delta_j)_{0\le j<i}$;
      $(\hat{\lambda}_{0i}, \hat{\sigma}_{0i}, \hat{\lambda}_{ri}, \hat{\sigma}_{ri}) \leftarrow (\lambda_{0i}, \sigma_{0i}, \lambda_{ri}, \sigma_{ri})$;
    end
    $\hat{\delta}_i \leftarrow \delta_i - (\hat{\eta}_i^0 - \eta_i^0)$;
    $h_r \leftarrow \min(\hat{\lambda}_{ri}, \hat{\sigma}_{ri} + \hat{\delta}_i)$;
    $temp \leftarrow \hat{\eta}_i^0$;
    if $h_r > 0$ **then**
      $(\hat{x}_j)_{0\le j\le i}, (\hat{\delta}_j)_{0\le j<i} \leftarrow$ updatePath ($(\hat{x}_j)_{0\le j\le i}$, $(\hat{\delta}_j)_{0\le j<i}$, h_r, r) ;
    end
    $\hat{\delta}_i \leftarrow \hat{\delta}_i - (\hat{\eta}_i^0 - temp)$;
    $\hat{x}_{i+1} \leftarrow f(\hat{x}_i, (v_{i+1}, \hat{\delta}_i))$;
    Use $(\hat{x}_j)_{0\le j\le i+1}$, $(\hat{\delta}_j)_{0\le j<i+1}$ to create a new label;
    Insert new label into *OPEN*;
  end
end

Appendix M: Path Choice

The choice of paths between clients to consider in the problem can affect both computational effort and solution quality. In this experiment, we analyze the effects of a simple heuristic that can be used to define these paths. It is important to note that trucks already face a number of restrictions regarding which roads they can use, such as weight and size limits. Therefore, before anything else, restricted roads need not be considered. As we are dealing with long-haul trucking, the focus is on large highways as opposed to small local streets. The networks used in this experiment have only a single client and are organized in layers, as the networks generated for Section Randomized Networks Experiments. The parameters were set so that the networks generated represent trips where, before the branches linking to TPLs are added, the shortest routes have around 40-44 hours of driving time and thousands of paths are available.

The heuristic used to reduce the networks in this experiment is the following:

l) Estimate a trip duration upper bound (u_t) if one is not known.

1. Before including the TPLs, calculate the shortest path to the destination.
2. Remove all edges that are not part of the shortest path.
3. Include the TPLs in the remaining edges.
4. Solve several instances of this problem and take the worst trip duration as an approximate upper bound.

m) Estimate a driving time upper bound (u_d) based on u_t.

1. Appendix I presented a method that, given a driving time x, calculates a lower bound $\mathcal{L}_l(x)$ (defined in (231)) for the trip duration when only HOS rules are considered. That method can be used to find the largest driving time that generates a lower bound smaller or equal to u_t, i.e., $u_d = \sup(\{x|\mathcal{L}_l(x) \leq u_t\})$.

n) Calculate the k shortest paths with driving time not exceeding the driving time upper bound u_d.

o) Remove all edges not present in any of the k paths calculated previously.

p) Include the TPLs on the remaining edges.

Note that all paths generated by the remaining edges will be considered during execution, even if they are not among the k shortest paths. Step 3 could simply take the k shortest paths without restricting their drive time, allowing steps 1 and 2 to be skipped. On the other hand, step 1 can use a larger number of paths to get a tighter bound.

These experiments were run on a different computer from the ones in the section Randomized Networks Experiments, and the networks used are also different, so the running times are not directly comparable. The intent is to study the path choice's impact on running time and solution quality. Figure 147, Figure 148, Figure 149 show, for 4 different networks, the trip duration, running time and driving time when the edges included in the network are chosen

using the previous heuristic with k equal to 1, 5, 10, 50, 100, 500 and 1000. It can be seen that, except for network 0, the improvements in trip duration stop when more than 50 paths are considered. So, in the networks tested, it would be reasonable to assume that stopping at k equals 50 would provide a good balance between solution quality and running time. Naturally, this behavior depends on the characteristics of the networks in question. Similar experiments should be performed on the network of interest to define the appropriate number of paths to be used according to the user's sensitivity to solution quality and running time. For example, the running time increases with the number of paths considered for most networks in Figure 148. However, for network 1, it actually decreased when going from 10 to 50 paths. For the other networks, users need to consider whether the expected cost decrease is worth the running time increase, but for network 1, there is no reason to use only 5 or 10 paths, instead of 50. In this case, as can be seen from the cost drop in Figure 147, it is likely that one or more edges not part of the 10 shortest paths had convenient parking locations that allowed significantly more efficient solutions. The drop in running time seen in Figure 148 is likely due to the new optimal solutions being able to dominate other solutions faster than before. User knowledge can also be used to supplement the heuristic. Roads deemed promising by the user can be included regardless of being part of the k shortest paths.

When considering multi-client trips, a possible approach is to perform this heuristic for each pair of consecutive clients with some small adjustments. If the trip leg always starts with all HOS counters set to zero, the trip durations found will be too optimistic as they do not take into account resources spent in previous trip legs. In this case, as these bounds depend on the time and HOS counters at the trip leg's start, it makes more sense to skip the calculation of trip duration and driving time upper bounds, and just take the k shortest paths. Another option would be to randomize the HOS resources and time at the start of the trip leg, calculate the driving time upper bound for each case accounting for the initial conditions (method from Appendix I can be used), then choose the driving time upper bound from among these case-specific ones.

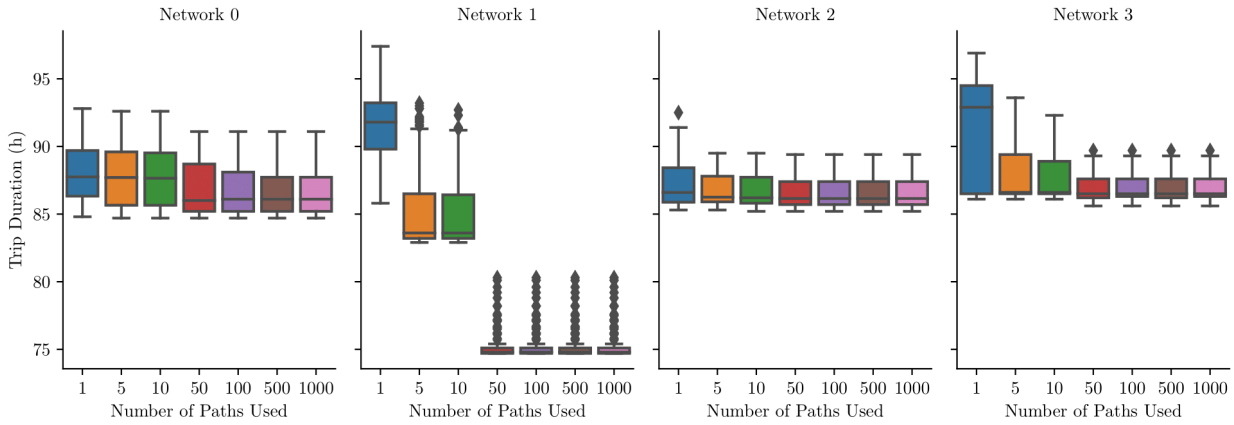


Figure 147. Trip duration, in hours, when only the edges of the k shortest paths are included in the networks.

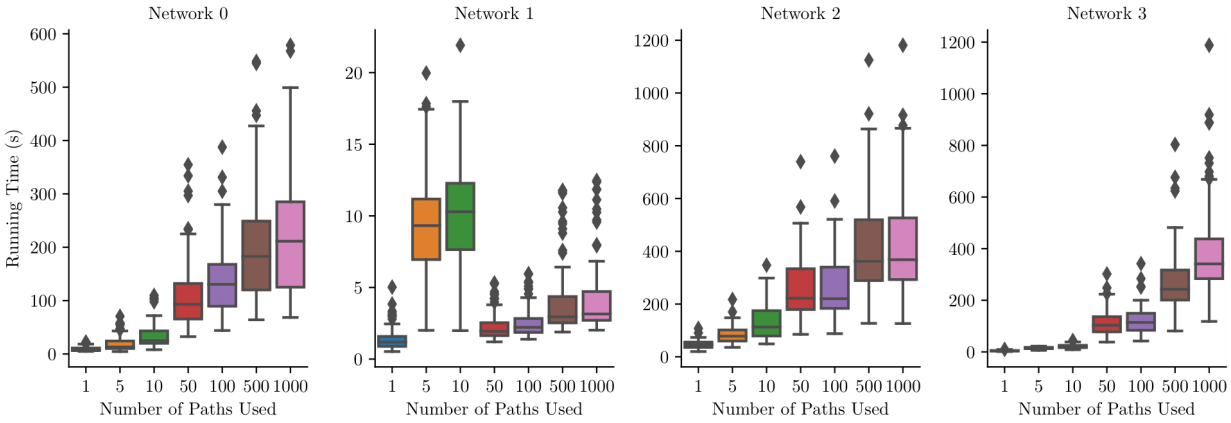


Figure 148. Running time, in seconds, when only the edges of the k shortest paths are included in the networks.

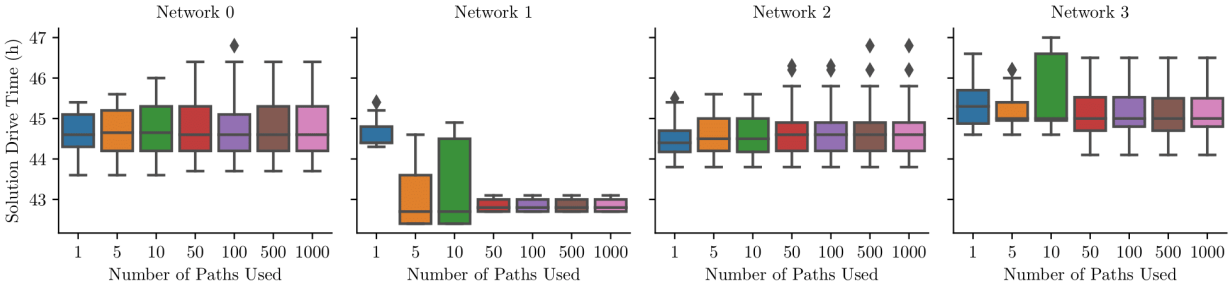


Figure 149. Solution's driving time, in hours, when only the edges of the k shortest paths are included in the networks.