Multi-Objective Optimization of Pricing Strategies for Sustainable Transportation

by

Jessica Lazarus

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil & Environmental Engineering

and the Designated Emphasis

in

Global Metropolitan Studies

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alexandre Bayen, Co-chair
Professor Susan Shaheen, Co-chair
Professor Daniel Rodriguez
Professor Joan Walker

Fall 2022

Multi-Objective Optimization of Pricing Strategies for Sustainable Transportation

Abstract

Multi-Objective Optimization of Pricing Strategies for Sustainable Transportation

by

Jessica Lazarus

Doctor of Philosophy in Engineering - Civil & Environmental Engineering

and the Designated Emphasis in

Global Metropolitan Studies

University of California, Berkeley

Professor Alexandre Bayen, Co-chair

Professor Susan Shaheen, Co-chair

The digitization and automation of transportation systems is fundamentally transforming the transportation landscape, creating important opportunities and challenges for improving the sustainability of transportation worldwide. There are more travel options to choose from than ever before, with real-time information on the cost and time trade-offs between alternative routes and modes of travel available in the palm of a hand. In particular, shared on-demand mobility services such as transportation network companies (TNC) (e.g., Lyft, Uber), bikesharing, and microtransit offer affordable and convenient alternatives to personal auto ownership that can complement or supplement existing public transit services. In addition, these services may aid in reducing existing inequities in access to fast and reliable transportation. However, despite the potential that innovative shared mobility service models bring forth to improve the sustainability of transportation, the inefficiencies of fleet-based services such as TNCs and the low adoption rates of pooled rides that transport multiple travelers in the same vehicle have contributed to worsening congestion in several regions across the United States. Meanwhile, the design and deployment of transportation demand management (TDM) strategies has not kept pace of disruption nor the corresponding evolution in travel behavior.

In light of the pressing need to improve the sustainability of a rapidly evolving transportation ecosystem, this dissertation contributes to the theory, methodology, and state of knowledge of optimal mechanism design for multi-faceted TDM strategies. With a focus on congestion pricing strategies, this research aims to facilitate the design and analysis of data-driven TDM

strategies that incorporate a multitude of policy levers (e.g., congestion pricing, multi-modal incentives, public transit operations) using a simulation-based multi-objective optimization approach to inform decision-makers about the inherent trade-offs between congestion and emission reductions, economic feasibility, and transportation equity.

In particular, I focus on the optimization of congestion pricing and targeted incentive schemes in multi-modal transportation networks including pooled ride options. Congestion pricing aims to reduce congestion by charging road users for driving on congested roads. A review of the various aspects of the transportation pricing optimization problem as studied in the literature is presented in Chapter 2, including the specification and analysis of various dimensions of demand sensitivity, congestion pricing structure and charging zone design, optimization objectives, optimization approaches, and transportation equity analysis. Studies of the optimization of pricing structures and charging zones for congestion pricing schemes have established that greater toll levels and charging zone coverage produce greater reductions in driving, which is deemed beneficial with respect to total system-wide travel time reductions. However, as is pointed out by public acceptance studies and literature on the equity implications of congestion pricing, the cost burden of congestion mitigation is disparately borne by lower-income individuals who are the most likely to be financially incentivized to adopt less desirable alternatives to driving. Few congestion pricing optimization studies have incorporated transportation equity objectives; none have included equity in addition to other efficiency and environmental objectives. The contributions of this dissertation to the literature on congestion pricing optimization span the theory of optimal mechanism design for multi-objective congestion pricing strategies, methodology for simulation-based multi-objective optimization of multi-faceted TDM strategies, and empirical understanding of congestion pricing strategies optimized with respect to multiple policy objectives.

Mechanism Design for Optimal Congestion Pricing Policies

Chapter 3 establishes that equity-based objectives and the inclusion of monetary incentives for the adoption of driving alternatives are feasible strategies for tackling the equity issues inherent in congestion pricing optimization (i.e., the disparate distributions of increased costs for lower income drivers and reduced travel times for higher income drivers). In this chapter, I formulate a bi-level optimization problem to compute optimal link- and mode-specific tolls and targeted mode-specific incentives (i.e., direct monetary transfers) using aggregate-level information on the flows of network users using various modes of transportation. This work contributes to the theory of congestion pricing optimization by proving the existence of optimal multi-modal congestion pricing schemes including both tolls and monetary incentives that are optimized with respect to equity-focused objectives defined on the basis of the distributional impacts of the pricing scheme across travelers. Several functions inspired by different theories of justice are presented as alternative social objective functions, including:

1. Utilitarian: maximize the sum of individual utility (i.e., a quasilinear function of travel

time, cost, and other factors weighted by the individual demand sensitivity to each) of travel,

2. Egalitarian: maximize the sum of individual utility using the average demand sensitivity to weight travel time cost,
3. Equality of Opportunity: maximize a weighted sum of the individual utility, again using the average demand sensitivity, with weights applied to the overall utility that are scaled according to a societal ordering of disadvantage across network users (e.g., weights decrease with income, weights increase for residents of historically underserved neighborhoods or members of historically underserved communities), and
4. Rawlsian: maximize the minimum individual utility of travel, again using the average demand sensitivity.

Consideration of the three latter social choice functions offers a deviation from the traditionally utilitarian view on congestion pricing optimization. Implementation of the bi-level optimization problem presented in chapter 3 and analysis of the impacts of each of the alternative objective functions on the optimal traffic assignment, toll, and incentive scheme are under development for future work.

The Berkeley Integrated System for TRansportation Optimization (BISTRO)

In chapter 4, I present a methodological framework for regional-scale multi-objective optimization of transportation systems using agent-based simulation (ABS), called the Berkeley Integrated System for TRansportation Optimization (BISTRO). BISTRO is an open source transportation planning decision support system that enables the simultaneous optimization of multiple transportation system interventions, including adjustment to public transit operations and vehicle fleet mixes as well as pricing strategies, using state-of-the-art activity-based travel models and ABS to explore the individual- and system-level impacts of packages of TDM strategies. In contrast to the state of the practice in transportation planning of analyzing a select few strategies across a discrete set of scenarios, the simulation-based multi-objective optimization approach that I explore in this dissertation using BISTRO leverages the machine learning principles of data exploration and exploitation to explore the interactions of policy design parameters and their outcomes to a much greater extent, thereby producing richer insights about the optimal trade-offs between the various competing objectives of policy design and implementation. BISTRO was developed in partnership with researchers at the Institute of Transportation Studies at UC Berkeley and Lawrence Berkeley National Laboratory as well as a project team at Uber, Inc. I led the development of the objective function design for BISTRO by translating municipal and regional transportation goals into quantitative optimization objectives with consideration for social, environmental, and economic sustainability implications and testing the efficacy with which these key performance

indicators (KPIs) functioned when applied in simulation-based multi-objective optimization. In addition, I led the design of the output analysis and visualization platform for BISTRO. My individual contributions are presented in the Scoring function design, Inputs, and Output analysis and visualization sections of chapter 4 in addition to the corresponding sections of Appendix A. Chapter 4 also presents the results of a pilot study of BISTRO that was conducted as a machine learning competition hosted within Uber Technologies, Inc. The key lessons learned from the design and execution of this pilot study demonstrated BISTRO's utility as a human-in-the-loop cyber-physical system: one that uses scenario-based optimization algorithms as a feedback mechanism to assist urban planners with iteratively refining objective function and constraint specification on multi-faceted intervention strategies (e.g., pricing, public transit scheduling, vehicle fleet mix). The remainder of my dissertation research focuses on the application of BISTRO for the design of congestion pricing strategies with mode-specific incentives.

Congestion Pricing Optimization Case Study I: Sioux Faux

The fifth chapter of this dissertation presents a case study of the optimization of congestion pricing policy design using BISTRO and the Sioux Faux benchmark model presented in Chapter 4. The study exemplifies how the granularity offered by activity-based travel models can be leveraged to enhance the interpretability of multi-objective transportation policy optimization through rich analyses of the effects of policy design on both individual- and system-level outcomes. The location and size of a circular charging zone as well as two different pricing schemes (a cordon fee and a cordoned mileage fee) were encoded as inputs to an ABS with an activity based travel model of 15,000 travelers. Through an analysis of the effects of various weighting schemes across KPIs representing congestion- (i.e., total VMT, average vehicle hours of delay (VD) per passenger trip, and total GHG emissions), social- (i.e., average generalized travel cost burden of work and secondary trips), and revenue-based (i.e., total toll revenue) objectives, I developed a method for interpretation of the inherent trade-offs in transportation policy optimization and demonstrate the importance of cultivating transparency in policy decision support systems that use black-box optimization in order to produce explainable, defensible policy strategies.

I find that cordoned mileage fees Pareto-dominate cordon tolls, meaning they produce greater improvements across all objectives studied. I estimated an empirical 3-dimensional Laffer curve for each pricing strategy, representing the concave relationship between the main components of the pricing scheme design - the average toll paid per driving/TNC trip and the coverage of the cordon (in % of trips affected) - with the toll revenue and the driving/TNC mode share. I analyzed the relationships between these design parameters and the other KPIs, demonstrating the utility of simulation-based multi-objective optimization for human-in-the-loop transportation policy design. I find that the prioritization of congestion reduction poses a challenge for congestion pricing optimization in that it may result in unnecessarily large mode shifts away from driving. This significantly worsens the cost burden of travel

(i.e., the total generalized cost of travel - including both travel cost and VoT-weighted travel time - as a portion of individual income) by disproportionately increasing the travel times of lower income individuals in a manner that may arguably be incommensurate with the benefits of congestion mitigation. Finally, I explore the role of weighting schemes in shifting the priority of an optimal congestion pricing scheme to social equity while maintaining congestion mitigation and toll revenue, the latter of which may be used to further improve the transportation system.

Congestion Pricing Optimization Case Study II: The San Francisco Bay Area

In chapter 6, I build upon the Sioux Faux case study by designing a Pareto-based optimization of a cordoned mileage fee and incentive scheme for the City and County of San Francisco using BISTRO. I developed and calibrated an activity-based model of 50,000 commuters in the San Francisco Bay Area for the case study, including both ride alone and pooled on-demand rides in 7-seater vehicles and a multinomial logit mode choice model with coefficients estimated from a general population stated preference survey of San Francisco Bay Area residents that I conducted in 2018. I applied the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm to optimize the hypervolume of the Pareto frontier of various transportation system objectives, including the minimization of the total VMT, average VD per passenger trip, total GHG and $PM_{2.5}$ emissions, average cost burden, and net public revenue resulting from the congestion pricing and incentive scheme. Rather than produce a single 'optimal' policy design, the approach taken in this study produces rich insights regarding the correlations between policy design parameters and the corresponding outcomes and produces a set of policy design options representing the optimal trade-offs between policy objectives.

I find that the inclusion of monetary incentives for pooled TNCs and public transit in a congestion pricing scheme for San Francisco improves all KPIs under consideration except for the net public revenue produced by the scheme, which may still be improved with respect to a baseline with neither tolls nor incentives. Incentives serve to amplify the operational and environmental benefits of congestion pricing by offering an additional positive financial incentive for the use of pooled modes of transportation (in addition to the negative incentive posed by the congestion charge), while reducing the average cost burden of travel across the region. This is particularly important in the San Francisco Bay Area where a majority of commuters traveling in the City of San Francisco either live or work in one of the other eight counties of the region where there is comparatively poor access to alternatives to driving. Incentives aid in compensating the most cost-sensitive drivers for the considerable travel time or cost increases they experience when shifting to public transit or pooled on-demand rides, respectively. This study demonstrates the value of Pareto-based optimization using ABS for developing dynamic TDM strategies capable of adapting to ever-changing travel behavior and policy priorities. I conclude this chapter with a discussion of actionable

insights regarding the potential of congestion pricing and incentive schemes to achieve a variety of policy objectives, particularly in the presence of high-capacity pooling services such as microtransit.

Summary of Key Findings and Research Directions

This dissertation develops theoretical, methodological, and empirical contributions to the field of transportation engineering, specifically for the design and optimization of sustainable and equitable congestion pricing strategies. The multi-objective simulation-based optimization framework I have developed and applied using BISTRO lays the foundation for the continued development of algorithmic transportation policy decision support tools.

The use of activity-based travel models and ABS enables the examination of the network effects of pricing strategies, which are particularly relevant when considering 1) the regional impacts of pricing on mode shifts, since behavioral responses in one sub-region can influence those of another through the secondary effects of the pricing strategy on travel times and the level of service of alternative modes (e.g., reliability of public transit or on-demand ride services), 2) on-demand vehicle fleet operations, which are designed to optimize the revenue generated from the strategic positioning and dispatching of vehicles across a market, and 3) the economies of scale of pooling, which generate a virtuous cycle of higher pooled ride match rates, better pooled ride service reliability and affordability, and reduced costs with respect to greater density of pooled rides in any given area and time period. This dissertation presents the first multi-objective congestion pricing and incentive scheme optimization studies using ABS with an activity-based travel model that includes numerous travel options to access/egress public transit and, in the second case study, includes pooled on-demand ride services with survey-based estimates of the sensitivity of demand for pooling incorporated into the behavioral dynamics of the model.

This research has produced applicable insights for both the immediate- and long-term. While congestion pricing policies are currently under development in cities across the globe, there is also a broad trend toward 'smart city' technologies that has been underway for over a decade, with increasing integration of sensors and connected control mechanisms to automate the operation of certain civil systems such as water, energy, and traffic management. In addition to updating traffic control systems (i.e., automated ramp metering, traffic lights, etc.), transportation agencies are integrating fare payment across transportation services and facilities and considering mobility as a service (MaaS) models that bundle transportation services and enable the seamless allocation of credits and incentives to nudge travel behavior.

The results of the two congestion pricing optimization case studies (chapters 5 and 6) contribute valuable insights for the ongoing development of transportation pricing strategies such as congestion pricing and MaaS. The San Francisco Bay Area case study in chapter 6 suggests that low-income incentive schemes can drastically improve the efficacy of congestion pricing to achieve congestion mitigation and environmental emissions reductions objectives

while also reducing the average cost burden of travel. These benefits come at the cost of reduced public revenues that could otherwise be used to expand and improve public transit access. Targeted public transit improvements may also serve to reduce the cost burden of congestion pricing schemes by reducing the travel time for individuals who shift away from driving due to the congestion charges. However, in a region as large and varied as the San Francisco Bay Area, strategizing which investments to make and where to make them is a challenging task. The optimization of targeted reinvestment of congestion charging revenue into public transit is an area ripe for future work using BISTRO. As was demonstrated in the pilot BISTRO study (chapter 4), BISTRO can be used to optimize the public transit vehicle fleet mix, route schedules and headways, and fares in addition to the pricing parameters optimized in the congestion pricing case studies.

In addition to generating practical insights for the design of congestion pricing schemes, this dissertation demonstrates the challenges and opportunities for algorithmic policy design. Building upon the lessons learned from the BISTRO pilot study presented in chapter 4, the methodological approach developed in chapters 5 and 6 focuses on improving the explainability and interpretability of simulation-based multi-objective optimization by using BISTRO to explore the various aspects of the transportation system optimization problem, including the:

- design of policy input parameters and an appropriate search space that is both large enough to enable exploration of potentially unexpected solutions and constrained enough to avoid the unnecessary use of resources exploring extreme solutions with undesirable outcomes,
- specification of KPIs that are both mathematically sound and representative of realistic policy objectives (spanning congestion mitigation, equity, and financial feasibility) with which to guide the optimization algorithm to find insightful solutions,
- development of a realistic and meaningful travel model that includes key behavioral, physical, and operational dynamics relevant to the optimization problem at hand (e.g., configuring the mix of fuel consumption and seating capacity in the public transit and on-demand vehicle fleets),
- implications of scalarization schemes and the impacts of weighting conflicting policy objectives on the results of optimization, and
- application of nuanced analysis to interpret and communicate the implications of the optimization results at both the individual- and system-levels.

The research presented in this dissertation offers a launching point for future work in the field of transportation policy design and optimization. BISTRO offers the capability to expand the scope of the congestion pricing optimization problem I have studied thus far to include

aspects such as induced demand, vehicle-based charging, dynamic pricing structures, parking pricing, and public transit investment. In addition, BISTRO may be used as a tool to develop methodological approaches for *robust* transportation system optimization that produces strategies that perform optimally across a variety of scenarios. Such scenario-based optimization may be conducted across multiple baselines (as opposed to a single BAU), each representing a potential variation in the distribution of land use, mobility preferences, or other factors that may affect the distribution of transportation supply or demand. BISTRO can also be used to develop algorithms for transfer learning of policies across varying scenarios and geographies. Such methodologies would characterize hyper-parameters representing factors in the transportation network, services, and demand profile that are significant across scenarios or geographies, thereby increasing the efficiency and robustness of the optimization.

Lastly, a policy optimization framework relies heavily on the design of the objective function. The case studies in this dissertation applied just one of the three alternative social choice functions presented in chapter 3. More work is needed to continue developing an equitable approach for defining and operationalizing the notion of social optimality with respect to both equity and sustainability. The level of service KPIs implemented in BISTRO, including measures of public transit vehicle crowding, transportation cost burden, and accessibility are a first step in this regard, and the research I have conducted thus far on the impacts of weighted and Pareto-based optimization have illuminated the challenges in automating public policy decision-making.

For Elena and Sara

Contents

List of Figures

List of Tables

Acknowledgments

The research presented in this dissertation is the result of the collaboration, mentorship, and support of the communities of scholars, professionals, friends and family that I have the immeasurable pleasure of being a part of.

I am forever grateful to my Ph.D. advisors, Professors Alexandre Bayen and Susan Shaheen, for their mentorship and collaboration throughout my graduate studies. Their leadership, knowledge, compassion, and commitment to making a meaningful impact on the future of transportation have inspired me and guided me to become the researcher and person that I am today.

Many thanks are also due to Professors Daniel Rodriguez and Joan Walker for their invaluable feedback, including thoughtful comments that guided me in making substantive improvements to this dissertation. I am also grateful to both Professors Walker and Rodriguez for their support and mentorship of the student-led graduate seminars on transportation economics and transportation equity that I had the pleasure of participating in, all of which had a profound impact on my dissertation research.

The development of BISTRO was a collaborative effort, to which I had the great honor of contributing alongside a multidisciplinary team including Professor Bayen, Sidney Feygin, Teddy Forscher, and Valentine Golfier-Vetterli at UC Berkeley, Colin Sheppard and Rashid Waraich at the Lawrence Berkeley National Laboratory (LBNL), and Abhishek Gupta and Jonny Lee at Uber, Inc. Many thanks are also due to the 400 participants of the 2019 Uber ML Hackathon, particularly those who provided feedback throughout the competition and those who participated in the post-competition interviews whose input contributed immensely to the improvement of BISTRO and the cultivation of lessons learned from the pilot study.

The ongoing development and application of BISTRO for the study of congestion pricing optimization was made possible by the collaborative effort of the talented students I have had the great pleasure of mentoring over the course of my Ph.D. This includes two teams of U.C. Berkeley Masters of Engineering students whose capstone projects contributed significantly to the development and execution of the congestion pricing optimization methodology applied in this dissertation as well as the other student researchers that supported this research. Thank you to Léo Toulet, Zangnan Yu, Anyi Chen, and Xinyi Tang for your work on the Sioux Faux case study; Haochong Xia, Christina Liu, and Albert Loekman for your work piloting the San Francisco Bay Area case study; Mackena Schwinn for your contributions to the Sioux Faux case study and research on Pareto-based optimization approaches for BISTRO; Jarvis Yuan for your contributions to the BISTRO starter kit, calibration of the SFSim model, and collaboration in designing the San Francisco Bay Area case study; and Timothé Kasriel for your contributions to the visualizations in the Sioux Faux case study. I am also grateful to the members of the BEAM team at LBNL that have provided technical support for the integration of BEAM and BISTRO, including Colin Sheppard, Rashid Waraich, Zachary Needell, Nikolay Lin, Artavazd Balaian, and Xuan Jiang.

The San Francisco Bay Area congestion pricing optimization case study was the culmination of several research studies, including collaborations with Juan Caecedo, Jeffrey Greenblatt, and Gordon Bauer to whom I am grateful for their contributions. I am also thankful to the members of the Transportation Sustainability Research Center and Bayen Lab that have helped to pilot surveys and have provided feedback on my research over the years.

The opportunity to pursue this research was made possible through generous funding from the Dwight David Eisenhower Transportation Fellowship Program, the Transportation Sustainability Research Center, and the University of California Institute of Transportation Studies.

To my colleagues in the UC Berkeley Institute of Transportation Studies that have provided invaluable support and camaraderie over the course of my graduate studies, thank you. To my writing partner, Doctor Harmon, who accompanied me through the final stretch of writing this dissertation, I am forever grateful. Finally, thank you to my wonderful friends and family for your love and support.

Chapter 1

Introduction

1.1 Motivation

Transportation systems across the globe are increasingly failing to provide environmentally, economically, and socially sustainable mobility. Road transportation alone accounts for about 18% of global CO_2 emissions each year [1]. Although efforts to improve fuel efficiency and reduce emissions have been accelerating in recent years, the transition from internal combustion engine (ICE) vehicles to electric vehicles (EVs) will not necessarily reduce the number of miles driven nor the number of vehicles produced, both of which are key determinants of energy consumption and traffic congestion. Population growth and increasing suburbanization have driven the growth of vehicle miles traveled (VMT) per capita across the United States by about three percent from 2009 to 2019, amounting to a ten percent increase in total VMT [2]. As of 2019, the average American spent about 55 minutes commuting per day, amounting to an estimated 99 hours lost to congestion over the year per person [3, 4]. Lastly, significant disparities in the affordability, accessibility, and convenience of transportation are persistent, creating inequities in the distribution of both the positive and negative impacts of transportation systems. Even though tremendous technological development in the transportation field has created opportunities to advance towards many of the most ambitious sustainability goals that have been set forth for decades, there is still much to be desired in the public's ability to steer the course of that development in a sustainable direction.

This dissertation develops a holistic approach for transportation system design and demand management to guide the development of data-driven strategies involving numerous policy interventions that systematically improve the environmental, economic, and social sustainability of transportation systems. The aim of this research is to advance data-driven policy-making by leveraging simulation-based optimization to investigate the trade-offs across conflicting policy objectives and their impacts on the design and resulting outcomes of multi-faceted sustainable transportation strategies. In particular, I focus on the optimization of congestion pricing and targeted incentive schemes in multi-modal transporta-

tion networks including pooled ride options. Congestion pricing aims to reduce congestion by charging road users for driving on congested roads. When optimized purely for the objective of congestion reduction, pricing produces regressive financial impacts. However, congestion pricing also has the potential to reduce transportation inequities by raising revenue that may be distributed to individuals (i.e., as rebates or incentives) or strategically re-invested in the transportation system (e.g., to increase public transit service quality or coverage). In the following sections of this chapter, I provide a background of the emerging trends affecting transportation sustainability and the role of pricing. A brief overview of the gaps in the congestion pricing optimization literature is also provided, followed by a summary of the theoretical, methodological, and empirical contributions made in this dissertation and a roadmap of the organization of the following chapters.

1.2 Emerging trends

The digitization and automation of transportation systems is fundamentally transforming both the supply and demand of personal transportation. Transportation services are increasingly automated and centrally operated with unprecedented access to information delivered via the internet of things (i.e., by mobile phones and other GPS-enabled devices as well as other sensing technology). In particular, shared on-demand mobility services, including shared micromobility (e.g., bikesharing, scootersharing) and ridesourcing (also known as transportation network company (TNC)) services, have emerged as popular, mostly privately operated, innovative mobility service options with the potential to reduce VMT and greenhouse gas (GHG) emissions by complementing traditional public transit services and reducing personal auto ownership. Their growing adoption reflects important changes in travel behavior, including an apparent willingness to pay for the relative convenience, reliability, and other intangible factors that on-demand mobility services offer. Moreover, travelers - particularly those in dense urban areas - have access to more travel options than ever before and are able to trade off the levels of service and costs of each option in real-time, enabling dynamic, highly flexible decision-making. In return, service operators are able to leverage user- and system-level data to dynamically optimize fleet management and pricing.

However, unlike public transit and traffic operations, private transportation services are not responsible for optimizing for social good. For example, the rapid deployment of TNC services (e.g., Lyft, Uber) has contributed significantly to increases in congestion in several large metropolitan areas of the U.S. due to increases in both the number of vehicle trips and VMT per trip [5, 6, 7, 8]. The total VMT produced by TNCs includes the miles driven en route to drivers' markets of choice, as well as those driven while roaming and unreserved, driving to pickup a passenger, and driving with a passenger in tow. The former three phases of service represent 'deadheading,' or miles driven without a passenger in the vehicle. An estimated 20 to 45 percent of miles driven by TNC vehicles are accounted for by deadheading while awaiting a ride request and driving to the passenger pickup point [6, 8, 9, 10, 11].

Although the automation and electrification of transportation systems is expected to

bring about reductions in VMT and emissions from road use, the possibilities of further increases in demand for auto travel and continuing inefficiency in road use caused by zero- and single-occupant travel pose significant challenges for realizing the greater benefits from these strategies. Simulation-based studies of the potential impacts of shared, automated vehicle (SAV) fleets project that adoption of SAVs in favor of personally owned vehicles can reduce VMT and increase vehicle throughput in urban areas [12, 13]. Moreover, the electrification of such fleets is projected to bring about significant reductions in energy consumption and environmental emissions from road use [14, 15, 16]. However, it is important to note that these projected benefits depend heavily on the SAV adoption rate as well as the occupancy rates of the vehicles [14, 15, 12, 16, 13].

Increased adoption of *pooling*, in which multiple travelers with similar trips share a ride in the same vehicle, can mitigate the increased VMT due to deadheading and induced demand by increasing the average occupancy of vehicles on the road network [17, 18]. The emergence of on-demand ride services has spurred the adoption of app-based pooling services and created the potential to drastically expand the market for pooling. Although pooling, which includes carpooling and vanpooling, is not a new concept, app-based pooling has broadened the primarily commuter-based market for pooling to a more diverse set of users and use cases. Moreover, in comparison to traditional carpooling/vanpooling in which drivers and passengers are primarily matched by employer-based programs and informal pickup/dropoff locations, app-based pooling enables travelers to consider pooling on-the-fly with minimal effort needed to coordinate the logistics of sharing a ride. Prior to the COVID-19 pandemic, numerous app-based pooling services were in operation across the U.S., including pooled TNC services (e.g., UberPool, Lyft Shares Rides), app-based carpooling/vanpooling (e.g., Waze Carpool, Scoop), and microtransit services (e.g., Via). The growing adoption of these services presented a unique opportunity to limit congestion, energy use, and emissions through increased vehicle occupancy.

However, the success of pooled ride services is largely dependent on the decisions of individual travelers to pool or not to pool, which are heavily influenced by the perceived differences in the levels of service offered by pooling versus riding or driving alone [19]. The COVID-19 pandemic drastically impacted travel behavior with immediate and potential long-term effects on the willingness of travelers to share a ride. The trends in the reduction of public transit ridership during the pandemic vary by socio-economic status, with ridership decreasing the least among lower-income captive riders and essential workers and the most among higher-income, higher-educated populations [20, 21]. In addition, many pooled TNC services were suspended during the pandemic and have not been re-introduced in all of the regions where they previously operated. Yet, even prior to the COVID-19 pandemic, the rates of pooled ride requests among users of the TNC services Lyft and Uber were relatively low, resulting in negligible impacts to overall vehicle occupancy [9, 7, 17]. In a survey distributed across four California metropolitan regions in 2018, [19] found that only about 30 percent of TNC users consider requesting a pooled ride more than half the time they use TNCs. In 2018 in New York City, only about 22 percent of requested Lyft Line (now Lyft Shared) rides and 23 percent of Uber Pool rides resulted in matched trips [7].

1.3 The role of pricing

Some jurisdictions have addressed the negative impacts of TNCs by imposing surcharges on rides while providing discounts for pooled TNC trips that can help reduce VMT. Examples include the: 1) New York State Congestion Surcharge, which applies a $2.75 fee to all ride-alone TNC trips but only $0.75 to all pooled TNC trips that start, end, or pass through Manhattan south of 96th Street; 2) San Francisco Rideshare Tax, which applies a 3.25 percent surcharge to all ride-alone trips but a 1.5 percent surcharge to pooled TNC trips that start in San Francisco; and 3) City of Chicago congestion pricing, which applies a $3 surcharge to all ride-alone TNC trips but a $1.25 surcharge to pooled trips that start or end in a designated downtown zone during weekday peak hours (between 6 AM and 10 PM) and applies a $1.25 surcharge on all other ride-alone trips but a $0.65 surcharge to pooled TNC trips. The disposition of funds from state and local TNC taxes and fees includes general funds, 'congestion mitigation' funds, and even public school funds. However, it remains to be seen whether these pricing policies are effective in curbing congestion and whether their effects are distributed equitably across the population.

In several regions across the U.S., a broader pricing strategy for congestion mitigation is under development: congestion pricing. Congestion pricing aims to reduce congestion during peak travel periods by creating a monetary incentive for road users to shift their preferred time or mode of travel [22]. This strategy is deeply rooted in economic theory which establishes that congestion pricing can reduce the inefficiencies of selfish travel behavior by charging road users for the externalities that they impose on the system (and other drivers) [22, 23]. When applied to *high-occupancy toll* (HOT) lanes (also known as *express* or *managed* lanes), congestion pricing is typically dynamic, with a distance-based rate that is optimized to maintain a desired level of service (e.g., free-flow speed) on the tolled lane. Zonal congestion pricing, including *cordon-* and *area-based* pricing schemes, charges vehicles for entry into or movement within a designated area, respectively. Several successful zonal congestion pricing policies have been in place across the globe for years, including in Singapore, London, United Kingdom, and Stockholm, Sweden. These policies have brought about significant reductions in road traffic and increased public transit ridership [24, 25, 26]. In the U.S., there are about 53 express and HOT lane facilities, amounting to about 1,858 lane-miles across 12 states [27]. In addition, the cities of New York, San Francisco, and Los Angeles are in various phases of the development of zonal congestion pricing policies, with New York being the first North American city to codify the strategy in law. However, concerns about the equity implications of pricing road users out of driving pose significant challenges for the public acceptance and political feasibility of congestion pricing.

By design, pricing-based transportation demand management (TDM) strategies leverage the heterogeneity in the sensitivity of demand to travel time and cost to elicit a desirable shift in behavior. The most price-sensitive travelers are likely to be the most effected by the increased costs. When possible, they will shift their travel behavior to use other modes (e.g., carpool, public transit, bike or walk) or to travel during off-peak times (i.e., before/after the times at which charges apply). In the event that alternatives are unavailable or inacces-

sible, they will have to either forego travel altogether or pay the increased cost and likely bear a greater cost burden for doing so compared to other travelers [28]. Thus, congestion pricing poses a considerable risk for worsening disparities in mobility and accessibility across groups of travelers with varying geographic and demographic characteristics. Underserved communities, including low-income residents and people of color (POC) living in areas with increasing costs of living and poor access to public or active transportation, are among the most car-dependent travelers. As such, they are the most likely to have to make a mode shift to an already undesirable alternative or to take on the additional costs of congestion pricing in order to maintain the level of mobility they rely on for important trip purposes such as commuting, doing groceries, and going to medical appointments. This is exemplified by the findings of [19], which showed that low-income, POC, and non-vehicle owners were among the most frequent TNC users across four metropolitan regions in California regardless of the greater cost of traveling by TNC compared to public transit.

Zonal congestion pricing strategies are typically developed as part of the long-range transportation planning process in which regional travel models are used to forecast the impacts of the policy under a set of potential future scenarios. Agent-based simulations (ABS) using activity-based travel models, which simulate daily travel decisions (e.g., departure time, mode, and routing choices) at an individual person- or vehicle-level are increasingly being adopted for this purpose due their ability to represent the network effects of system interventions such as congestion pricing at a granular level [29]. ABS is sensitive to the impacts of system-level interventions to individual-level decisions, enabling the analyses of detailed forecasts of the behavioral and operational outcomes of each strategy option. However, the time and resource intensity of simulating thousands of individual agents constrains the set of policy design options that can be fully analyzed by public agencies. Only a discrete subset of options are typically simulated under each scenario considered, resulting in point estimates of relative performance metrics. Input from the public and other stakeholders may further complicate the interpretation and application of these results, which are likely to reflect conflicts between various policy objectives.

1.4 Gaps in the literature

The optimization of road pricing schemes and congestion pricing in particular has been studied widely in the transportation literature. The vast majority of the congestion pricing optimization literature employs the objective of social welfare maximization, typically defined as the total toll revenue minus the sum of total user costs (e.g., travel time and/or cost) and social costs (e.g., operational costs). While many studies have examined the distributional equity implications of this efficiency-focused approach, few have incorporated notions of equity into the congestion pricing optimization problem. Moreover, the inclusion of multiple competing objectives in congestion pricing optimization is limited to congestion, emissions, and toll revenue objectives, with no studies incorporating more than two trade-offs in the optimization objective at a time.

In addition, there is a lack of consideration for competition across multiple modes of transportation, with most studies focusing solely on private auto users, some incorporating mode choices between auto and public transit, and few considering the role of carpooling in congestion pricing optimization. Although on-demand ride services have been in the market for almost a decade, congestion pricing optimization studies have not yet begun to reflect the impact of these dynamic ride services or app-based pooling options on congestion pricing optimization from a public perspective.

1.5 Contributions

This dissertation aims to advance the state of the art in transportation policy design and analysis by developing a scalable mechanism design-based approach for formulating, optimizing, and analyzing transportation pricing policies. I develop a framework for the multi-objective optimization of multi-faceted TDM strategies with a focus on congestion pricing and incentives for reducing congestion and environmental emissions while producing net public revenues and reducing the disparities in the cost burden of transportation. Using a game theoretic approach, I first establish a theoretical foundation for the optimization of road pricing and incentive schemes in a multi-modal network with heterogeneous road users by formulating a bi-level optimization problem for the minimization of equity-based social objectives. I establish the existence of an optimal road pricing strategy with mode- and link-specific tolls and mode- and individual-specific incentives with respect to a social objective function of the estimated travel costs of each user group in the network (defined by their origin, destination, and individual characteristics) and propose various objective functions representing different theories of justice, including Utilitarian, Egalitarian, Rawlsian, and Equality of Opportunity. Each of these alternative objective functions have implications on the distributional impacts of road pricing schemes.

Next, I develop a methodological approach for regional-scale optimization of multi-faceted TDM strategies within the context of on-demand ride services. I co-developed the Berkeley Integrated System for TRansportation Optimization (BISTRO), an open-source transportation planning decision support system that applies a simulation-based optimization framework to leverage the power of ABS, activity-based travel models, and machine learning to assist stakeholders in designing optimal system-level TDM strategies [30]. I conducted two case studies of congestion pricing optimization using BISTRO. The first is focused on developing the methodology for cordon-based congestion pricing optimization as well as investigating the impacts of policy and objective function design on optimization outcomes in a benchmark scenario called Sioux Faux. Two pricing schemes are optimized and compared: 1) a simple cordon toll charging drivers and TNCs upon entry/exit to the cordon, and 2) a cordoned mileage fee that charges drivers and TNCs per mile driven within a circular cordon. The methodology is refined in the second case study, which implements a Pareto-based optimization of a cordoned mileage fee with multi-modal incentives in a realistic regional model of the San Francisco Bay Area.

1.6 Organization

The following chapter presents background on the academic literature related to road pricing optimization. Next, chapter 3 formulates a mechanism design problem for optimal road pricing and incentive schemes in multi-modal transportation networks with heterogeneous users. Chapter 4 presents BISTRO, including the design and software implementation details as well as the results and key takeaways from a pilot machine learning competition using BISTRO that was conducted among employees of Uber, Inc in 2019. Next, chapters 5 and 6 present the two case studies of congestion pricing optimization, including the development of the methodological approach, analysis of results, and discussion of the implications of the findings for further research and policy development. Finally, chapter 7 summarizes the key contributions of this dissertation and discusses the key takeaways for policy recommendations and future research directions.

Chapter 2

Transportation Pricing Optimization

Road pricing charges road users for access to a particular road or area within a road network. Road pricing that is devised to elicit a reduction in the volume of vehicles in congested areas and/or time periods, known as *congestion pricing*, can take various forms. When applied to *high-occupancy toll* (HOT) lanes (also known as *express* or *managed* lanes), congestion pricing is typically dynamic, with a distance-based rate that is optimized to maintain a desired level of service (e.g., a target speed) on the tolled lane. Zonal congestion pricing, including *cordon-* and *area-based* pricing schemes, charges vehicles for entry into or movement within a designated area, respectively. Area-based pricing includes both flat-rate (e.g., the daily London Congestion Charge) and *distance-based* pricing schemes (e.g., Singapore's Electronic Road Pricing 2.0).

The optimization of road pricing schemes has been studied widely in the transportation literature. Historically, the optimization of road tolls has been studied using analytical models of vehicle dynamics on a single road or in a simplified road network, with an emphasis on the application of the theory of *marginal-cost pricing* wherein road users are charged according to the marginal social cost of their individual consumption of a resource (i.e., a road) [23, 22]. The literature focuses on determining the optimal rate to charge under various assumptions regarding the behavioral and physical dynamics of the transportation system. Under assumptions that a *first-best* price equal to the short-run marginal cost of road use can be achieved, the resulting marginal costs and marginal benefits of the pricing scheme are equal, which is considered a *Pareto efficient* outcome [31, 32]. Given the unrealistic nature of such assumptions (e.g., zero implementation costs, ubiquity of charges, perfect information), most of the contemporary literature on road pricing optimization focuses on *second-best* pricing schemes in which market distortions are represented by constraints imposed on the optimization problem [32, 33]. Increasing in complexity from the standard model of a single congestible road, the study of second-best pricing schemes includes, but is not limited to, the optimization of road pricing in the context untolled alternatives (e.g., managed lanes), *multi-commodity* networks (i.e., with demand across multiple origin-destination (OD) pairs), fully or partially private information about the heterogeneity of road user preferences, and competition with other travel modes. The following sections provide an overview of the

various aspects of the congestion pricing optimization problem as studied in the literature, including the specification and analysis of various dimensions of demand sensitivity, pricing structure and cordon design, optimization objectives, optimization approaches, and equity analysis.

2.1 Demand sensitivity

The literature on congestion pricing optimization initially focused on *single commodity* networks (e.g., a single origin and destination) with fixed demand. The *latency*, or cost, of alternative routes determine the distribution of demand, typically according to assumptions of *utility-maximizing* (i.e., cost-minimizing) decision-making. In the simplest cases, travel time and cost are traded off using a *value of time (VoT)* parameter which is assumed to be uniform across the population. However, several studies investigating the role of the heterogeneity of travel demand in optimal tolling demonstrate that ignoring heterogeneity in VoT can result in drastic underestimation of the welfare benefits of second-best tolling policies [34, 35, 36, 37]. In addition to being a key determinant of the equilibrium distribution of routing choices, heterogeneity is essential for determining the elasticity of demand with respect to travel time (e.g., departure time, arrival time), mode (e.g., public transit, walking, biking, choosing not to travel), and even longer term decisions such as home or work location choices that may also be sensitive to the distribution of costs in a transportation network [35, 38, 39, 40, 41, 42, 43].

Small (2004) examines the impacts of increased travel speeds and pubic transit ridership on public transit service levels (number of routes and service frequency), operator costs, and user costs optimized from the perspective of the public transit operator, demonstrating that congestion pricing can induce a "virtuous cycle" of mode shifts, service level improvements, and cost reductions for both users and operators that can result in considerably greater improvements to consumer surplus than is projected by models of the impacts to private vehicles alone [43]. Although the model developed by [43] incorporates changes in speed and public transit ridership as exogenously determined variables, several other studies that explicitly model the choices between multiple modes support these findings [35, 38, 40, 44]. Glazer and Niskanen (2021) and Liu et al. (2009) study revenue-neutral pricing schemes in which revenue from a toll applied to cars in a network with public transit alternatives is entirely redistributed as subsidies for public transit fares [41, 35]. They concluded that any revenue-neutral pricing scheme of this type is Pareto-improving as long as it minimizes total system travel time. Similarly, [40] analyzed consumer surplus using an equilibrium model with road tolls, transit fares, and transit scheduling all optimized by a central agency, concluding that gains in consumer surplus are highly dependent on the portion of travel demand that is dependent on public transit as well as the potential impact of the congestion pricing scheme to significantly reduce public transit costs through reduced congestion. Armelius (2005) demonstrates the importance of incorporating departure time sensitivity in such contexts in order to reflect the relative impacts of pricing schemes on shifts in preferred

travel times and/or modes in response to the changes in the travel times and costs of each alternative [38]. Moreover, the consideration of the sensitivity and heterogeneity of demand is essential for the evaluation of the *equity* implications of optimal tolling schemes [45, 46, 47, 48]

Traditional carpooling, as incentivized by high occupancy vehicle (HOV) and high occupancy toll (HOT) lanes has been examined in the congestion pricing optimization literature, typically in comparison to ubiquitous first-best tolling schemes. Yang and Huang (1999) demonstrate that, in the absence of HOV lanes, price differentiation between single-occupant vehicles (SOV) and HOVs is necessary to achieve a first-best tolling scheme [49]. In an analysis of various optimal pricing schemes, [46] find that the introduction of HOV lanes improves efficiency by encouraging carpooling, which significantly reduces travel times on the managed lanes. Moreover, they show that most of the gains of first-best pricing (tolling all lanes on a highway) can be achieved by offering a carpool exemption for tolls, which lessens some of the undesirable distributional effects of ubiquitous tolling. However, they find that the magnitude of the potential benefits from carpool exemptions - including discounts for carpools based on the number of passengers (e.g., HOV3+) - depend heavily on the initial share of carpooling in the study area. Observational studies have found that HOV to HOT lane conversions have had negligible and even slight negative effects on carpooling and that the share of HOVs on a facility is generally insensitive to travel time savings [50, 51]. More recently, the adoption of transportation network companies (TNCs) and the development of automated vehicle (AV) technology has spurred a renewed interest in the potential impacts of pooling rates on overall transportation system efficiency. Ostrovsky and Schwarz (2019) postulates that the provision of pooled on-demand ride services by shared AV (SAV) fleets in the presence of ubiquitous tolling will produce the optimal conditions to achieve the economies of scale in pooling demand necessary to support high match rates [18]. However, it remains to be seen whether pricing mechanisms and automation alone will engender sufficient demand for pooling, particularly given several key socio-economic issues including the willingness to share, curb management, social equity, personal safety, and labor considerations [52].

2.2 Pricing structure and charging zone design

The optimization of congestion pricing schemes may also entail the design of the charging zone itself. Network-scale studies predominantly formulate link-based tolling schemes. While it is possible that the optimization of link-based tolls may result in schemes with discernible tolled and untolled areas, it is unlikely that such a result would be produced without the application of explicit constraints on the optimization problem. Examination of spatially and/or geometrically constrained tolling schemes is valuable for understanding the trade-offs in infrastructure investment decisions for pricing schemes that are more technically and practically feasible than unconstrained link-based schemes [42, 44, 53].

In comparison to traditional cordon tolls in which drivers are charged upon entry to

a charging zone, area-based, multi-cordon, and distance-based schemes have been found to produce greater social welfare gains by increasing the coverage of the tolling zone and achieving a greater correlation with the marginal congestion contributed by each tolled vehicle [39, 42, 45, 54]. Generally, greater toll levels are found to produce greater reductions in driving which lead to greater social benefits [42], as does greater coverage of the charging zone [**maruyama˙efficiency˙2007** , 37, 42]. Distance-based and time-varying schemes are also found to outperform simple cordon tolls with respect to social welfare [39, 54, 55]. However, greater complexity and charge rates generally lead to lower public acceptance rates [42]. May et al. (2002) emphasize simplicity and ease of understanding as qualitative design objectives for politically feasible pricing schemes in addition to aiming for a charge level that is acceptable and perceived as fair to the public [42].

2.3 Optimization objectives

The design of optimal tolls is inherently dependent on the definition of objectives used in the optimization. A vast majority of the literature defines optimality with respect to social welfare, typically defined as the total user benefit (i.e., the total generalized travel cost) minus the total social cost (i.e., the total travel time plus the total toll revenue) [34, 36, 45, 46, 56]. Mayet and Hansen (2021) demonstrate the significance of the definition of the social welfare function itself, with a social welfare function defined in monetary terms resulting in higher tolls (due to assumptions about the heterogeneity in marginal utility of income) while a function defined in time units results in lower tolls (due to assumptions on uniform marginal utilities of time) [36].

Few studies consider other objectives for toll optimization. Pricing optimization on managed lanes tend to consider throughput and/or revenue maximization as primary objectives, which follow from assumptions of the purpose of such facilities [37, 57]. The correlation of efficiency objectives with environmental objectives depend strongly on the modeling assumptions made. Some studies find that minimizing congestion results in emissions reductions due to the reduction in vehicle miles traveled achieved by mode shifts [42, 58]. Yet other studies conclude that the two objectives are at odds due to the relationship of travel speeds with optimal emission rates of road travel, though optimal tolling schemes with respect to specified trade-offs between congestion and emissions are feasible [59].

2.4 Optimization approach

Low-dimensional models of pricing optimization are often employed for the purposes of simplicity and interpretability. Such models can be solved numerically using mathematical approaches. However, when modeling higher dimensional networks, computing the optimal toll and traffic assignments using numerical methods becomes infeasible. A multitude of methods are used to overcome the challenge of simultaneously optimizing road user choices

and toll optimization in order to compute an equilibrium, including heuristic algorithms [37, 47, 53], genetic algorithms [60, 61], and macroscopic fundamental diagram (MFD) models [39, 62].

A few studies have employed agent-based simulation (ABS) engines for this purpose, which explicitly model the decisions and movements of individual road users throughout discrete and/or continuous time horizons. May and Milne (2000) demonstrated the importance of considering rerouting effects in congestion pricing design by utilizing a static traffic assignment and simulation framework [54]. De Palma, et al. (2005) used the METROPOLIS simulator, which incorporated departure time, mode, and route choice models with uniform VoTs [55]. A combination of heuristics and response surfaces were utilized to iteratively update and simulate the tolling schemes until converging to optimal solutions for six variations of link-tolling schemes. Some studies have used surrogate-based techniques to optimize tolls using dynamic traffic assignment [63, 64]. These techniques approach the optimization as a black-box problem, in which the outcomes of the simulation are considered stochastic and a-priori unknown.

2.5 Equity analysis

Many studies have assessed the equity of outcomes from optimal congestion pricing schemes, both in theory and in practice. Distributional analyses of the travel times and/or tolls paid by heterogeneous road users under an optimized scheme find that lower income road users are generally worse off than higher income users as a result of lower and higher values of time, respectively [37, 45, 48, 56]. Verhoef and Small (2004) assess the distributions of social welfare effects of various optimal tolling schemes across a population with a continuous distribution of VOT, finding that the average loss in consumer surplus (measured as the change in generalized travel cost compared to a no toll scenario) is smallest for drivers with the highest VOT and greatest for those with the lowest VOT [37].

Santos and Rojey (2004) examine the role of geographic differences in the spatial distribution of heterogeneous demand with respect to the placement of a cordon toll, finding that the toll can be progressive even without revenue distribution, particularly if higher income groups are the most likely to be subject to the toll [56]. Maruyama and Sumalee (2007) find that the Gini coefficient of consumer surplus grouped by trip chains and ODs increases (worsens) slightly with respect to the total consumer surplus produced by optimal tolls (which was found to be positively correlated with the coverage and toll level of the cordon) [45].

Yang and Zhang (2002) propose spatial and social equity constraints on an optimal second-best tolling scheme that bound the changes in generalized travel costs by those of the no-toll and first-best tolling equilibria with respect to OD pairs and user classes [48]. The result is to constrain the relative percentage changes in generalized OD travel costs across OD pairs and income groups according to the specified level of inequity determined by the constraints. However, across all examples shown, the optimal tolls still produce the most

negative effects on the lowest income groups while high income groups are impacted the least, and in some cases, receive benefits [48].

2.6 Gap analysis

The vast majority of the congestion pricing optimization literature employs the objective of social welfare maximization, typically defined as the total toll revenue minus the sum of total user costs and social costs. While many studies have examined the distributional equity implications of this efficiency-focused approach, few have incorporated notions of equity into the congestion pricing optimization problem or redefined the primary objective of the problem in these terms. Moreover, the inclusion of multiple competing objectives in congestion pricing optimization is limited to congestion, emissions, and toll revenue objectives, with no studies incorporating more than two trade-offs in the optimization objective at a time.

In addition, there is a lack of consideration for induced travel demand nor competition across multiple modes of transportation, with most studies focusing solely on private auto users, some incorporating mode choices between auto and public transit, and very few considering the role of carpooling in pricing optimization. Although on-demand ride services have now been in the market for almost a decade, congestion pricing optimization studies have not yet begun to reflect the impact of these dynamic ride services or app-based pooling options on congestion pricing optimization from a public perspective.

Chapter 3

Mechanism Design for Optimal Transportation Pricing Policies

3.1 Introduction

Road pricing policies, including congestion pricing and mileage-based road user fees, are gaining momentum as potential strategies to mitigate growing congestion and the need for more sustainable sources of transportation revenue. However, when optimized purely for efficiency in terms of travel times, costs, and/or toll revenues, road pricing can produce regressive impacts, with the lowest income road users taking on the greatest cost burden of the strategy. In order to address these inequities and their ensuing negative impact on the public perception and acceptance of road pricing policies, the redistribution of revenue generated by pricing schemes as financial incentives or subsidies for targeted groups is increasingly being considered.

As the equity implications of congestion pricing gain priority in the policy-making process, mechanisms for optimal congestion pricing that include a strategy for the distribution of revenue are needed. Moreover, mechanisms that optimize pricing schemes with respect to equity-focused objectives are needed, particularly in light of the primarily utilitarian approach that has been applied in the transportation literature. Utilitarianism emphasizes that the maximization of social welfare, as measured by the total individual "utility" of a particular action or policy, achieves justice regardless of the differences in the preferences or circumstances across the population. This approach, as applied in the optimization of congestion pricing, leads to inequities in the distribution of impacts of the pricing scheme.

This chapter formulates a bi-level optimization problem for socially optimal toll and incentive schemes in multimodal multi-commodity networks with heterogeneous users. This work extends that of [65] with the inclusion of mode- and individual-specific incentives and a budget-balancing condition in the optimal tolling scheme and establishes the existence of tolls that enforce a socially optimal multimodal traffic assignment with mode- and individual-specific incentives. In addition, I formulate four different social choice functions for which

such a scheme may be optimized, each representing different theories of justice: utilitarian, egalitarian, equality of opportunity, and Rawlsian.

3.2 Related Literature

Congestion pricing optimization

The study of optimal congestion pricing stems from public economic theory regarding marginal cost pricing of public goods, whereby consumers are charged according to the marginal social cost of their individual consumption of a resources (i.e., a road) [23, 22]. The optimization of congestion pricing in network systems spans various disciplines. It has been studied in the context of internet, electricity, and transportation systems, among others. In these settings, users of capacity-constrained networks seek to find an optimal path between a particular origin-destination (OD) pair with respect to their individual preferences or objectives. The resulting *user-optimal* routing scheme is considered a *Nash* or *Wardrop* equilibrium if no individual has any incentive to deviate from their chosen path. Numerous works have demonstrated the efficiency losses resulting from such *selfish routing*, as measured by the differences in aggregate system-level objectives (i.e., various measures of social welfare) produced by user-optimal versus *system-optimal* routing. Optimal tolling schemes generally seek to elicit system-optimal routing at equilibrium by increasing the individual perceived costs of network usage [65, 66, 67].

In congestion pricing optimization, a non-decreasing function of the latency (e.g., travel time) and cost of each link represents individual trade offs between money and time, typically using a parameter called the *value of time* (VoT) that converts latency from units of time to money. The importance of representing the heterogeneity in user preferences in the optimization and analysis of congestion pricing schemes has been underscored by numerous works [36, 37, 67, 68, 69]. The existence of link-based tolls for heterogeneous users in a congested network was proven by [65] and [66]; the latter paper also establishes existence in networks with multiple commodities (i.e., multiple OD pairs). Both of these works assert the existence of optimal tolls with respect to a secondary objective, such as minimizing total tolls paid by network users or the weighted sum of latencies. In the context of mechanism design, this secondary objective is referred to as the *social objective*.

The most prominent social objectives for optimal tolling schemes include the total travel time and total *generalized travel cost* - the sum over all users of the product of the VoT and travel time for each user plus the travel cost [36, 37, 48, 68, 69]. Depending on assumptions on the redistribution of toll revenues, the total tolls paid may also be included in the social objective as a social benefit [36, 37, 70]. Other social objectives have included measures of the environmental impacts of road traffic (e.g., greenhouse gas (GHG) emissions [59, 71], energy consumption [72], and measures of fairness and equity (e.g., pareto optimality, Gini coefficient, minimax individual costs) [73, 74].

Transportation equity and congestion pricing

Many studies have assessed the equity of outcomes from optimal congestion pricing schemes, both in theory and in practice. Distributional analyses of the travel times and/or tolls paid by heterogeneous road users under an optimized tolling scheme find that lower income road users are generally worse off than higher income users as a result of lower and higher VoTs, respectively [37, 45, 48, 56]. In a simulation-based study of several cities in the United Kingdom, [56] demonstrates that the regressivity of congestion pricing may vary by geography, depending on the spatial and temporal distributions of travel demand with varying socio-economic characteristics. When low-income road users are reliant on driving within a charging zone or negatively impacted by the financial pressure to shift their travel behavior, congestion pricing is likely to be regressive [56].

Congestion pricing can also have secondary short-run effects on the distribution of demand across time and alternative modes of transportation, which add complexity to the assessment of the corresponding equity implications. Long-run effects, such as the impact of congestion pricing on vehicle ownership rates and activity generation, may also be significant, yet are generally considered outside the scope of tolling optimization problems. Models of multi-modal transportation networks that include alternatives to driving (e.g., public transit modes, carpooling, on-demand ride services) enable the investigation of modal shifts as a result of tolling, which can have significant impacts on secondary objectives such as total vehicle miles traveled (VMT), environmental emissions, public transit ridership and revenues. Small (2004) examines the impacts of increased travel speeds and pubic transit ridership on public transit service levels (i.e., number of routes and service frequency), operator costs, and user costs optimized from the perspective of the public transit operator, demonstrating that congestion pricing can induce a "virtuous cycle" of mode shifts, service level improvements, and cost reductions for both users and operators that can result in considerably greater improvements to consumer surplus than is projected by models of the impacts to private vehicles alone [43]. Although the model developed by [43] incorporates changes in speed and public transit ridership as exogenously determined variables, several other studies that explicitly model the choices between multiple modes support these findings [35, 38, 40, 44]. Glazer and Niskanen (2021) find that a policy involving a toll on car travel only will unambiguously hurt the users of the public transport system when both modes are subject to congestion and there is no redistribution of toll revenues [41].

Various revenue redistribution schemes have been investigated with the aim of addressing the disparate impacts of tolling across road users [35, 40, 41, 44, 74, 75, 76]. Jalota, et al. (2021) prove the existence of congestion pricing and revenue refunding (CPRR) schemes that reduce total system efficiency (measured as the sum of VoT-weighted travel times), while not increasing wealth inequality [74]. Glazer and Niskanen (2021) and Liu, et al. (2009) study a revenue-neutral pricing scheme in which revenue from a toll applied to cars in a network with public transit alternatives is entirely redistributed as incentives for public transit (i.e., fare subsidies) [35, 41]. They concluded that any revenue-neutral pricing scheme of this type is Pareto-improving as long as it minimizes total system travel time. Similarly, Ferrari

(2005) analyzed consumer surplus using an equilibrium model with road tolls, transit fares, and transit scheduling all optimized by a central agency, concluding that gains in consumer surplus are highly dependent on the portion of travel demand that is dependent on public transit as well as the potential of the congestion pricing scheme to significantly reduce public transit costs through reduced congestion [40].

Bi-level optimization

This study presents a bi-level program to optimize link- and mode-specific tolls with mode- and individual-specific incentives in a multi-modal transportation network with heterogeneous road users. In general, bi-level programming problems have the following form:

$$\min_{x \in X,\, y} \quad F(x,y) \tag{3.2.1a}$$

$$\text{s.t.} \quad y \in \{\arg\min_{y \in Y} f(x,y) : g(x,y) \leq 0\} \tag{3.2.1b}$$

$$G(x,y) \leq 0 \tag{3.2.1c}$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the upper level objective, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the lower level objective, $G : X \times Y \to \mathbb{R}$ are the upper level constraints and $g : X \times Y \to \mathbb{R}, j = 1, \ldots, J$ are the lower level constraints [77]. The program presented in this paper is an extension to the bi-level program developed by [65], in which a single-level reduction method is suggested to optimize link-based tolls in a multi-commodity network with heterogeneous users with respect to any arbitrary social objective function (that is a nondecreasing in each of its arguments). In this method, the lower-level problem is replaced by it's Karush-Kuhn-Tucker (KKT) conditions, producing a single-level optimization problem [77].

3.3 Problem formulation

Preliminaries

Consider a network $\mathcal{G}(\mathcal{V}, \mathcal{L})$ with OD pairs $w = (w_o, w_d) \in \mathcal{W} \subseteq \mathcal{V} \times \mathcal{V}$. The set $\mathcal{P} = \{\mathcal{P}_w\}_{w \in \mathcal{W}}$ contains the sets of paths (i.e., ordered sets of directed links) connecting each OD pair. The set $\mathcal{M} = \{\mathcal{M}_p\}_{p \in \mathcal{P}}$ denotes the travel modes available on each path.

The users of the network are distinguished by membership to various discrete groups Θ, which may represent demographic or other individual characteristics that factor into their travel preferences. For example, $\Theta \in \mathcal{I} \times \mathcal{A}$ where $\mathcal{I}$ and $\mathcal{A}$ represent income and age groups, respectively. A demand profile $D \in \mathbb{R}_+^{|\mathcal{W}| \times |\Theta|}$ encodes the number of network users of each group $\theta \in \Theta$ that will travel each OD pair $w \in \mathcal{W}$ where $\forall w, \theta : \; d_{w\theta} \geq 0$.

Each network user chooses a strategy $S \in \mathcal{S}_{w\theta} = \{(p,m)\}_{p \in \mathcal{P}_w, m \in \mathcal{M}_p}$, consisting of a travel mode and path, from among the set of available strategies for the corresponding group and OD pair, $\mathcal{S}_{w\theta}$. The available strategy set may vary by user group, enabling the representation

of user groups that may not have access to certain modes of transportation (e.g., non-vehicle owners, persons with physical conditions that restrict travel). The resulting strategy profiles $X_{w\theta} \in \mathbb{R}_+^{|\mathcal{P}_w|\times|\mathcal{M}|}$ are the quantities of each user group and OD pair that choose each strategy. A *feasible strategy profile* is one that fulfills the total demand of each group as follows:

$$\mathcal{X} := \{X_{w\theta} \in \mathbb{R}_+^{|\mathcal{P}_w|\times|\mathcal{M}|} : \sum_{S\in\mathcal{S}_{w\theta}} x_{w\theta S} = d_{w\theta}\}_{w\in\mathcal{W},\theta\in\Theta}.$$

where $x_{w\theta S}$ is one element of the vector $X_{w\theta}$ representing the flow of user group θ traveling OD pair w using strategy S. The individual cost of each strategy is determined by a non-decreasing continuous function of link and mode flows $F_{w\theta S}(g) : \mathbb{R}_+^{|\mathcal{L}||\mathcal{M}|} \to \mathbb{R}$, where $g_{lm} = \sum_w \sum_\theta \sum_{S\in\mathcal{S}_{w\theta}|(l,m)\in S} x_{w\theta S}$ is the total user flow on each link and mode in the network. The cost of each strategy is the weighted sum of the total operational cost and travel time along each link using the corresponding path and mode, as follows:

$$F_{w\theta S}(g, \psi_{\theta m}) = \beta_m^\theta + \alpha_m^\theta \sum_{l\in p} tt_{lm}(g) + tc_{pm}$$

where $S = (p, m)$, β_m^θ is a constant representing the relative preference of user group θ for each mode m, and $\alpha_m^\theta \in \mathbb{R}_+$ is the VoT multiplier specific to group θ and mode m which converts travel time to monetary units. The non-decreasing continuous function $tt_{lm}(g) : \mathbb{R}_+^{|\mathcal{L}||\mathcal{M}|} \to \mathbb{R}_+$ estimates the travel time for mode m on link l and the constant $tc_{pm} \in \mathbb{R}_+$ estimates the travel cost of using mode m on path p.

The model assumes that network users behave selfishly, seeking to minimize the total individual costs of traveling. This selfish behavior leads to a user equilibrium in $\mathcal{G}$, $E(\mathcal{G})$, with the familiar *Nash equilibrium* condition:

$$E(\mathcal{G}) := \{x_{w\theta} \in \mathcal{X}_{w\theta} : \; x_{w\theta S} > 0 \iff F_{w\theta S}(g) \leq F_{w\theta S'}(g) \; \forall S' \in \mathcal{S}_{w\theta}\}_{w\in\mathcal{W},\theta\in\Theta}$$

The existence and uniqueness of a Nash equilibrium strategy profile is well established [66].

Socially optimal traffic assignment and incentive schemes

In the absence of tolls, the Nash equilibrium in $\mathcal{G}$ leads to inefficiencies with respect to the aggregate costs of all network users [78]. Consider that a central authority is charged with the task of setting tolls $\tau_{lm} \in \mathbb{R}_+$ for each mode and link in the network as well as monetary incentives $\psi_{\theta m} \in \mathbb{R}_-$ for each user group and mode in order to minimize a social objective function, $f(g, \tau, \psi) : \mathbb{R}_+^{|\mathcal{L}||\mathcal{M}|} \times \mathbb{R}_+^{|\mathcal{L}||\mathcal{M}|} \times \mathbb{R}_-^{|\Theta||\mathcal{M}|} \to \mathbb{R}$. Note that the social cost function is a function of g, the aggregate link-level flows by mode, and not of the individual strategy profiles. As in reality (in most transportation systems), the central planner does not have access to information on the distribution of traffic flow at an individual level. The task of the central planner is to optimize the aggregate link- and mode-level flows, hereby referred

to as the socially optimal traffic assignment (SOTA), with respect to the social objective function.

A SOTA g is *feasible* if there exists a feasible strategy profile x such that the total flow of users on each link and mode is less than or equal to the corresponding SOTA flow:

$$\forall l \in \mathcal{L}, m \in \mathcal{M} : \sum_{w} \sum_{\theta} \sum_{S=(p,m) \in \mathcal{S}_{w\theta} | l \in p} x_{w\theta S} \leq g_{lm}.$$

Furthermore, a SOTA and incentive scheme (g, ψ) is *enforceable* if there is a set of non-negative tolls τ such that the strategy profile induced by the Nash equilibrium of $\mathcal{G}$ with respect to τ and ψ, $\mathcal{G}^{\tau\psi}$, produces link-level flows equal to g. More specifically, the Nash equilibrium $E(\mathcal{G}^{\tau\psi})$ is:

$$E(\mathcal{G}^{\tau\psi}) := \{x_{w\theta} \in \mathcal{X}_{w\theta} : \; x_{w\theta S} > 0 \iff F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l \in p} \tau_{lm} \leq$$
$$F_{w\theta S'}(g) + \psi_{\theta m'} + \sum_{l' \in p'} \tau_{l'm'} \; \forall S' = (p', m') \in \mathcal{S}_{w\theta}\}_{w \in \mathcal{W}, \theta \in \Theta}$$

Furthermore, $x = E(\mathcal{G}^{\tau\psi}) \implies \sum_w \sum_\theta \sum_{S=(p,m) \in \mathcal{S}_{w\theta} | l \in p} x_{w\theta S} = g_{lm} \; \forall l \in \mathcal{L}, m \in \mathcal{M}$.

Fleischer, et al. (2004) proved the existence of tolls that enforce a SOTA that minimizes any arbitrary function from $\mathbb{R}^{\mathcal{V}} \to \mathbb{R}$ that is non-decreasing in each of its arguments [65]. This chapter establishes the existence of tolls that enforce a SOTA and incentive scheme.

3.4 Existence of an optimal pricing strategy

This section proves that tolls exist to enforce a socially optimal traffic assignment and incentive (SOTAI) scheme. The proof extends the theorems and corollaries presented in [65] for the inclusion of individual- and mode-specific incentives in optimal road pricing mechanisms.

Given a SOTAI scheme (g, ψ) to enforce, the following linear program $\mathbf{P}_{g\psi}$ characterizes the Nash equilibrium strategy profiles:

$$\min_{x} \quad \sum_{w} \sum_{\theta} \sum_{S \in \mathcal{S}_{w\theta}} (F_{w\theta S}(g) + \psi_{\theta m}) x_{w\theta S} \tag{3.4.1a}$$

$$\text{s.t.} \quad \forall w, \theta : \sum_{S \in \mathcal{S}_{w\theta}} x_{w\theta S} = d_{w\theta} \tag{3.4.1b}$$

$$\forall l, m : \sum_{w} \sum_{\theta} \sum_{S=(p,m) \in \mathcal{S}_{w\theta} | l \in p} x_{w\theta S} \leq g_{lm} \tag{3.4.1c}$$

$$\forall w, \theta, S \in \mathcal{S}_{w\theta} : x_{w\theta S} \geq 0 \tag{3.4.1d}$$

where equation 3.4.1a minimizes the total individual costs, equation 3.4.1b ensures the traffic assignment in the SOTAI is feasible, equation 3.4.1c ensures that the total demand is

fulfilled by the strategy profile, and equation 3.4.1d constrains the strategy profiles to be non-negative.

The dual of this problem, $\mathbf{D}_{g\psi}$, is as follows:

$$\max_{z,\tau} \quad \sum_{w}\sum_{\theta} d_{w\theta} z_{w\theta} - \sum_{l}\sum_{m} g_{lm}\tau_{lm} \tag{3.4.2a}$$

$$\text{s.t.} \quad \forall w,\theta,S=(p,m)\in\mathcal{S}_{w\theta} : z_{w\theta} \le F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l\in p}\tau_{lm} \tag{3.4.2b}$$

$$\forall l,m : \tau_{lm} \ge 0 \tag{3.4.2c}$$

where the non-negative dual variable corresponding to the SOTAI feasibility constraint in the primal problem (equation 3.4.1b) is the toll vector, τ, and the dual variable corresponding to the demand-fulfilling constraint (equation 3.4.1c) is the vector of total travel costs in the Nash equilibrium, z. The complementary slackness condition is equivalent to the Nash equilibrium condition for $E(\mathcal{G}^{\tau\psi})$. In particular, if x^* and (z^*,τ^*) are the optimal solutions to $\mathbf{P}_{g\psi}$ and $\mathbf{D}_{g\psi}$, respectively, then:

$$\forall w,\theta,S=(p,m)\in\mathcal{S}_{w\theta} : x^*_{w\theta S} > 0 \implies z^*_{w\theta} = F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l\in p}\tau^*_{lm}.$$

Next, I prove the existence of optimal tolls to enforce a SOTAI. The following proof is analogous to the corresponding proof in [65] of the enforceability of a traffic assignment since the inclusion of incentives in the SOTAI scheme results in a simple affine transformation of the objective of $\mathbf{P}_{g\psi}$ (with respect to $\mathbf{P}_g$).

Theorem 3.4.1 (Enforceable traffic assignment and incentive schemes)**.** *A feasible traffic assignment and incentive scheme (g,ψ) is enforceable if and only if the linear program $\boldsymbol{P}_{g\psi}$ has an optimal solution in which, for every link $l\in\mathcal{L}$ and mode $m\in\mathcal{M}$, the inequality 3.4.1b is tight.*

Proof. First, we prove that the enforceability of a SOTAI scheme (g,ψ) implies that an optimal solution to $\mathbf{P}_{g\psi}$ exists in which equation 3.4.1b is tight. Let (g,ψ) be an enforceable SOTAI scheme. Then there is a set of nonnegative tolls τ such that equation 3.4.1b is tight given the Nash equilibrium strategy profile $x = E(\mathcal{G}^{\tau,\psi})$. By definition of $E(\mathcal{G}^{\tau,\psi})$, for every OD pair w and group θ, the total latency $F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l\in p}\tau_{lm}$ of any available strategy $S=(p,m)\in\mathcal{S}_{w\theta}$ with nonzero flow (i.e., $x_{w\theta S}>0$) must be equal; let that value be $z_{w\theta}$. The total latency of any strategy not being used must be greater than $z_{w\theta}$ such that $z_{w\theta} \le F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l\in p}\tau_{lm}$ for all strategy profiles $S=(p,m)$ available to any OD pair and group. Thus, x and (τ,z) are feasible solutions to $\mathbf{P}_{g\psi}$ and $\mathbf{D}_{g\psi}$, respectively, and they satisfy the complementary slackness conditions. Thus, x is an optimal solution for $\mathbf{P}_{g\psi}$.

Next, we prove that if there exists an optimal solution to $\mathbf{P}_{g\psi}$ in which equation 3.4.1b is tight, (g,ψ) is an enforceable SOTAI scheme. Let x^* be an optimal solution for $\mathbf{P}_{g\psi}$ such that, for every mode m and link l, the sum of flows for strategies using l and m is equal to g_{lm}

(inequality 3.4.1b is tight). Furthermore, let (τ^*, z^*) be the corresponding optimal solution for $\mathbf{D}_{g\psi}$. By the complementary slackness condition of $\mathbf{P}_{g\psi}$, for every OD pair w and group θ, the inequality 3.4.2b corresponding to any strategy $S = (p, m) \in \mathcal{S}_{w\theta}$ with nonzero flow (i.e., $x^*_{w\theta S} > 0$) must be tight: $z^*_{w\theta} = F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l\in p} \tau^*_{lm}$. For any strategy that is not being used (i.e., $x^*_{w\theta S} = 0$), $z^*_{w\theta} \leq F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l\in p} \tau^*_{lm}$. Thus, $x^* = E(\mathcal{G}^{\tau\psi})$, which means that g is enforceable.

□

The following proofs show the existence of tolls to enforce an *optimal* SOTAI scheme that minimizes a social objective function with arguments in the dimension of either the: 1) network facilities (i.e., the links and modes) or, 2) demand characteristics (i.e., the OD pairs and user groups).

Theorem 3.4.2 (Existence of tolls to enforce a flow-based socially optimal traffic assignment and incentive schemes)**.** *For every multi-modal multicommodity network in the discrete setting, there are tolls that enforce a feasible traffic assignment and incentive scheme that optimizes a function $f : \mathbb{R}^{\mathcal{L}} \times \mathbb{R}^{\mathcal{M}} \to \mathbb{R}$ that is non-decreasing in each of its arguments.*

Proof. Let $(g^*, \psi^*) \in \arg\min_{g,\psi} f(g)$ be the optimal solution to the function $f : \mathbb{R}^{\mathcal{L}} \times \mathbb{R}^{\mathcal{M}} \to \mathbb{R}$ that is non-decreasing in each of its arguments. Suppose that x is the optimal solution to P_{g^*,ψ^*} and there is some link l and mode m such that $\sum_w \sum_\theta \sum_{S=(p,m)\in\mathcal{S}_{w\theta}|l\in p} x_{w\theta S} < g^*_{lm}$. Thus, there exists a feasible g' such that $g'_{lm} < g^*_{lm}$. However, $f(g') < f(g^*)$, which contradicts the optimality of g^*. □

A similar argument can be made to prove the existence of tolls to optimize a function of tolls. For the next proof, I establish the notion of a *minimal network*: a network $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ in which there is no link or mode for which the Nash Equilibrium strategy given a feasible SOTAI is zero:

$$\mathcal{G} = (\mathcal{V}, \mathcal{L}) : \ \forall l \in \mathcal{L}, m \in \mathcal{M} : \exists x = E(\mathcal{G}^{\tau,\psi}), \sum_w \sum_\theta \sum_{S=(p,m)\in\mathcal{S}_{w\theta}|l\in p} x_{w\theta S} > 0$$

Theorem 3.4.3 (Existence of tolls to enforce a user-based socially optimal traffic assignment and incentive schemes)**.** *For every minimal multi-modal multicommodity network in the discrete setting, there are tolls that enforce a feasible traffic assignment and incentive scheme that optimizes a function $f : \mathbb{R}^{\mathcal{W}} \times \mathbb{R}^{\Theta} \to \mathbb{R}$ that is non-decreasing in each of its arguments.*

Proof. Let $(g^*, \psi^*, z^*) \in \arg\min_{g,\psi,z} f(z)$ be the optimal solution to the function $f : \mathbb{R}^{\mathcal{W}} \times \mathbb{R}^{\Theta} \to \mathbb{R}$ that is non-decreasing in each of its arguments. Furthermore, suppose that x and (z^*, τ) are the optimal solutions to P_{g^*,ψ^*} and D_{g^*,ψ^*}, respectively, and there is some link l and mode m such that $\sum_w \sum_\theta \sum_{S=(p,m)\in\mathcal{S}_{w\theta}|l\in p} x_{w\theta S} < g^*_{lm}$. Thus, there exists a feasible g' such that $g'_{lm} < g^*_{lm}$. There are two cases to consider: $g'_{lm} > 0$ and $g'_{lm} = 0$. In the case that g'_{lm} is nonzero, then there is at least one user group (w, θ) using a strategy $S = (p, m) \in \mathcal{S}_{w\theta}$ that

uses mode m and link l. By the complementary slackness condition for the optimality of x and (z^*, τ), the total latency of that strategy must be equal to $z^*_{w\theta} = F_{w\theta S}(g^*) + \psi^*_{\theta m} + \sum_{l \in p} \tau_{lm}$. However, since $F_{w\theta S}(g') < F_{w\theta S}(g^*)$, $z^*_{w\theta} > F_{w\theta S}(g') + \psi^*_{\theta m} + \sum_{l \in p} \tau_{lm}$, a contradiction. In the second case, $\sum_w \sum_\theta \sum_{S=(p,m) \in \mathcal{S}_{w\theta} | l \in p} x_{w\theta S} = g'_{lm} = 0$. This contradicts the condition that the network is minimal. □

3.5 Bi-level problem formulation

The SOTAI design problem is thus composed of a social planner's decision problem, $\mathbf{C}_{g\psi}$, in which the optimal policy (g, ψ, τ), is determined according to a real-valued social choice function f. While g is constrained by the condition of enforceability that there must exist an optimal solution to the linear program $\mathbf{P}_{g\psi}$ for which the total user flow on each link and mode is exactly equal to the corresponding value of g, it is necessary to design a constraint on the value of ψ. One option is to enforce the condition of *budget balancing* on the SOTAI optimization problem by constraining the sum of the total tolls paid and incentives used to be greater than or equal to zero. The following problem is one such instantiation:

$$\min_{g,\psi} \quad f \tag{3.5.1a}$$

$$\text{s.t.} \quad \sum_w \sum_\theta \sum_{S \in \mathcal{S}_w | m \in S} \psi_{\theta m} x_{w\theta S}(g, \psi) + \sum_l \sum_m g_{lm} \tau_{lm}(g, \psi) \geq 0 \tag{3.5.1b}$$

$$\forall l, m : \sum_w \sum_\theta \sum_{S=(p,m) \in \mathcal{S}_{w\theta} | l \in p} x_{w\theta S}(g, \psi) = g_{lm} \tag{3.5.1c}$$

$$\forall l, m : 0 < g_{lm} \leq \kappa_{lm} \tag{3.5.1d}$$

$$\forall \theta, m : \psi_{\theta m} \leq 0 \tag{3.5.1e}$$

$$x(g, \psi) = \arg\min \mathbf{P}_{g\psi} \tag{3.5.1f}$$

$$(z(g, \psi), \tau(g, \psi)) = \arg\max \mathbf{D}_{g\psi} \tag{3.5.1g}$$

where constraint 3.5.1b balances the total subsidies distributed with the total toll revenue generated and 3.5.1d ensures that any capacity constraints of the network are met and that the network is minimal. In addition, $x(g, \psi)$, $z(g, \psi)$, and $\tau(g, \psi)$ are the optimal values of the primal and dual problems, $\mathbf{P}_{g\psi}$ (equations 3.4.1a - 3.4.1d) and $\mathbf{D}_{g\psi}$ (equations 3.4.2a - 3.4.2c), respectively.

To solve this problem via the single-level reduction method, the stationarity constraint (equation 3.4.2b) and complementary slackness conditions are added, as follows:

$$\min_{g,\psi,z,\tau,x} \quad f \tag{3.5.2a}$$

$$\text{s.t.} \quad \sum_w \sum_\theta \sum_{S \in \mathcal{S}_w | m \in S} \psi_{\theta m} x_{w\theta S} + \sum_l \sum_m g_{lm} \tau_{lm} \geq 0 \tag{3.5.2b}$$

$$\forall l, m : \sum_w \sum_\theta \sum_{S=(p,m) \in \mathcal{S}_{w\theta} | l \in p} x_{w\theta S} = g_{lm} \tag{3.5.2c}$$

$$\forall w, \theta, S = (p, m) \in \mathcal{S}_{w\theta} : z_{w\theta} \leq F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l \in p} \tau_{lm} \tag{3.5.2d}$$

$$\forall w, \theta, S = (p, m) \in \mathcal{S}_{w\theta} : x_{w\theta S}(z_{w\theta} - F_{w\theta S}(g) + \psi_{\theta m} + \sum_{l \in p} \tau_{lm}) = 0 \tag{3.5.2e}$$

$$\forall l, m : 0 < g_{lm} \leq \kappa_{lm} \tag{3.5.2f}$$

$$\forall \theta, m : \psi_{\theta m} \leq 0 \tag{3.5.2g}$$

$$\forall w, \theta, S \in \mathcal{S}_{w\theta} : x_{w\theta S} \geq 0 \tag{3.5.2h}$$

$$\forall l, m : \tau_{lm} \geq 0 \tag{3.5.2i}$$

Social choice functions

Theorems 3.4.2 and 3.4.3 establish that any social choice function from the space of the network links and modes or of the user groups and ODs that is nondecreasing in each of its arguments can be optimized to produce an enforceable SOTAI. The following subsections present several potential specifications of social choice functions that implement different theories of justice.

Utilitarian

The most prominent philosophy applied in the valuation of transportation system interventions is utilitarianism, which is based on the view that the maximization of social welfare, as measured by the total individual "utility" of a particular action or policy, achieves justice regardless of the differences in the preferences or circumstances across the population [79]. The utilitarian approach is directly applied in cost-benefit analyses, in which alternatives are evaluated by estimating the total costs and benefits - including the total dollar-valued utility - of each alternative. A utilitarian social choice function in the context of the SOTAI optimization problem is what is typically referred to in the traffic assignment and road pricing optimization literature as the social optimum: the sum of the total cost across all user groups and OD pairs, as follows:

$$f(z) = \sum_w \sum_\theta d_{w\theta} z_{w\theta} \tag{3.5.3}$$

The existence of optimal tolls to enforce a utilitarian SOTAI mechanism follows directly from theorem 3.4.3. By employing the individual payoffs of network users, a utilitarian social choice function implicitly weighs the impacts of the SOTAI scheme by the relative VoT of each user group. As a result, the travel time reductions experienced by the individuals with the highest VoT will weigh more heavily than those experienced by individuals with lower VoT.

Egalitarian

Egalitarianism prioritizes *equality* of treatment, suggesting that the outcomes of a set of alternative actions or policies should be evaluated with equal weight across all individuals. In the context of the SOTAI optimization problem, I define an egalitarian social choice function as the sum of the total *average* cost across all users in the network, as follows:

$$f(\bar{z}) = \sum_{w} \sum_{\theta} d_{w\theta} \bar{z}_{w\theta} \tag{3.5.4}$$

where $\bar{z}_w$ depends on a modified individual cost function, $\bar{F}(x)$ that estimates the cost of a particular strategy to the average network user. Specifically, in the modified individual cost function $\bar{F}_{w\theta S}$, the parameters β^θ and α^θ of $F_{w\theta S}$ are replaced by estimators of the population averages, $\bar{\beta}$ and $\bar{\alpha}$, respectively, as follows: $\bar{F}_{wS}(g, \psi_{\theta m}) = \bar{\beta}_m + \bar{\alpha}_m \sum_{l \in p} tt_{lm}(g) + tc_{p,m}$. The Nash equilibrium conditions that define the value of z are also applicable to $\bar{z}$ and must be included in the constraints of the egalitarian SOTAI problem, as follows:

$$\forall w, \theta, S = (p, m) \in \mathcal{S}_{w\theta} : \ \bar{z}_w \leq \bar{F}_{wS}(g) + \psi_{\theta m} + \sum_{l \in p} \tau_{lm} \tag{3.5.5}$$

$$\forall w, \theta, S = (p, m) \in \mathcal{S}_{w\theta} : \ x_{w\theta S}(\bar{z}_w - \bar{F}_{wS}(g) + \psi_{\theta m} + \sum_{l \in p} \tau_{lm}) = 0 \tag{3.5.6}$$

Equality of Opportunity

The equality opportunity view of justice emphasizes that all individuals have equal access to the same opportunities regardless of their position in society [80]. Transportation plays an integral role in providing access to opportunities by connecting individuals to destinations and activities that contribute to their quality of life and social mobility. However, access to transportation itself varies widely across spatial, temporal, economic, and physiological dimensions of the population [81]. While the equilibrium cost of any particular strategy under a SOTAI scheme may be individually optimal given an individual's preferences and the costs of the alternatives, the burden of that cost with respect to the individual's circumstances (e.g., income, age, physical condition, health or safety concerns, etc.) may be disproportionate to the burden of the same strategy for an individual of different circumstances. Thus, I interpret the theory of equality of opportunity as justifying the application of OD- and/or

group-specific weights in proportion to factors impacting individual equality of opportunity as follows:

$$f_\omega(\bar{z}) = \sum_w \sum_\theta \omega_{w\theta} d_{w\theta} \bar{z}_{w\theta} \tag{3.5.7}$$

where ω is a vector of weights that scales up with the relative level of disadvantage of each user group θ and/or OD pair w, as determined by the social planner. Again, $\bar{z}$ and $\bar{F}$ are used in order to avoid implicitly weighing the costs of each user group differently, thus allowing the vector ω to be the sole weighting factor in this social choice function.

Rawlsian

Rawls' theory of justice is comprised of two main principles: 1) rules that define individuals' basic rights and liberties should apply equally across individuals and enable the maximum amount of freedom as long as no individual's freedom infringes on that of another, and 2) inequalities in the social and/or economic impacts of a particular action or policy can only be considered fair if they both a) provide equality of opportunity, and b) work to the benefit of the least well off [79]. The latter part of the second principle is called the *difference principle*; it prioritizes mitigation of existing inequities by placing higher value on outcomes that 'raise the floor' of the distribution of outcomes. The maximin criterion has been proposed to implement the difference principle: maximize the welfare of the individual with the minimum welfare across the population [79]. The following social choice function implements the Rawlsian theory of justice in the SOTAI problem by minimizing the maximum cost across OD pairs and user groups:

$$\hat{f}(\bar{z}) = \max_{w\theta} \bar{z}_{w\theta} \tag{3.5.8}$$

where we again take an agnostic approach to the estimation of cost across user groups using $\bar{z}$ as opposed to z.

3.6 Conclusions

This chapter formulates a bi-level optimization problem for socially optimal toll and incentive schemes in multimodal multi-commodity networks with heterogeneous users. Whereas prior studies on optimal tolling schemes that have included monetary transfers (also called subsidies) have defined the transfers either in the same dimension as the tolls (typically link- or path-based) or simply by the dimension of the user heterogeneity, I develop a SOTAI scheme that defines incentives by mode and the dimensions of user heterogeneity. Thus, while the tolls are agnostic of the network user paying them, the incentives can target specific user groups.

This chapter proves the existence of tolls to enforce an optimal SOTAI scheme with respect to functions of the network links and modes or of the user groups and OD pairs. The latter enables the optimization of tolling and incentive schemes with respect to social choice functions that prioritize the mitigation of transportation equity issues that arise from

the disparate impacts of road pricing policies on certain user groups. This chapter presents four social choice function alternatives, the latter three of which deiate from the commonly assumed utilitarian view on road pricing. Analysis of the impact of these alternative social choice functions on the optimal traffic assignment, tolls, and incentives is in progress, including application of the bi-level optimization model for empirical experiments.

Chapter 4

The Berkeley Integrated System for Transportation Optimization

4.1 Introduction

In this chapter, I present the *Berkeley Integrated System for TRansportation Optimization* (BISTRO), an open-source simulation-based optimization framework and software toolkit designed to aid transportation researchers, policy-makers, and other stakeholders in developing multi-objective transportation system intervention strategies. This chapter was written in collaboration with Sidney Feygin, Edward Forscher, Valentine Golfier-Vetterli, Jonny Lee, Abhishek Gupta, Rashid Waraich, Colin Sheppard, and Alexandre Bayen. It was published in *ACM Transactions on Intelligent Systems and Technology* in July, 2020. I led the development of the objective function design for BISTRO by translating municipal and regional transportation goals into quantitative optimization objectives with consideration for social, environmental, and economic sustainability implications and testing the efficacy with which these key performance indicators (KPIs) functioned when applied in simulation-based multi-objective optimization. In addition, I led the design of the output analysis and visualization platform for BISTRO. My individual contributions are presented in the Scoring function design, Inputs, and Output analysis and visualization sections of this chapter in addition to the corresponding sections of Appendix A.

As modern transportation systems undergo a period of intense technological change, researchers, practitioners, and policymakers are seeking to understand how long-term trends towards vehicle digitalization, automation, electrification, as well as the emerging sharing economy will shape the future day-to-day dynamics of human mobility in cities worldwide.Likewise, rapid advances in computing as well as the advent of metropolitan scale data mining and pattern recognition (*i.e.*, machine learning) have, for the first time, made it possible to characterize urban traffic flows based on movement traces from millions of individual travelers. These methods fuse passively collected spatiotemporal trajectories derived from smartphone data with static census data in order to map travel patterns to sociodemo-

graphic characteristics. Specialized traffic models can thereby be trained using feature-rich representations of human mobility.

While such predictive models have proven effective in nearcasting congestion events [82], purely data-driven methods often fail to generalize when applied to the task of forecasting the effect of transportation policy strategies on future demand–particularly when anticipated changes in urban sociodemographics, regulatory policy, and mobility technologies as well as interactions between these layers cannot themselves be predicted with a high degree of certainty [83]. In other words, the covariates of statistical models may not capture the numerous interacting physical and behavioral components (*i.e.*, *network effects*) influencing transportation supply and demand.

Agent-based modeling and simulation (ABMS) techniques and software, on the other hand, are beginning to be used as flexible decision-support tools by planning groups, technologists, and regulatory agencies to facilitate forecasting the short- and long-term implications of transportation system interventions. That is, by simulating models of traveler decision-making in the context of a landscape of future plausible "states of the world", stakeholders can better resolve uncertainty over how transportation system interventions may fare. For example, ABMS enables the investigation of the extent to which citizens may adopt on-demand autonomous vehicles, as well as how to achieve sustainable fleet operation [84, 85].

Even with these advances, selecting interventions that balance competing transportation system policy objectives remains a difficult and contentious process. Current methods to identify the best alternative under a given scenario often involves the use of Monte Carlo methods to evaluate different policy options. Such approaches can be extremely computationally expensive for large ABMS implementations and are unlikely to yield an optimal solution. Moreover, transportation researchers and practitioners often develop scenario-specific and/or geography-specific simulation models with limited integration of the efficient and highly generalizable methods developed in the *artificial intelligence* (AI) and *machine learning* (ML) communities.

BISTRO is an open-source framework and software toolkit designed to address the increasingly complex problems arising in transportation systems worldwide. BISTRO does this through an innovative human-machine collaboration approach, using state-of-the-art optimization algorithms to efficiently automate the search for policy interventions that achieve good performance over diverse transportation system scenarios and stakeholder objectives. BISTRO includes an ABMS system and scenario development pipeline to build empirically-calibrated simulations of travel demand in metropolitan transportation systems. After a calibrated scenario has been developed, BISTRO enables use of state-of-the-art optimization algorithms to identify system interventions that best align with policy and planning objectives. Users can deploy BISTRO to enable distributed development of algorithms that rapidly optimize a feasible set of policy and investment decisions. Once one or more desirable solutions are found, BISTRO provides a suite of analysis and visualization tools to empower citizens, transportation system planners and engineers, private entities, and governments to better understand and collaborate on developing strategies that achieve equitable access to and sustainable use of current and emerging mobility services.

The rest of this chapter is organized as follows: Section 4.2 provides a brief overview of the current state of transportation planning, ABMS tools, and transportation optimization tools, respectively; Section 4.3 presents BISTRO, covering the system architecture, scoring function design, inputs, outputs, analysis capabilities, and performance characteristics; Section 4.4 details an initial pilot study, updates to BISTRO based upon the pilot, and a few algorithmic solution approaches; and Section 4.5 offers a short conclusion.

4.2 Background

Transportation Planning and Policy Decision Support Systems

In the United States, the primary outcome of efforts to design, model, and communicate the impacts of proposed policy interventions and infrastructure investments on transportation systems are Metropolitan *Regional Transportation Plans* (RTPs) and State *Long-Range Transportation Plans* (LRTPs). These plans are forward-looking, long-term (20+ year time horizons) and have been a federally mandated task since the Federal-Aid Highway Act of 1962. Recently, state departments of transportation and *Metropolitan Planning Organizations* (MPOs)—the entities tasked with producing RTPs every four-to-five years—have been shifting towards performance-based planning and programming (PBPP) frameworks [86].

As part of RTP development using PBPP guidelines, an MPO or related agency will typically conduct a community engagement process to identify one or more *visions* or *goals* shared by stakeholders and the public describing a desired future state of the regional transportation system (e.g., safe roadways, accessible transit, environmental stewardship) [87, 88, 86]. Measurable *objectives* are defined together with quantitative *performance indicators* in order to evaluate the extent to which alternative *strategies* consisting of policy interventions or infrastructure investments could make progress towards achievement of a goal [87, 86]. Typically, the impact of these projects on *key performance indicators* (KPIs) of transportation system performance will be forecast using an analytic, data-driven model of travel demand[1].

The domain expertise necessary to understand the functionality of models of transportation demand may result in recommendations that are frustratingly opaque to public interpretation [88]. Explanations of the inner workings of the modeling process are not presented during public collaboration meetings and questions regarding the validity of models have led to lawsuits over the lack of publicly available information on model specifics [89]. Often, contractual obligations or proprietary data formats used by the consulting firms tasked with developing planning software further restrict public access. The inability or unwillingness on

[1]Traditionally, the travel demand modeling process consists of four main steps: 1) trip generation to and from all analysis zones, 2) trip distribution (or matching origins and destinations, often using a gravity model), 3) assigning traveler mode choice based upon individual preferences and alternative characteristics, and 4) route assignment, of trips onto physical network links; this is referred to as the four-step model. Many MPOs and other agencies are moving towards disaggregate, activity-based, or person-centric models of daily activity, rather than aggregate approaches operating on the zonal level [87].

the part of MPOs to promote flexible open-source software licensing strategies discourages independent investigations (by, for example, academic or public interest groups) to verify that the theory, equations, algorithms, and data comprising a model match its software implementation. This lack of operational transparency not only subverts accountability, but also impedes the rapid transfer of innovative modeling methods and technologies across and even within agencies [90, 91].

To address criticism concerning model transparency and explainability, recent transportation and land-use planning organizations have begun to pilot software that visualizes the impacts of alternatives and provide fora for public input via collaborative simulation platforms [87, 92, 93, 94, 95]. An interesting example is the UrbanAPI project, which includes a web-based 3-dimensional virtual reality visualization of the impacts of urban growth scenarios at the scale of individual neighborhoods [94]. While initial efforts have focused on the usability requirements of these tools, their broader impact has been limited due to narrowly defined geographic contexts or specialization for regional system objectives [96].

One project that does have broad support across several MPOs is ActivitySim, which currently being developed by the Association of Metropolitan Planning Organizations as an open source activity-based travel modeling platform [97]. ActivitySim may be used in concert with synthetic households, synthetic persons, employment data, land use data, and network performance data (e.g., travel times by mode and time of day, costs, and transfers) to generate a full-scale model for a city's travel demand. ActivitySim may be seen as a complementary project to our own in that ActivitySim model outputs can be used as input data for a BISTRO scenario.

ABMS of Transportation Systems

Developing models that replicate how urban systems operate and evolve has been a major focus of transportation engineers, urban planners, and geographers. Trip-based methods, such as the traditional four-step model used by MPOs, are specified at aggregate geographic or categorical levels rather than at the level of individual decision-makers. This level of aggregation can limit the ability of such models to explain complex individual decisions. Some MPOs have begun to adopt a more behaviorally-descriptive activity-based approach [98]. In contrast to trip-based models, activity-based methods represent more comprehensive links between activity scheduling, mode choice, social interaction, and spatiotemporal constraints [98]. Agent-based models and simulations (ABMS) of transportation demand are capable of replicating observed *macroscopic* traffic patterns by simulating the *microscopic* decision-making behavior of a synthetic population of software agents as they execute their daily travel plans on a virtual model of the transportation system. When *calibrated* to ground-truth data by modifying only global parameters characterizing the embedded choice model, agent-based frameworks represent parsimonious descriptions of regional travel demand. Consequently, ABMS are capable of accurately capturing shifts in macro patterns when forecast changes in transportation infrastructure, policies, demographics, and vehicle ownership are introduced to the virtual travel environment [83].

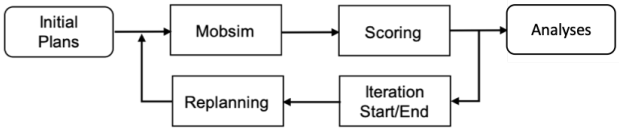


Figure 4.1: Conceptual process of MATSim. It iteratively evaluates and mutates a proportion of agent plans until the utility of plans no longer improves. At this point, the system is said to have reached a *stochastic user equilibrium*. For further details, see [84].

In the past three decades, several multi-agent frameworks such as TRANSIMS [99], MATSim [84], SUMO [100], and POLARIS [92] have been developed and widely adapted for numerous applications in transportation and land use planning and research [101, 102, 83, 96]. MPOs often couple urban development simulation models[2] with microscopic agent-based transport models to better understand how predicted changes in population growth, land-use, real-estate development, and resource markets will co-evolve with changes in the transportation system [105, 106]. ABMS of transportation systems can take different population configuration files as inputs, giving planners and modelers the ability to simulate how long-term urbanization processes can be shaped by the daily transportation decisions of individuals (possibly in the presence of alternative policy interventions and new mobility technologies) [107, 106]. Increasingly, these open-source platforms are enabling use of publicly available data to create transparent and replicable input preparation pipelines, reducing the cost and effort of a variety of urban planning tasks [108]. Since BISTRO primarily relies on BEAM and BEAM itself incorporates many aspects of MATSim, the remainder of this section takes a closer look at these two software frameworks' purpose, functionality, and computational characteristics.

MATSim MATSim is an ABMS framework developed by teams at ETH Zürich and TU Berlin [84]. MATSim enables simulating the travel behavior of millions of individual agents, representing a synthetic population of urban travelers. At the heart of MATSim is a co-evolutionary algorithm that iteratively executes, evaluates, and mutates (i.e., replans) the *daily* activity schedules of agents (see Figure 4.1). The end result of this process is an equilibrium between network supply and travel demand–resulting in realistic congestion patterns as agents compete for limited space on a virtual road network.

Besides road network and transit data, the key input to MATSim is the agent *population*. This file encodes a set of unrealized plans (one for each agent) consisting of the start times, types (e.g., "Home", "Work", "Shopping", "School", etc.) and locations of various significant activities[3]. A mobility simulation (`MobSim`) executes these plans on a virtual road network. While, at first, agents only drive and/or walk to activities, additional modes may

[2]Examples of these include the *Integrated Transportation, Land Use, and Environment* model (ILUTE, [103]) and UrbanSim [104].

[3]The population file is often generated from the output of an activity-based travel demand model.

be introduced during plan mutation (explained below). Each agent's plan is then scored[4]. Copies of evaluated plans are stored in a limited-size array, representing the agent's memory. At the start of the subsequent iteration, a portion of agents in the population are chosen to have a randomly selected mutation strategy applied to a plan drawn from each of their memories. Examples of plan mutations include changing activity start time, mode of travel for a tour (or subtour), and selecting a different route based on previously experienced link travel times. Optional extensions to MATSim model further behavioral dimensions such as parking choice, group travel, and vehicle sharing.

The algorithm converges once agents are no longer able to improve the score of plans in their memory, at which point MATSim produces a series of statistics and outputs describing the aggregate performance of system components as well as a snapshot of all events that occurred over the course of the simulation. Events can be processed to derive the actual paths and travel times realized by each agent and each vehicle, as well as other data reflecting simulation performance.

BEAM The *Behavior Energy Autonomy and Mobility* (BEAM) framework is a multi-agent travel demand simulation framework developed at *Lawrence Berkeley National Laboratory* (LBNL) [109]. While the overall learning and traffic assignment mechanism is similar to MATSim's co-evolutionary algorithm, added functionality in BEAM is specifically focused on helping users understand the impacts of new and emerging travel modes on limited capacity resource markets. This subsection describes essential features of BEAM that led to its selection as the core simulation engine of BISTRO.

BEAM is closely integrated with the transit service capabilities of the R5 routing engine, which include *General Transit Feed Specification* (GTFS) file processing and routing based on multiobjective variations of the RAPTOR algorithm [110]. Transit may be combined with other modes modeled in BEAM such as autonomous vehicles, on-demand rides, e-bikes, and scooters, enabling agents to make realistic, multimodal mobility decisions. In order to provide agents with information about the time and monetary costs of different travel options, R5 computes the lowest generalized cost path (based on travel time estimates from the mobility simulation) for the corresponding mode(s) available to the agent for the trip.[5]

Unlike the replanning mechanism in MATSim, which only mutates plans *between* consecutive iterations, agents in BEAM are designed to adapt to changing conditions *during* an iteration according to what is known as a *within-day* or *online* model[6]. Thus agents can make unplanned and time-sensitive choices about how to maximize the score of their travel plans while competing for limited resources that vary in availability over time. For example, an agent that chooses a transit mode may be denied access to an overfull bus, requiring the agent to make a mid-trip change to their itinerary. The agent could then choose to wait

[4]This score can be interpreted as econometric utility. It is measured by a linear model that assigns negative value to time spent traveling and positive value to time spent at activities.

[5]For detailed information about the R5 router, see [111].

[6]While MATSim has a within-day mode, much of the functionality enabled by extension modules representing emerging mobility technologies assumes that replanning happens between iterations.

for the next bus or hail a ride from the point of departure (if a driver is available nearby). The BEAM software architecture addresses the performance and complexity challenges of integrating new models of within-day dynamics by implementing agents as *actors*, as defined within the *actor-based model of concurrency*[7].

Like MATSim, BEAM enables users to model realistic variations in travel preferences predicated on agent characteristics (which are themselves derived from sociodemographic statistics computed on census data and travel surveys). However, the mechanism for selection of travel alternatives differs significantly from MATSim's in order to model within-day decision-making. Specifically, the probability of selecting an alternative (route, mode, parking choice, and refueling decision) is represented in BEAM according to a multinomial logit model [113, 114]. That is, among several distinct travel schedules, agents are exponentially more likely to select the option that maximizes their enjoyment of important activities while reducing time and money spent traveling between activity locations. The choices made by BEAM agents and their corresponding scores comprise a BEAM plan, which is stored in memory following its execution. Thus, in the transit choice scenario described above, an agent's response to a full bus could incorporate multiple downstream choices within one iteration rather than wait for a score penalty to propagate an optimal response through the selected plan (i.e., using plan mutations that take place over the course of multiple iterations)[8].

Simulation-based Optimization of Transportation Systems

Optimization-based formulation of the planning problem

The problem class solved by the BISTRO framework can be characterized as simulation-based optimization of large urban transportation systems. It can be symbolically formulated as an optimization problem:

$$\underset{\vec{d}\in\mathcal{D}}{\text{minimize}}\ f(\vec{d},\vec{x};\vec{z}) \equiv \mathbb{E}[F(\vec{d},\vec{x};\vec{z})] \tag{4.2.1}$$

[7]Like objects in the object-oriented programming paradigm, actors encapsulate state and behavior. However, unlike the object model, actors do not share computer memory. Instead, each actor encapsulates its own thread of execution and interacts with one other actors using messages. An actor may send message to other actors without blocking. Each actor processes messages synchronously in the order received; however, computation is scheduled asynchronously over multiple actors. Thus, the actor-based model of computation obviates the need for locking mechanisms commonly used to synchronize state among interdependent objects. Consequently, reasoning about agent behavior using actors can allow researcher developers to focus on implementing novel models and applications rather than debugging threads and locks [112]

[8]As in MATSim, the highest scoring BEAM plans are more likely to be re-evaluated and possibly selected for mutation. However, in contrast to MATSim, plans selected for mutation are cleared of all choices and re-evaluated within the context of the current iteration's transient state. Empirically, we find that this approach reduces the number of iterations needed to reach equilibrium.

constrained by simulation outcomes and design constraints, *i.e.*,

$$\begin{cases} \vec{x} = B(\vec{d}; \vec{z}), \\ g(\vec{d}; \vec{z}) = 0, \end{cases} \tag{4.2.2}$$

respectively, where the objective, f, is defined as the expected value of a stochastic performance measurement function, F. The deterministic decision vector, $\vec{d}$, is chosen from a search space $\mathcal{D}$, which may be continuous, categorical, combinatorial, or conditional. In the BISTRO context, the decision variables, $\vec{d}$, are the user-defined inputs that control policy levers within the transportation system. For example, $\vec{d}$ may specify the fare or vehicle types for specific public transit routes. The exogenous variables, $\vec{z}$, are the configuration inputs that determine the parameters of the population synthesis, the parameters of the transportation network, and the parameters governing supply of transportation services. The endogenous variables, $\vec{x}$, are the outcomes of the simulation run using $\vec{d}$ and $\vec{z}$ as input. The vector $\vec{x}$ contains the details of agent and vehicle movements throughout the simulation run, such as mode choices, travel times, travel costs, and vehicle path traversals, that were realized during the simulation run, *i.e.*, $\vec{x} = B(\vec{d}; \vec{z})$, where B represents the BEAM simulator. It is assumed that the iterative simulation process described in Section 4.2 has achieved stationarity.[9]

An important goal of BISTRO is to define objective functions that guide algorithms towards a range of solutions that represent *interpretable* and implementable policy decisions. A critical safeguard against unrealistic outcomes is implemented in BISTRO by translating business rules about inputs into mathematical constraint functions, g, parameterized by the decision vector, $\hat{d}$. Finally, F, is computed as a convex combination of the score components that guide solutions towards the *system objective* (as defined in Section 4.2).[10] The score components are evaluated from key performance indicators (KPIs) of the system performance, which are calculated using the simulation outputs, and relevant inputs. Following [115], we approximate the objective as the sample average of r independent realizations of F:

$$\hat{f}(\vec{d}, \vec{x}; \hat{z}) = \frac{1}{r} \sum_{i=1}^{r} F_i(\vec{d}, \vec{x}; \hat{z}) \tag{4.2.3}$$

The state space dynamics that govern the simulation-based optimization of an urban transportation system are highly complex and nonlinear, potentially containing several local

[9] Individual optimization algorithms may relax this constraint in order to reduce compute time while potentially trading off reduced accuracy or increased stochasticity of simulation output statistics.

[10] Due to variable amounts of nondeterminism and stochasticity inherent in ABMS, given fixed $\vec{d}$ and $\vec{z}$, the distribution of f can be approximated using n realizations of F as $\hat{f}\left(\vec{d}, \vec{x}; \vec{z}\right) = \frac{1}{n} \sum_{i=1}^{n} F_i\left(\vec{d}_i, \vec{x}_i; \vec{z}\right)$. In practice, optimization usually proceeds with $n = 1$ in order to identify promising (i.e., close to optimal) subsets of $\mathcal{D}$; however, when reporting final scores, one must carefully select n such that variability in output values is adequately captured.

minima. The lack of closed-form solutions to this class of optimization problem together with the computational expense associated with evaluating a single decision point make this class of problems particularly difficult to solve. In the next subsection, we describe several possible approaches to address this complexity.

Optimizing complex simulated systems: challenges and approaches

As described in Section 4.2, agent-based micro- or meso-scale simulations of transportation systems model the interdependent choices of rational individuals as they navigate virtual representations of physical and human geographies. Calibrating such models to high-resolution GPS traces and other sensor data embedded in infrastructure makes them highly suitable for evaluating the outcomes of location-specific policy alternatives. The trade-offs in accounting for the heterogeneous preferences of millions of agents, are that 1) the simulation model is expensive to evaluate for different settings of $\vec{d}$, and 2) the complex relationship between network dynamics and agent behavior lead to stochastic, non-convex specifications of performance measure, F. Consequently, the efficient gradient-based methods used to optimize closed-form relaxations of mobility dynamics as well as data-driven models derived from historical movement patterns do not apply [115]. Instead, generalized *stochastic optimization* (SO) algorithms treat the simulator as a black box. Commonly used derivative-free SO approaches include grid search, random search, ranking and selection, metaheuristic, and *metamodeling* techniques [116, 117].

Metamodeling algorithms encompass a broad class of simulation-based optimization approaches. These approximate F using a *surrogate model*, $\mathcal{Q}$ that is less costly to evaluate. Flexible and computationally tractable representations such as polynomial splines are able to approximate any objective function; however, many simulation runs are still required to accurately fit the response surface of the underlying system [116, 117].

Sequential model-based optimization (SMBO) is a general metamodeling formalism that, given a history of previous evaluations, $\mathcal{H} = \left\{\left(\vec{d_1}, y_1\right), \ldots, \left(\vec{d_i}, y_i\right)\right\}$, of observations $y_i = F\left(\vec{d_i}, \vec{x_i}; \vec{z}\right)$ at sample points in $\mathcal{D}$, selects the optimal next point $\vec{d}_{i+1}$ based on an approximation of F. To initialize SMBO, a small set of samples, $\left\{\vec{d_1}, \ldots, \vec{d_i}\right\}$ from $\mathcal{D}$ are selected using various experimental design techniques (e.g., random or Latin hypercube sampling). For each $\vec{d_i}$, evaluations of the expensive objective function, F form an observation, which, together with $\vec{d_i}$ are appended to a historical dataset $\mathcal{H}$. Once $\mathcal{H}$ is initialized, SMBO then proceeds iteratively: First, a regression model, $\mathcal{Q}$, is fitted to the current dataset, $\mathcal{H}$, yielding a surrogate model for F at the current iteration, which may be denoted $\mathcal{Q}_i$. Based on Q_i, the next input, $\vec{d}_{i+1}$ to F is selected by optimizing an *acquisition function*, $\alpha : \mathcal{D} \mapsto \mathbb{R}$ over $\mathcal{D}$, which measures the utility gained from evaluating F at $\vec{d}_{i+1}$. Following evaluation of $F\left(\vec{d}_{i+1}, \vec{x}_i; \vec{z}\right)$, $\mathcal{H}$, is updated as $\mathcal{H} = \mathcal{H} \cup \left(\vec{d_i}, y_i\right)$. The SMBO process continues until a predetermined time or computation budget is exhausted.

SMBO techniques are typically distinguished by the forms of the surrogate model, Q, and the acquisition function, α. In Bayesian optimization (BayesOpt), a Gaussian Process (GP, [118]) is typically used to model a prior over Q, which, at each iteration, is updated using previously observed data $\mathcal{H}$ to give a posterior predictive distribution $p\left(y \mid \vec{d}, \mathcal{H}\right)$ [119, 120]. Several methods using GPs as surrogate models may be distinguished according to the form of the covariance kernel parameterizing the GP [119, 121]. In lieu of GPs, BayesOpt algorithms have also used random forests [122] and *tree-based Parzen estimators* (TPE) [123, 124] as priors over Q. Acquisition functions are chosen to balance *exploration* and *exploitation* in the sample domain. The most common acquisition function used by these methods is based on an expected improvement criterion [125]; however, newer methods use variations on knowledge gradients [126, 127].

Many SMBO algorithms can run trials in parallel, which may yield reduced wall clock time (although the total number of trials required may be identical to that used by sequential implementations) [128]. One method to further reduce the running time of SMBO *trials* when model evaluations require an inner iterative loop to achieve stationarity (as in BISTRO) is to incorporate an early stopping rule for simulation evaluations that are likely to eventually be extremely suboptimal. An example of such an approach is "freeze-thaw Bayesian Optimization" [129].

Recent efforts in transportation science and operations research have also sought to develop tractable simplifications of scenario-based optimization of simulated urban transportation systems. One vein of research concentrates on deriving deterministic analytic equations describing system dynamics at equilibrium from static information, $\vec{z}$ (*e.g.*, network topology and bus schedules) to inform purely functional metamodels [130, 131, 132]. For example, [115] combine a computationally tractable model of congested traffic based on queuing theory with a detailed local approximation using a linear combination of basis functions from a parametric family. Alternatively–and analogously to the "freeze-thaw" Bayesian optimization setting highlighted above–some approaches use information about the *process* by which the stochastic simulation achieves stationarity to develop techniques that rapidly evaluate different settings of the decision variable vector, $\vec{d}$, while avoiding the need to reach convergence [133, 134].

The present work refrains from prescribing a single best approach to solve optimization program Equations (4.2.1) and (4.2.2). Instead, the intent of BISTRO is to enable *replicable* future research in this area by providing a platform and problem setting that is generalizable across different planning contexts as well as approachable and of research interest to the ML/AI community. Problem characteristics such as the propensity for competing metrics to be present in system objectives, preemptive stopping of inner optimization loops, the high dimensionality of the search space, and the potential for hybridization of functional and physical metamodels are expected to provide challenging, scalable, and, critically, *explainable* solution approaches. Algorithms combining data-driven dimensionality reduction techniques as well as efficient experiment design can result in repeatable protocols to effectively constrain more general local and global search techniques. Following a presentation of the framework

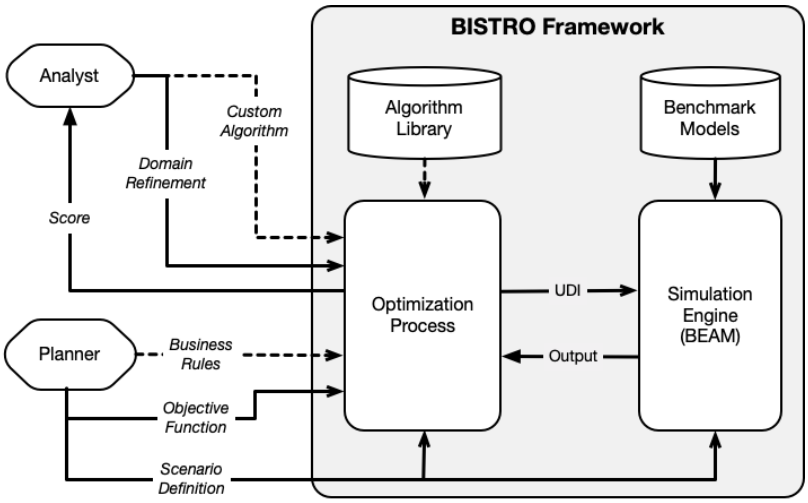


Figure 4.2: BISTRO software architecture, illustrating how the optimization process modulates the flow of information between the BEAM simulation as well the two primary user types. The distinction between the *planner* and the *analyst* is critical in that we do not expect the analyst (an expert in applied ML/AI-based optimization methods) to have transportation or planning background, yet still they should be able to develop generalizable algorithms that can be used to optimize transportation system objectives set by the planning organization.

architecture, in Section 4.4, we empirically explore the effectiveness of some of these solution strategies as well as the interpretability of their outputs.

4.3 Berkeley Integrated System for TRansportation Optimization (BISTRO)

BISTRO is a new analysis and evaluation platform that works in concert with an ABMS (BEAM) to enable the open-sourced development and evaluation of transportation optimization methods in response to given policy priorities. This section gives an in-depth description of the BISTRO framework and all of its major components, providing an overview of their purpose, use, and functionality as well as calling attention to the most novel aspects of its design.

System Architecture

As indicated in Section 4.2, BISTRO implements elements of the travel demand planning process coupled with components of an automated simulation-based optimization system. This section describes the high-level overall architecture of BISTRO (depicted in Figure 4.2),

focusing on the conceptual distinction between features relevant to scenario developers and those more appropriate to algorithm designers.

A BISTRO run environment is configured using a set of fixed input data defining the required transportation system supply elements (*e.g.*, road network, transit schedule, on-demand ride fleet) and demand elements (*e.g.*, synthetic population, activity plans, and mode choice function parameters). Precisely which aspects of the virtual transportation system should be represented in the simulation model depends on the strategic goals and system objectives defined as part of the planning and analysis process motivating a particular BISTRO use case. An example of the set of raw inputs and pre-processing steps is illustrated in Figure 4.3.

A boundary separates external, exogenously defined inputs from the BISTRO simulation optimization pipeline. Outside of the boundary, the *user-defined inputs* (UDIs) represent the investment, incentive, and policy levers applicable to and available for the study at hand. Concretely, algorithm developers encode solutions as numeric values that represent vector-valued variables controlling aspects of the initialization and evolution of the simulation. For example, a UDI that alters the frequency of buses on a route must specify a target transit agency, a route, a start time, an end time, and the desired headway.

While BISTRO maintains a library of available interventions compatible with BEAM, scenario designers, policy makers, and other stakeholders will often want assurance that infeasible, regressive, or otherwise undesirable input combinations are prevented from being selected as "optimal." Together with syntactic and schematic validation of inputs, flexibly-defined *business rules* can effectively act as constraints on the search space—enhancing the interpretability and, thereby, the rhetorical and communicative value of BISTRO-derived solutions.

Just as UDIs from previously conducted BISTRO-based studies are actively maintained and made available to scenario designers, the BISTRO community contributes to a growing library of recommended optimization algorithms that, when evaluated across multiple BISTRO benchmark scenarios, demonstrate desirable performance characteristics. Thus, users lacking resources or expertise to develop optimization routines in-house can still benefit from what, we anticipate, will be cutting-edge research on algorithms and strategies to optimize the simulation of demand-responsive cyberphysical infrastructure.

Project owners of BISTRO deployments may work with stakeholders to develop representative models that will be used to benchmark optimization algorithms. Enabling a well-defined benchmark mechanism permits data on the performance of user-supplied algorithms to be compared. These comparisons can be used to assist in identification of design patterns and computational strategies that advance the state of the art in simulation-based optimization of urban transportation systems.

Scoring Function Design

Transportation system intervention alternatives are scored in BISTRO based on a function of score components evaluated using *key performance indicators* (KPIs) of the simulation. KPIs

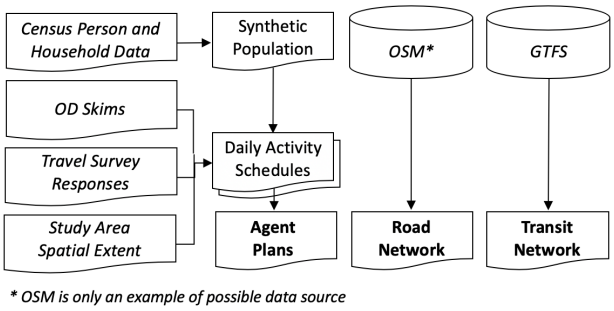


Figure 4.3: Generation of fixed inputs. Italicized entities represent the necessary data to generate fixed inputs (in bold).

for a given simulation should be selected in accordance with the operational, environmental, and social goals, or, *system objectives* developed as part of the participatory planning process described in Section 4.2. BISTRO project planners may select KPIs to include in the scoring function from an existing library of options, or may choose to develop additional KPIs, as appropriate, for the goals and system objectives of the project. Additionally, the form of the scoring function may be designed by the analyst in consultation with the project planner.

Key Performance Indicators

KPI overview There are two general types of KPIs developed in BISTRO: 1) KPIs that measure the operational efficiency of the transportation system (*e.g.*, *vehicle miles traveled* [VMT], vehicle delay, operational costs, revenues) and 2) KPIs that evaluate the experience of transportation system users (*e.g.*, generalized travel expenditure, bus crowding experienced, accessibility). KPIs can be aggregated or disaggregated into score components to support particular policy objectives. For example, the accessibility KPI (detailed below) may be disaggregated by activity type, time period, mode used, and/or sociodemographics in order to evaluate the distributional equity of access provided across different opportunities at varying times of day and/or across population segments of concern.

In practice, any KPI that may be evaluated from the set of output variables (see Section 4.3) produced by a BISTRO simulation run may be included as a score component in the scoring function. However, careful consideration of candidate KPIs must include an evaluation of the sensitivity of the metric to the UDIs of interest as well as the efficiency of the KPI in providing the desired feedback regarding the optimality of outcomes of alternative UDI values. For example, *person miles traveled* (PMT) is a commonly used metric in transportation system performance measurement to gauge the amount of mobility delivered by the system. Yet, PMT is highly invariant within a scenario in BISTRO due to the fact that agent plans are fixed. Thus agents will make the same trips regardless of the UDI values and the miles traveled by each agent will only vary in so much as the networks available for each mode offer more or less direct paths to travel from the origin to destination of each

trip.[11]

Implemented KPIs The following items represent categories of KPIs that have been developed and implemented in BISTRO at the time of publication of this dissertation:

1. *Accessibility.* In an urban transportation planning setting, accessibility has often been defined as a measure of the ease and feasibility with which opportunities or points of interest can be reached via available modes of travel. It is quantified in BISTRO as the sum of the average number of points of interest (of a specific type of activity) reachable within a given duration of time, with functionality also provided to measure mode-specific accessibility.

2. *Trip Expenditure and Generalized Transportation Cost Burden.* The socio-demographic and spatial heterogeneity of travel behavior within BISTRO enables a variety of equity-focused impact analyses. Two such metrics have been implemented in BISTRO: average trip expenditure and average generalized transportation cost burden. While the former is the average monetary cost incurred by agents per trip, the latter is computed as the sum of the travel expenditures of the trip (costs of fuel and fares minus incentives, as applicable) and the monetary value of the duration of the trip (the product of the total trip duration and the population *average value of time* (VOT)), divided by the household income of the agent completing the trip. the monetary value of the trip duration is calculated by multiplying total duration by the population average VOT. Both KPIs may be disaggregated to emphasize particular equity goals (e.g., by socio-demographic groups, trip purpose, mode, etc.).

3. *Bus Crowding.* The *level of service* (LoS) experienced by public transit passengers has a direct influence on short- and long-term demand for public transit service. In addition to cost and travel time factors, the available capacity on a transit vehicle affects whether or not a passenger can board a public transit vehicle at their desired time as well as the level of comfort they experience during the trip. Though the LoS of public transit may be measured in BISTRO by any one of the factors mentioned, BISTRO includes a ready-made example of an LoS KPI related to passenger comfort: average bus crowding experienced. This metric is computed as the average over all transit legs of the total passenger-hours weighted by VOT multipliers corresponding to the load factor (the ratio of total passengers to the seating capacity) of the bus during the leg.

4. *Vehicle Miles Traveled (VMT) and Delay.* The BISTRO KPI library includes three examples of congestion score components that provide insight into the destination- or opportunity-independent level of mobility on a network, the overall network performance, and efficiency: total VMT by all motorized vehicles in the transportation

[11]For example, a transit mode choice for a particular trip may result in more PMT than a walk mode choice for the same trip, as the sidewalk network may enable a more direct path.

system, total vehicle delay, and average vehicle delay experienced per passenger trip. Total vehicle delay is calculated as the sum over all path traversals of the difference between the realized duration and the free flow travel time of the traversal. Vehicle delay experienced per passenger trip is calculated as the total difference between the realized duration and free flow travel time of all legs of a trip completed by modes subject to congestion.

5. *Financial sustainability.* Most system interventions will have some impact on the flow of funds in or out of the transportation system. Including a KPI that helps stakeholders understand the general financial impacts of such interventions is necessary. The financial sustainability metric provided in the BISTRO KPI library is the sum of all public transit fares collected minus all incentives distributed (if any) and all operational costs of the public transit system [12]. In the event that a BISTRO project does not alter public transit service, the operational costs may be omitted from the KPI, if desired.

6. *Environmental sustainability.* The environmental sustainability of a transportation system intervention may be measured as the local and/or global impacts to the system. In addition to VMT- and fuel efficiency-based estimates, BEAM enables estimation of emissions directly from the simulated fuel consumption, based on the realized speeds traveled by each vehicle throughout a simulation run[13]. The VMT-based fine particulate emissions ($PM_{2.5}$) KPI captures local environmental sustainability via a mileage-based measure of air quality impacts based upon vehicle type. Additional local emissions KPIs may easily be included using the appropriate emissions factors[14]. A *greenhouse gas* (GHG) emissions KPI allows the optimization to explicitly account for fuel-consumption-based global environmental sustainability [15].

Scoring Function

The BISTRO scoring function serves as the objective function by which the UDIs are optimized. The selection and/or definition of the objective function in accordance with the

[12]The operational costs include the total costs of fuel consumed, and hourly variable costs of bus operations (see Table 4.3 for an example of operational costs). Hourly variable costs include estimated labor, maintenance and operational costs. The rates for each of these factors is specified in the vehicle fleet configuration variables.

[13]For more information on the methodology followed to estimate fuel consumption, please refer to the BEAM documentation `https://beam.readthedocs.io/en/latest/index.html`.

[14]For more information on the methodology followed to develop this metric, please refer to the California Air Resources Board documentation, `https://www.arb.ca.gov/cc/capandtrade/auctionproceeds/cci_emissionfactordatabase_documentation.pdf`.

[15]It is important to note that the GHG emissions KPI will be correlated with VMT and fine particulate emissions. Thus, inclusion of all three KPIs creates a suite of environmental sustainability metrics that may apply disproportionate weight on environmentally-related objectives, which may or may not be desirable for certain policy agendas. Project planners may choose to apply scaling factors (as described in Section 4.3) to balance the influence of the environmental sustainability score components.

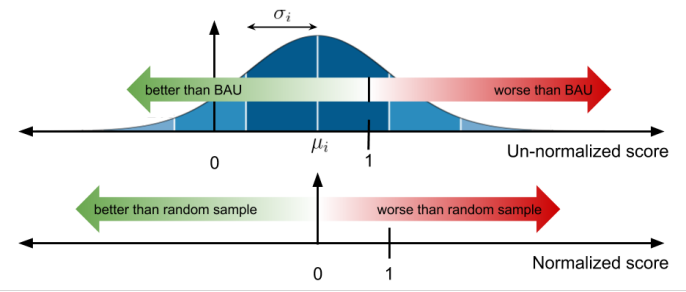


Figure 4.4: A visual representation of the normalization procedure for a hypothetical score component, i. The ratio of the submission score and BAU score (depicted in the upper plot) is normalized by taking its z-score (depicted in the lower plot) relative to a random input sample.

project directives is considered to be the responsibility of the project planner. Herein, a general structure is defined to facilitate the creation of custom objective functions. Multiple project objectives (referred to here as *score components*) may be included in the scoring function–either as individual elements within a vector of scalar-valued score components to be minimized, or as parameters to a function that aggregates the objectives into a one-dimensional scalar score. The score components are computed as the normalized ratio of the value of the corresponding KPI in the given simulation run to the value of the same KPI in the *business-as-usual* (BAU) run[16]. The improvement ratios are normalized using KPI values produced by a randomized sample of the UDI space, the size of which can be defined by the BISTRO project owner. This normalization (depicted graphically in Figure 4.4) accounts for differences in variance across KPIs, thus allowing the score components to provide meaningful feedback on the improvement achieved for each KPI relative to the distribution of the ratios of KPI to BAU produced by the random search. The composite score is thus a function of the normalized relative improvements of the candidate input to the BAU in each metric, as follows:

$$F\left(\vec{C}_s, \vec{K}, \vec{\sigma}, \vec{\mu}, \vec{\alpha}\right) = f\left(\vec{z}, \vec{\alpha}\right), \tag{4.3.1}$$

where $\vec{K}$ is the vector of all KPIs evaluated for a given set of inputs, $\vec{C}_s$; $\vec{\mu}$ and $\vec{\sigma}$ are the vectors of normalization parameters; and $\vec{z}$ is a vector of each KPI's z-scores, *i.e.*,

$$z_i = \frac{\frac{K_i(C_s)}{K_i(C_{BAU})} - \mu_i}{\sigma_i}, \tag{4.3.2}$$

for the i-th KPI. The value of the i-th score component in the BAU case is simply $K_i(C_{BAU})$.

[16]In the BAU of a given scenario, the simulation is run without alteration from the initial configuration of that scenario.

The default objective is to minimize the composite score function, since an increase in many of the score components actually represents a scenario that is worse than the *status quo* (*e.g.*, decreasing VMT over BAU results in a lower unscaled score than increasing VMT). To maintain consistency in this regard, the scoring function may include an additional parameter $\vec{\alpha}$ to allow for transformation of score components that are positively related to desirable outcomes (*e.g.*, improvements in accessibility). For example, if the scoring function takes the form of a sum over all score components, the parameter $\vec{\alpha}$ may be used as a coefficient of each score component that determines whether the component will be summed or subtracted, as follows:

$$\alpha_i = \begin{cases} -1 \text{ if it is desirable for score component } i \text{ to increase} \\ 1 \text{ otherwise} \end{cases} \tag{4.3.3}$$

This approach is distinct from the one typically used for *Cost-Benefit Analysis* (CBA) tasks in urban planning practice, in that it seeks to optimize an aggregate function of the relative improvements in each KPI rather than optimizing the net improvement from all KPIs. While CBA often draws skepticism due to the discretion inherent in the process of converting all KPIs into a common unit such as time or money so that the net value of costs and benefits can be computed, the approach taken in the BISTRO scoring function does not require any such assumptions to be made. Rather, each score component represents the relative improvement over the BAU that is achieved by a simulation run using a particular set of UDIs. Objective function designers may choose to apply additional scaling factors to the score components using the $\vec{\alpha}$ parameter vector.

Inputs

Preparation of fixed inputs For each BISTRO study, a set of fixed inputs must be provided to BEAM. For a given study area, these typically include the road network, the transit schedule, and the demand profile. Depending on the system objectives, additional data may be necessary to fully configure the simulation. Figure 4.3 illustrates a schematic of the inputs for a typical simulation.

The *road network*, including the physical properties of its links and nodes, may be generated using *Open Street Maps* (OSM) data for the geography of interest. The *transit network* configuration follows the easily accessible *General Transit Feed Specification* (GTFS) format. *On-demand ride services* (*Transportation Network Companies* [TNCs] such as Uber, Lyft, and Via), are modeled as a fleet of vehicles driven by agents that are exogenous to the population, or may be driven autonomously. The initial locations of the vehicles may be sampled randomly or from a specified distribution in accordance with appropriate data. The price of on-demand rides is fixed, consisting of a distance-based and a duration-based component. The size of the on-demand ride service fleet is a proportion of the total number of agents in the simulation, as determined by a configuration parameter. Driver repositioning behavior when not currently driving to or serving a passenger can be configured to follow one of several repositioning algorithms defined within BEAM.

Table 4.1: Example of bus frequency adjustment input file.

route_id	start_time	end_time	headway_secs	exact_times
1,340	21,600	79,200	900	1
1,341	21,600	36,000	300	1
1,341	61,200	72,000	300	1

At the start of the simulation, a *synthetic population* of virtual agents and households is generated such that the sociodemographic attributes of these virtual entities are spatially distributed in accordance with real-world census and/or location-based data from the city of interest. Each agent follows a daily plan consisting of several activities throughout the day. As illustrated in Figure 4.3, these *daily activity schedules* are generated based on *origin-destination* (OD) skims (matrices that provides the number of trips between zones), travel surveys, and zonal boundary spatial data.

Calibration Prior to use in optimization runs, BEAM needs to be calibrated to empirical data by iteratively adjusting model parameters until a simulation outputs representing traffic patterns match their real-world counterparts with minimal error[17]. Calibration of BEAM for usage in BISTRO should be limited to adjustments of behavioral parameters controlling microscopic decision-making (e.g., mode choice intercepts, prices, regulation-driven incentives/tariffs)[18]. While it is possible to adjust many additional BEAM parameters to reduce calibration error, this practice should be discouraged, as it may result in models that are overfitted to a state of the world represented by a particular ground-truth dataset, thereby limiting the calibrated simulation model's use in predictive contexts. In addition to the representativeness of ground truth data, the quality and quantity of input data (*e.g.*, network data resolution or population spatial resolution), may influence the extent to which the model is able to achieve calibration end points.

Configuration of UDIs BISTRO provides a library of possible inputs for scenario designers to adapt to specific use cases. The selection of UDIs is intended to be compatible with the system objective. UDIs may represent, for example, the investment (*e.g.*, transit fleet mix modification, bus route modifications, parking supply, electric vehicle charge station locations, dynamic redistribution of e-bikes or on-demand vehicles), incentive (*e.g.*, incentives to specific socio-demographic groups for selected transportation modes, road pricing/toll roads,

[17]BEAM calibration is typically targeted at mode split, volumetric traffic counts, and travel distance distributions. The choice of which target(s) to use may depend on regulatory requirements, literature recommendations, or precedent [83, 135].

[18]Often, due to computational constraints, a sub-sample of a full population is simulated. The capacity of physical resources (e.g., road network link carrying capacity, maximum transit occupancy, number of electric vehicle charging plugs per station) may need to be adjusted based on the size of the population sample. For evaluation purposes, the outputs of a sub-sampled simulation are often scaled back up to the full population.

fuel tax), or policy/operational (*e.g.*, transit schedule adjustment, transit fare modification, parking pricing) levers applicable to the study at hand. The project owner may constrain the range of possible values upon which each UDI is valid by setting the corresponding input validation parameters and business rules. The example input file for bus scheduling shown in Table 4.1 defines alteration of the headway of a particular bus route during a particular service period (defined by its start and end times).

Output Analysis and Visualization

The *raw* outputs of a BEAM simulation include millions of events that reflect the microscopic actions of each agent as they make their way through the day. While this is a detailed history of what transpired during the day for each agent, it does not provide planners with explanatory insight into how perturbations of input variables influence model behavior leading to changes in output responses. To simplify exploration of alternatives to purely black box optimization methods, the BISTRO platform provides a suite of tools that process and organize event data from BEAM simulation runs into a relational format, permitting relationships between each person's activities, trips, path traversals to be queried at different levels of detail[19]. A Jupyter notebook then incorporates the post-processed outputs into a standard suite of multivariate analyses and visualizations, thereby facilitating interpretation and communication of the effect that policies have on system objectives.

Implementation Details and Performance Characteristics

Both BISTRO and BEAM are primarily implemented in Scala. Input files are read from a single directory and injected into the BEAM initialization routine. The system is containerized using Docker, which helps to facilitate OS-agnostic local and remote execution.

The runtime of BEAM depends on various inputs including population size, network resolution, transit network, ridehail fleet size as well as available compute resources such as number of processors, memory, etc. On a machine with 32 (2.5GHz Intel®Xeon®Platinum 8175) CPUs and 128GB RAM, the runtime for a 315,000 agent simulation of a San Francisco Bay Area scenario (representing a 25% sample of the approximate 2018 population) takes around 14 hours to complete 15 iterations. On a machine with identical compute resources, a 15,000 agent simulation for the the Sioux Faux scenario (see Section 4.4) takes approximately 30 minutes to complete 15 iterations.

Currently, the primary performance bottleneck in BEAM is routing. The routing engine generates millions of routes (reflecting multimodal options for agents to choose between) for a single simulation run. Some additional overhead considerations such as data availability, level of model resolution required, as well as the impact of augmented BEAM functionality must be balanced in light of available computational resources.

[19]Further details on the output of the parser including an entity relationship diagram are available at `http://bistro.its.berkeley.edu/assets/download/pdfs/General_System_Specification.pdf`

4.4 Initial Pilot Study and Launch

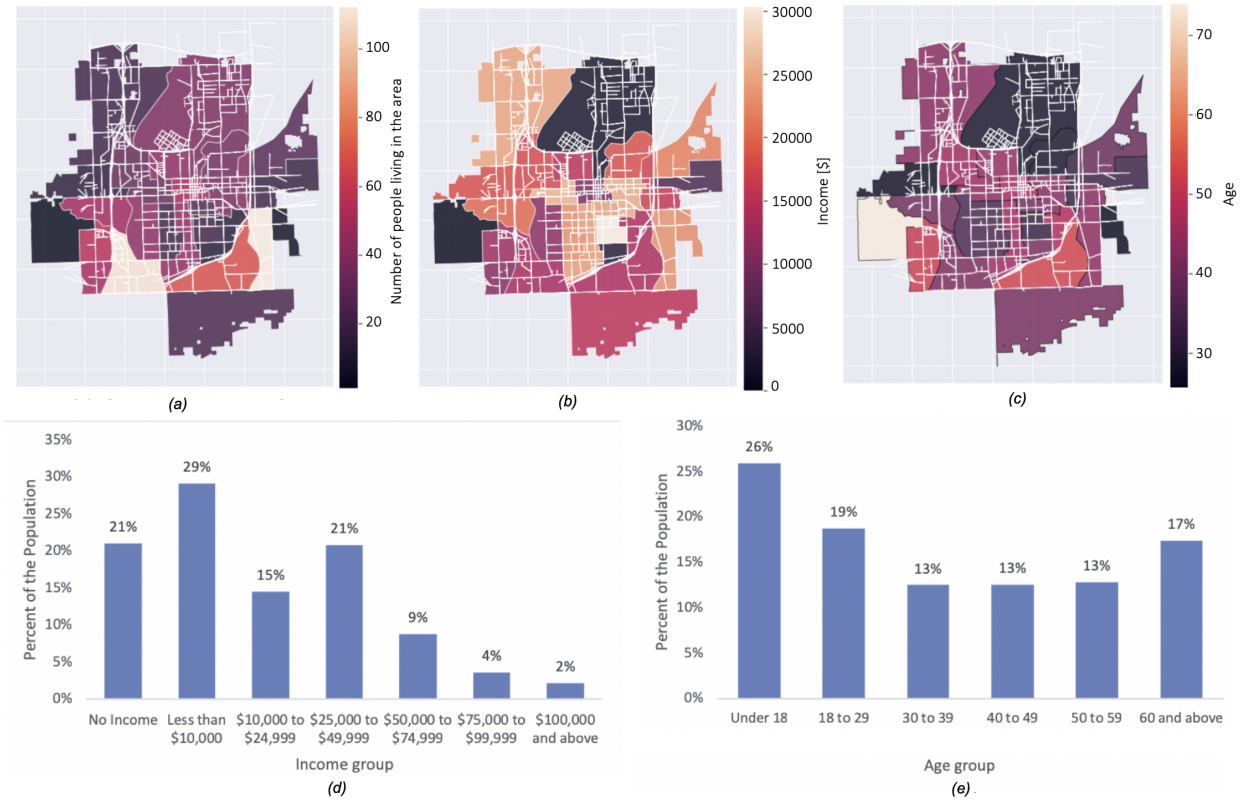


Figure 4.5: Demographics of Sioux Faux. (a) Overall distribution of the population per census tract. (b) Distribution of the median population income per census tract. (c) Distribution of the median population age per census tract. (d) Overall population income distribution. (e) Overall population age distribution.

Sioux Faux

An agent-based model of transport supply and demand inspired by the real city of Sioux Falls, South Dakota[20] was adapted for the purpose of developing and testing example scenarios within BISTRO. To underscore that for these purposes, such scenarios were not developed to be true replicas of the city of Sioux Falls, this benchmark BISTRO scenario is referred to as *Sioux Faux*. The scenario configuration, input specification, and scoring function were designed to support strategic objectives of financial and environmental sustainability, reduced congestion, and improved equity, accessibility, and transportation system level of service.

[20]The "Sioux Falls" scenario is a commonly used benchmark in ABMS research, see `https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls`

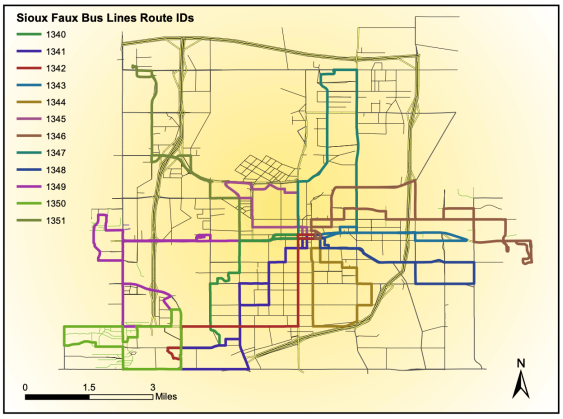


Figure 4.6: Sioux Faux bus and road networks.

Scenario Configuration

Population and plan synthesis The synthetic population of Sioux Faux was generated using publicly-available census data for the city of Sioux Falls, South Dakota as inputs to the Doppelganger library[21], a state-of-the-art population synthesis framework developed in Python. Specific inputs to Doppelganger used to generate the Sioux Faux population included household and individual *Public Use Microdata Sample* (PUMS) data for South Dakota from the 2012-2016 (5-year) *American Community Survey* (ACS), which is conducted annually by the US Census[22]. The *Public Use Microdata Area* (PUMA) for Sioux Falls constrains the state-wide survey data to our general area of interest. Population demographics derived from the synthetic population for Sioux Falls are shown in Figure 4.5.

An existing set of agent plans for Sioux Falls previously developed for MATSim simulations was used as the basis for the plans of our expanded Sioux Faux population[23]. After initial pilot testing to determine trade-offs between population size, behavioral realism, and computational complexity, we took a 15% sub-sample of the full synthetic population (approximately 15,000 agents). We used a spatially-constrained sampling mechanism in order to allocate plans to agents in accordance with household locations and census tract household and individual attribute distributions. The subsampling mechanism also enforces realistic constraints on agent plans and behavior using predicates such as "agents under the age of

[21]Doppelganger uses a novel Bayesian optimization approach combined with the hierarchical list-balancing algorithm developed as part of the PopSyn library [136]. For more information about the Doppelganger library, see `https://github.com/sidewalklabs/doppelganger`

[22]The 5-year PUMS comprises a 5% sample of the US population. It is computed as an aggregate of 1-year samples, which themselves aim to survey 1% of the US population

[23]More specifically, we modified the Sioux Falls 2016 scenario developed by Hörl ([85]), which is an update of a scenario prepared in 2014 by Chakirov and Fourie [137]. For more information on the Sioux Falls scenario, see `https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/901-1000/ab978.pdf`

18 should not have a work activity" and "agents under the age of 16 should not be allowed to drive".

Transportation Network The Sioux Faux transportation network includes a road network accessible to walking agents, personal vehicles, on-demand ride services (TNCs such as Uber and Lyft), and public buses providing fixed-route service[24]. The on-demand ride services implemented in this scenario include only single-passenger rides (*e.g.*, UberX, Lyft Classic) from a fleet of on-demand ride vehicles that was distributed randomly across the road network at the start of each simulation run. Driving alone is the most frequently used mode, comprising approximately 75% of the miles traveled for the BAU scenario.

User-Defined Input Specification

For this initial pilot study, a set of four UDIs were investigated: 1) bus fleet vehicle composition, 2) bus service frequency, 3) bus fare, and 4) multimodal incentive program for on-demand rides and public bus trips. In the bus fleet for the BAU scenario, all vehicles were set to a default bus type. Optimization of the bus fleet vehicle composition and service frequency offers the opportunity to improve the level of bus service by better matching the bus type with specific demand characteristics of each route. Four types of buses (including the default) were considered (see Table 4.3), each with different technical properties (seating and standing capacity) and cost characteristics (cost per hour, cost per mile, fuel type and fuel consumption rate).

A UDI was implemented to vary the bus schedule on each route, including the hours of service and the headway, or service frequency as shown in Table 4.1. Multiple service periods with varying headways on the same route were thus possible. The bus fare UDI allowed for the optimization of the fare on each route, segmented by passenger age groups. Finally, a multimodal incentive UDI was implemented to enable reimbursement for on-demand rides, walk to/from transit, or drive to/from transit trips to qualifying individuals based on age, income, or both.

Business Rules

In order to ensure that optimal solutions would be compliant with common policy and planning practices, four business rules were implemented: 1) there may be no more than five distinct bus service periods (this mimics a typical delineation of transit service provision: am peak, midday, pm peak, evening, late night/early morning), 2) bus route headways may be no more than 120 minutes and no fewer than 3 minutes, 3) bus fares and mode incentives may not isolate a single age, and 4) ages for both fares and incentives may be specified in segments no smaller than five years in range and income for incentives may be assigned in segments no smaller than $5,000 in range.

[24]The initial bus route scheduling is directly generated from the publicly available GTFS for Sioux Falls, which includes erratic headways across routes.

Table 4.2: Values used for α_i in each of the subsequent results sections. For score components that are positively related to desirable outcomes, negative α_i is provided to transform it consistent with a minimization problem.

KPI	KPI type	Contest	Post-Contest	New KPIs
accessible work locations	Accessibility	-1	-1	–
accessible secondary locations	Accessibility	-1	-1	–
accessible work locations by car	Accessibility	–	–	-1
accessible secondary locations by car	Accessibility	–	–	-1
accessible work locations by transit	Accessibility	–	–	-1
accessible secondary locations by transit	Accessibility	–	–	-1
average trip expenditure-work	LoS	1	1	–
average trip expenditure-secondary	LoS	1	1	–
average travel cost burden-work	Equity	–	–	1
average travel cost burden-secondary	Equity	–	–	1
average bus crowding experienced	LoS	1	1	1
total vehicle miles traveled	Congestion	1	1	1
average vehicle delay per passenger trip	Congestion	1	1	1
costs and benefits	Financial Sustainability	-1	-1*	-1*
total grams $PM_{2.5}$ emitted	Environmental Sustainability	1	1	1
total grams GHG_e emitted	Environmental Sustainability	–	–	1

*fixed KPI post-contest

Table 4.3: Transit vehicle types available for Sioux Faux bus fleet: (a) Fuel type, (b) Fuel consumption rate (J/m), (c) Operational cost (USD/hr), (d) Seating capacity, (e) Standing capacity.

Vehicle type, $c \in C$	a	b	c	d	e
BUS-DEFAULT	diesel	20048	89.88	37	20
BUS-SMALL-HD	diesel	18043.2	90.18	27	10
BUS-STD-HD	diesel	20048	90.18	35	20
BUS-STD-ART	diesel	26663.84	97.26	54	25

Scoring Function Design

The set of Sioux Faux UDIs have varying interconnected impacts on the operation of and access to public transit and on-demand ride service by agents. Thus, the scoring function upon which the inputs were optimized was designed to include a variety of metrics that relay feedback on the user experience and operational efficiency of the transportation system as a whole. Table 4.2 presents all of the KPIs used in Sioux Faux scenarios referenced in this text[25].

Five KPIs were developed to represent three main aspects of user experience: accessibility, travel expenditure, and transit passenger comfort. The accessibility and travel expenditure were both disaggregated by trip purpose such that score components for accessibility and travel expenditure to work and secondary activities were each included separately in the scoring function. Transit passenger comfort was measured as the average bus crowding experienced by bus passengers[26].

Four KPIs of operational efficiency were included to account for the congestion, environmental sustainability, and financial sustainability resulting from optimized inputs. Total VMT was included as a KPI for overall congestion while average vehicle delay per passenger trip served as a KPI of the average impact of congestion. The total amount of $PM_{2.5}$ emitted served as a KPI of the environmental impact resulting from each simulation run. Finally, the financial sustainability KPI was included to incentivize outcomes with minimal impact to the bottom line of the transit agency by taking into account the operational costs, incentives distributed, and revenues collected from any combination of transit fleet mix, scheduling, fare structure and incentive program.

All metrics are aggregated according to the following function:

$$F\left(\vec{C}_a, \vec{F}, \vec{\sigma}, \vec{\mu}, \vec{\alpha}\right) = \sum_{i \in \mathbb{K}} \frac{\left(\frac{K_i(C_s)}{K_i(C_{BAU})}\right)^{\alpha_i} - \mu_i}{\sigma_i} \tag{4.4.1}$$

where all variables are defined as described in Section 4.3, with the set of KPIs and corresponding α_i values as specified in Table 4.2. We executed 800 runs using randomly generated values of $\hat{d}$ to produce the normalizing statistics (i.e., μ_is and σ_is in Equation (4.3.2)) for each metric.

Pilot study results

Contest participation and results Over the course of 17 days, 487 people in teams of one to four (mostly consisting of engineers and data scientists with little to no domain expertise in transportation planning) effectively created nearly 1,000 different "city transportation

[25]Note that several of the KPIs in this table refer to two post-contest follow-on studies, see Section 4.4

[26]Average bus crowding in the Sioux Faux scenario was calculated as the average number of agent hours spent per transit trip in buses occupied above their seating capacity. This KPI has since been updated, see Section 4.3

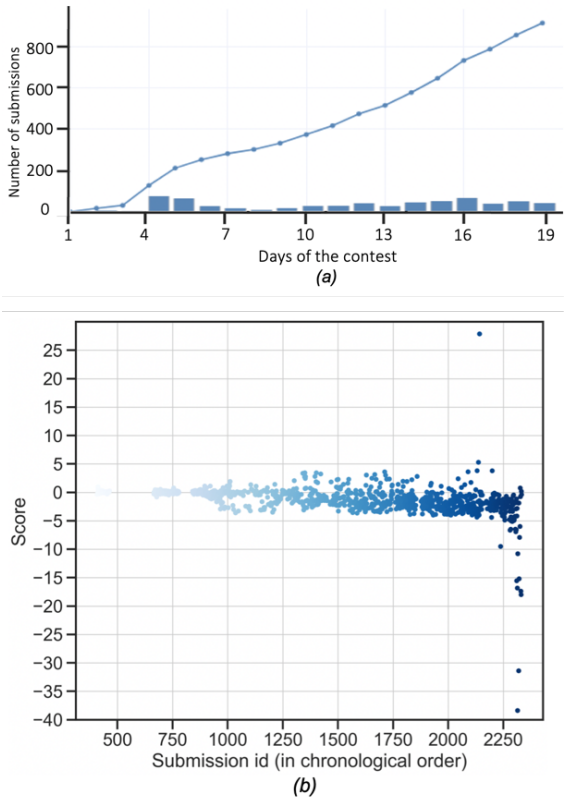


Figure 4.7: Participation history. (a) Number of solutions submitted over time (per day and cumulative). (b) Evolution of scores over time (see equation Equation (4.4.1)).

plans" for the Sioux Faux scenario consisting of the UDIs described in section Section 4.4 [27]. To be able to compare their results and scores with other participants, each team could submit up to five solutions per day and thus be ranked in a web-accessible leaderboard. While contestants trained algorithms online, final evaluation, leaderboard, and discussion boards were hosted by AICrowd.com[28]. Inputs from top teams were evaluated 5 times for 100 iterations each in order to achieve a consistent final score.

Figure 4.7 illustrates the evolution of submissions over time during the competition. Participation developed in two phases. During the first week, contestants became familiar with the BISTRO framework and the Sioux Faux transportation optimization problem. During the second phase, contestants continued to optimize their solutions. According to code submissions and a post-contest survey, the solutions that achieved the highest value of the objective function (averaged over 5 replications) followed similar strategies. Typically, they used domain-specific analysis to prune the large input space. For example, by sampling the

[27] Uber does not endorse any of the solutions presented.

[28] `http://www.aicrowd.com`

Table 4.4: Proportion of algorithmic approaches used, according to a survey conducted with contestant teams.

Approach	Proportion
Bayesian optimization	34%
Evolutionary algorithms	28%
Gradient based	14%
Meta-heuristics	7%
Plackett-Burman design	3%
Hill climbing	3%
Other	10%

age distribution within walking distance from a bus route's stops, a reasonable bound on the components of the fare UDI could be assigned. Following these informed factor screening steps, contestants used a variety of algorithmic approaches to more efficiently search the lower dimensional design space. As shown in Table 4.4, the black-box global optimization techniques used during the Contest primarily incorporated variants of Bayesian optimization, genetic/evolutionary algorithms, gradient-based techniques, and meta-heuristics methods.

Most teams managed to improve their scores by three standard deviations better than a random search benchmark (*i.e.*, with scores of approximately -3). Due to an insidious modeling deficiency, the financial sustainability score component could be optimized towards negative infinity. As such, any contributions from other score components would be relatively inconsequential. Two teams discovered input settings that took advantage of the lack of a lower bound on the financial sustainability score component and were thus able to reach extremely low scores of -30 or -40. This experience underlines the importance of developing careful theory supplemented by judicious testing when designing objective functions.

Post-Contest and New KPIs follow-on studies and results After the contest, two follow-on studies were conducted to interpret the solutions from the top algorithms in the context of improvements to the objective function. An initial set of improvements, hereby referred to in this text as the "Post-Contest" objective function study, simply addressed the unbounded financial sustainability score component as well as other minor problems discovered during the competition. The "New KPIs" objective refers to an expanded set of KPIs, summarized in Table 4.2. Two of the best-performing algorithms from the Contest—namely, Bayesian Optimization using *tree-based Parzen estimators* (TPE) [123] and *Genetic Algorithms* (GA) [138]—were adapted and re-implemented to run BISTRO on the Sioux Faux scenario with both of the updated objective functions. As a baseline algorithmic benchmark, *random search* (RS) was performed for 800 trials using both objective functions. The GA assessment on both the "Post-Contest" and "New KPIs" objectives utilized five parallel evolutionary trajectories, each drawing a random sample of seeds from a larger gene pool. Both TPE and GA were run over identical design spaces for 1,400 trials on the "Post-Contest"

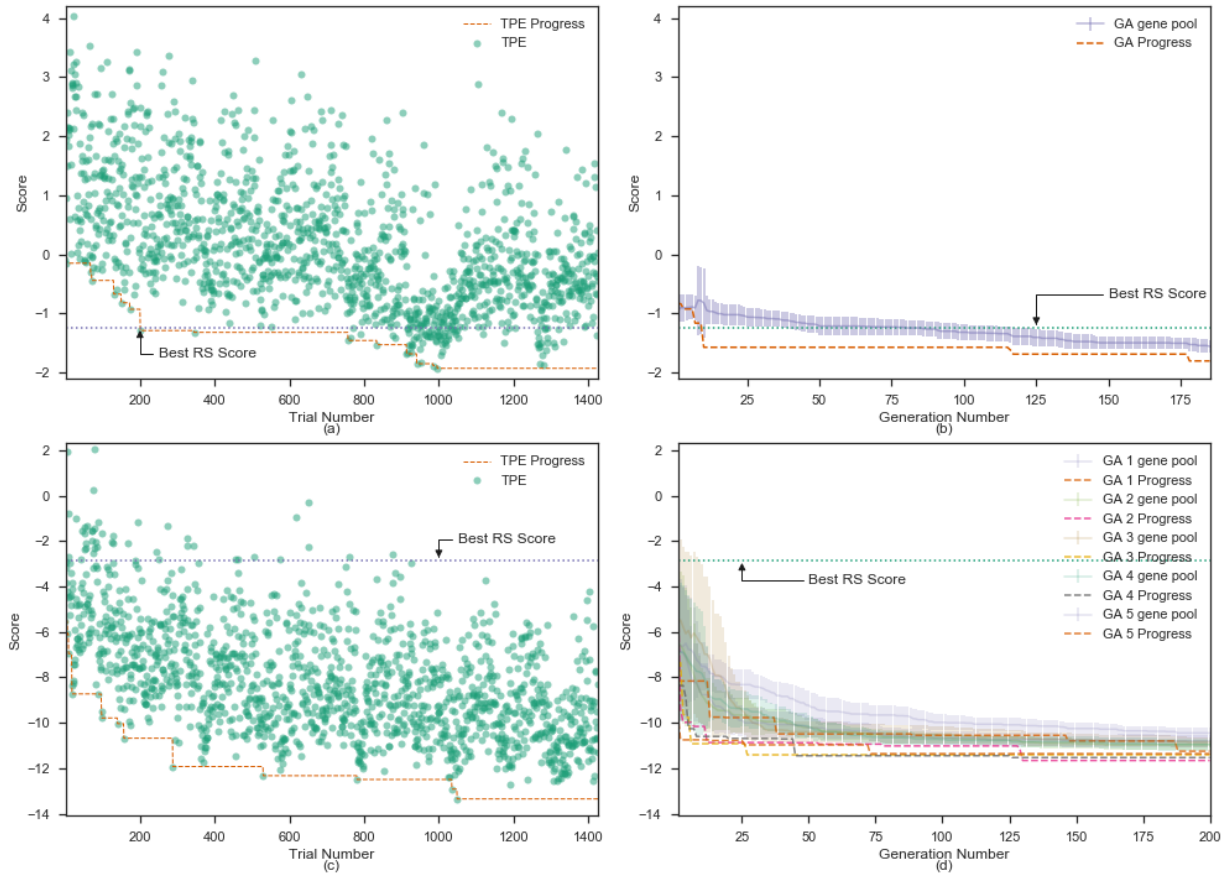


Figure 4.8: Optimization of the Sioux Faux 15k scenario with TPE (left) and GA (right) using "Post-Contest" (top) and "New KPIs" (bottom) objective function settings. The dashed line(s) across the bottom of each denotes the best (lowest) score achieved by an algorithm within the first N trials. Individual trial scores (at 40 iterations) are shown for TPE plots, whereas one standard deviation ranges of current gene pools are displayed in the GA plots. For the "Post-Contest" objective, the TPE and GA algorithms surpass the best score from 800 RS trials of 40 iterations (-1.24) within 200 and 10 trials, respectively. For the "New KPIs" objective, both algorithms significantly outperform the best result (-2.84) of an 800 trial, 40 iteration RS almost immediately.

and "New KPIs" studies[29].

[29]Partial convergence criteria of 40 iterations were used during initial search, as this was determined to be sufficient for establishing a trajectory consistent with a fully relaxed state.

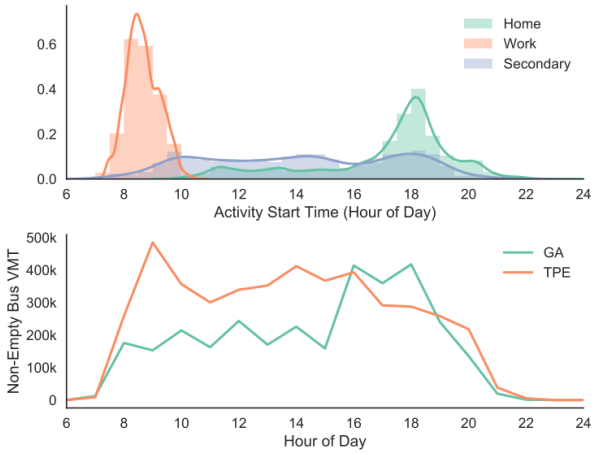


Figure 4.9: Example of output analysis for the "Post-Contest" case study. The upper plot shows the various activity start times of agents by activity type. The lower plot shows non-empty bus VMT for two competing algorithms.

Inputs corresponding to the trial yielding the top score for each algorithm were then simulated for 100 iterations with five replicates per trial. Figure 4.8 demonstrates that both GA and TPE produce input configurations that are superior to RS. We found that both GA and TPE achieve optimal solutions that reflect sensible yet distinct transportation system management strategies. For example, Figure 4.9 illustrates the relationship between activity start times (top subfigure) and bus utilization (bottom subfigure) as computed using outputs for the highest scoring solution for GA and TPE algorithms when using the "Post-Contest" objective function. This figure suggests that differences in optimal TPE and GA solutions for the "Post-Contest" objective arise from distinct transit usage patterns. Note that "Work" activity start times occur in the early morning (between 7:00 and 10:00 AM) and correspond to the highest period of utilization for buses under the optimal solution found for the TPE algorithm. In contrast, the GA solution ensures that buses are available during the evening peak commute time; that is, when agents travel back home from work and/or engage in secondary activities.

In the case of TPE, we found that the algorithm produced solutions that corresponded to sensible real-world policies. Figure 4.10 presents visualizations of input distributions for the top fifth percentile of TPE trials (corresponding to the top 70 of 1400 evaluated solutions by score, as plotted on Figure 4.8). This figure illustrates that the values of components of $\hat{d}$ for the best performing (lowest scoring) solutions using TPE occupy a narrow band in the design space. For example, in Figure 4.10(a), the highest scoring TPE input value sets evaluated in BISTRO under the "Post-Contest" objective suggest charging more expensive bus fares ($\$8 - \$10$) for adult citizens (16-60) than for youth (1-15) ($\$4 - \$6$) and elderly (60-120) ($\$5 - \$7$). The low variance of the components of $\hat{d}$ for these trial points is indicative of both objective function sensitivity to UDI definitions as well as robust algorithm convergence to

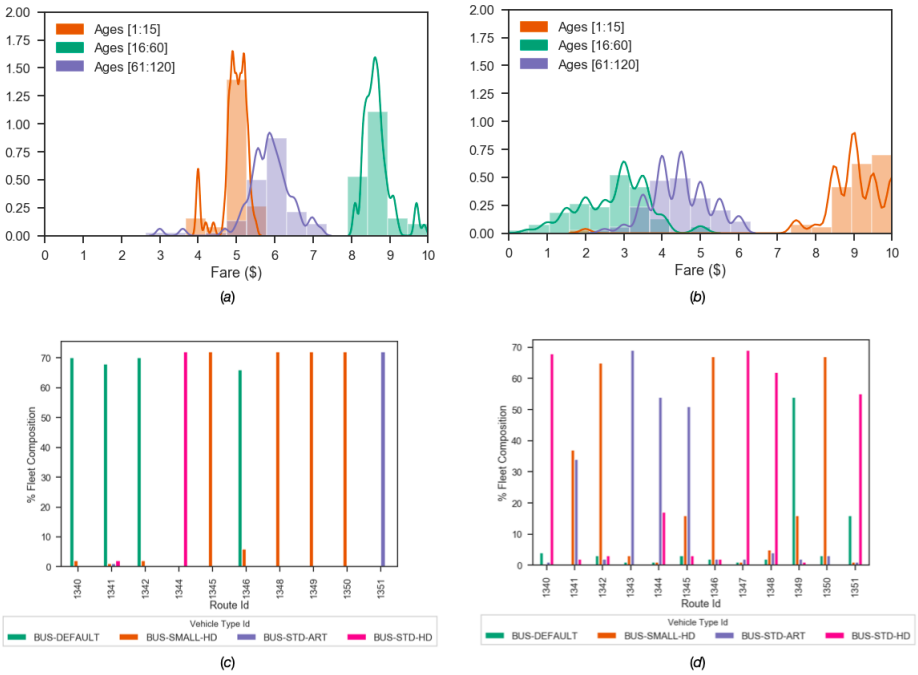


Figure 4.10: Distributions of bus fare by age (top) and vehicle fleet mix by route (bottom) for inputs representing the best fifth percentile scores among trials run for Sioux Faux 15k scenario using the TPE algorithm, shown for "Post-Contest" (left) and "New KPIs" (right) objective functions.

a (locally) minimal score value. The corresponding bus types on a given route suggested by these solutions are well-resolved and tend towards smaller vehicle models. In contrast, for near-optimal inputs evaluated using the "New KPIs" objective function, fares assigned to youth ($\$4 - \$10$) are, on average, higher than those assigned to adults ($\$0 - \$4$) and seniors ($\$5 - \$10$). The corresponding bus types by route are also more diverse among optimal solutions, indicating that the objective is less sensitive to the `VehicleFleetMix` input when evaluated using the "New KPI" objective function. Using the "New KPIs" objective, GA (not shown) also finds a tight distribution of fares for top-performing solutions but contrarily finds diversity in its `VehicleFleetMix` solutions.

4.5 Conclusion

This chapter has presented the design, software architecture, and preliminary evaluation of BISTRO: a general-purpose transportation policy decision support tool and scenario-based optimization framework supported by empirically-driven agent-based models. When combined with sensible guidance from experienced planning professionals, BISTRO can be used to identify more holistic, empirically-driven approaches to urban transportation planning

and management. In addition to overall system purpose, design, and software architecture, this work provides a concrete example of the process that BISTRO supports as implemented in the context of a scenario-based policy optimization "contest." While many participants had little or no prior expertise in the transportation science and policy analysis methods typically used in urban planning practice, over a dozen teams developed algorithms that found inputs, which, when evaluated in the simulator, achieved scores that surpassed both random search as well as human judgment. The mixed results of the competition led us to conclude that the optimization-based search techniques enabled by BISTRO should support an iterative approach that involves applying optimization algorithms to refinements of KPI specifications in order to better align objective functions with system goals.

Research conducted using BISTRO strives to meet the highest standard of reproducibility in computational experiments [139, 140, 141] as well as fact-based policymaking [142] by making all data, models, and algorithms freely available and open source[30]. One finding of post-contest reproducibility efforts was that different classes of algorithms appeared to converge to solutions that emphasized distinct policy strategies.

Our experience from this pilot study demonstrates that we have implemented a compelling platform to study human-in-the-loop design of expensive simulation-based optimization algorithms. This conclusion suggests that, in addition to its utility as a decision support system, BISTRO could serve as an exemplary testbed for multiple emerging streams of research (*e.g.*, freeze-thaw, multi-objective, multi-task, and multi-fidelity optimization) in SMBO and associated meta-model-based optimization methods. Should BISTRO be widely adopted as part of the urban planning toolkit, innovative algorithms and new theory developed as part of inquiry in these sub-domains will have the added benefit of directly serving a humanitarian purpose.

[30]All code and data used and referred in the chapter is available at `http://bistro.its.berkeley.edu`

Chapter 5

Congestion Pricing Optimization Case Study I: Sioux Faux

5.1 Introduction

In this chapter, I use the Berkeley Integrated System for TRansportation System Optimization (BISTRO) to develop a simulation-based methodology for multi-objective optimization of congestion pricing. The methodology is applied to a benchmark scenario to investigate the impacts of pricing structure and objective function design on congestion pricing optimization outcomes. This chapter was written in collaboration with Makena Schwinn, Léo Toulet, Zangnan Yu, Anyi Chen, Xinyi Tang, Timothé Kasriel and Alexandre Bayen. It is under review for publication in the *IEEE Transactions on Intelligent Transportation Systems*. I led the study design, including the development of the methodological and analytic approach, oversaw the implementation and execution of the study, and conducted all of the analysis presented in this chapter.

Congestion pricing has been garnering increasing attention as a promising *transportation demand management* (TDM) strategy to aid in addressing key issues on the transportation policy agenda, including transportation finance, congestion mitigation, energy consumption, and pollution. Congestion pricing schemes across the globe have proven effective in reducing congestion by charging vehicles for travel within a designated area that typically corresponds to the city center. In 2019, New York City became the first city in the U.S. to approve a congestion pricing plan, which was projected to lead to a 6.7% reduction in vehicle miles traveled (VMT) in the charging zone.

Several challenges arise in the pursuit of an optimal pricing policy that achieves operational and environmental objectives while improving mobility and accessibility in an equitable manner. While a utilitarian approach may suggest that pricing road access with respect to the marginal cost of an additional vehicle is optimal, the equity implications of doing so are highly dependent on the level of service provided by alternative transportation modes such as public transit, walking and biking [143, 144]. Although targeted reinvestment of

congestion pricing revenue has been deemed an important strategy for alleviating the potential inequities of "pricing out" lower income drivers from what is often the fastest and most convenient mode of transportation, the design of a pricing policy that directly optimizes for multiple competing social and operational objectives has not yet been proposed.

This chapter presents a case study of multi-objective congestion pricing optimization using the *Berkeley Integrated System for Transportation Optimization* (BISTRO), an open-sourced transportation planning and decision support system that uses an agent-based simulation and optimization framework to develop and evaluate multi-dimensional transportation system interventions. The aim of this study is to demonstrate the key opportunities and challenges in using agent-based simulation with activity-based travel models to optimize transportation policies across multiple conflicting objectives. This chapter demonstrates how BISTRO enhances the interpretability of multi-objective transportation policy optimization by enabling researchers and policy-makers to investigate the implications of both the objective function and policy lever design.

The case study implements a Bayesian Optimization algorithm on a four-dimensional pricing policy and a weighted multi-objective function consisting of six key performance indicators (KPIs). Two variants of cordon-based congestion pricing are optimized: 1) a flat cordon toll charged upon entry to a circular cordon, and 2) a cordoned mileage fee charged for all travel within a circular cordon. The location and size of the cordon as well as the rate charged in either policy variant are encoded as model inputs. The KPIs include total VMT, average vehicle hours of delay (VHD) per passenger trip, total greenhouse gas (GHG) emissions, average generalized travel cost burden (CB) by trip purpose, and total toll revenue. The aggregate behavioral response to variants of the pricing policies across trip lengths and geographies are analyzed in order to demonstrate the inherent trade-offs in transportation policy optimization. Although the results of this case study are not intended to be implemented in practice, the demonstrated methodology exemplifies the importance of 'digging under the hood' of black-box optimization in order to produce explainable data-driven policies.

5.2 Background

Transportation systems are highly complex civil systems, with interrelated dynamics constrained by the physics of the transportation network and the resources that make up the supply of both private and public transportation. Travel demand, which determines the spatio-temporal distribution of people and vehicles in the transportation network, is sensitive to the geographic distribution of the population, the spatio-temporal distribution of their travel needs, and the time and cost needed to travel using available transportation modes (e.g., personal vehicle, public bus, walk, etc.). The design of pricing strategies for TDM necessitate modeling techniques that represent the physical, operational, and behavioral dynamics of the transportation system in order to determine the optimal policy with respect to both system-level objectives and individual-level impacts. The following subsections provide

an overview of congestion pricing optimization and simulation-based optimization as well as an introduction to agent-based simulation and the BISTRO optimization framework used in this study.

Road Pricing Optimization

Road pricing charges road users for access to a particular road or area within a road network. Road pricing schemes that are devised to elicit a reduction in the volume of vehicles in specific congested areas and/or time periods, known as *congestion pricing*, can take various forms. When applied to *express* lanes, congestion pricing is typically dynamic, with a distance-based rate that is optimized to maintain a desired level of service on the tolled lane. Zonal congestion pricing, including *cordon* and *area-based* policies, charge vehicles for entry into or movement within a designated area, respectively. Area-based pricing includes flat-rate (e.g., the London Congestion Charge), dynamic, and distance-based pricing (e.g., Singapore's Electronic Road Pricing 2.0).

The optimization of congestion pricing schemes has been widely studied in the transportation literature. Historically, the optimization of road tolls has been studied using analytical models of vehicle dynamics on a single road or in a simplified road network, with an emphasis on the examination of the theory of marginal-cost pricing [22, 23]. This literature has since expanded, with an emphasis on the examination of the optimal location, price structure, and/or price level for congestion pricing schemes, with congestion minimization as the most common objective. When optimizing for congestion minimization, measures of total throughput [145], average travel times [63, 145, 146], critical network density [62, 64, 147], and the spatial spread of congestion [64] have been used. Few congestion pricing optimization studies consider the impact of the pricing scheme to congestion *outside* of the charging zone [62, 64, 147], and fewer actually incorporate network-wide indicators in the objective function for optimization [63, 146, 148]. Some studies have explored multi-objective optimization of congestion pricing, by incorporating minimization of user costs [39], maximization of toll revenue [145, 146], and minimization of emissions [58] as additional objectives to congestion minimization using scalarization techniques [58, 146, 149], constrained optimization [64, 145], or pareto-based acquisition functions [145].

Simulation-Based Optimization of Transportation Systems

Simulation enables the evaluation of hypothetical scenarios in which one or more aspects of the transportation system are optimized using models that represent the interaction of important variables of interest. Simulation-based optimization problems can be formulated as follows:

$$\min_{\mathbf{d}\in\mathcal{D}} f(\mathbf{d}, \mathbf{x}; \mathbf{z}) \tag{5.2.1}$$

$$\text{s.t. } \mathbf{x} = B(\mathbf{d}; \mathbf{z}) \tag{5.2.2}$$

$$g(\mathbf{d}; \mathbf{z}) = \mathbf{0} \tag{5.2.3}$$

where the decision vector $\mathbf{d}$ from the input space $\mathcal{D}$ is chosen such that it minimizes an objective f, a function of the decision vector, endogenous variables $\mathbf{x}$, and exogenous variables $\mathbf{z}$. The endogenous variables are measurable outcomes of the simulation, which may include agents' choices and their outcomes (e.g., travel times, travel costs, vehicle paths). Exogenous variables include configuration parameters that define the supply, demand, and dynamics of the system. The function B represents the simulation of outcomes given a decision vector and a set of exogenous variables, the latter of which includes parameters that determine the characteristics and activity of the population and/or parameters that define the physical and operational makeup of the transportation network. The function g represents design constraints, which can be implemented to enforce realism and/or interpretability in the simulation configuration and outcomes.

In problems where the simulation produces stochastic outcomes, as is the case in agent-based models using random utility models to reflect the stochasticity in human behavior (see section 5.2), f can be defined as the expected outcome of stochastic performance measures, F:

$$f \equiv \mathbb{E}[F(\mathbf{d}, \mathbf{x}; \mathbf{z})] \tag{5.2.4}$$

A variety of different simulation-based optimization approaches have been applied to optimize congestion pricing strategies. Some studies have used surrogate-based techniques to optimize toll zones and prices to minimize network congestion [63, 64]. A metamodel can also be used to approximate the simulation-based objective. Metamodels can be much more computationally efficient to solve, and thus a suitable approach for large-scale transportation problems [150]. Zheng, et al. (2012) used a macroscopic fundamental diagram to control toll rates in a cordoned scheme, in conjunction with an agent-based simulator [62]. However, the simulator's demand elasticity was limited to departure time and route choice, excluding mode choice. BISTRO, the optimization framework used in this study, uses an agent-based simulation (ABS) with an activity-based travel model that includes models for departure time, mode, and routing choices.

Agent-Based Simulation

ABS with activity-based travel models provides both increased interpretability and flexibility for transportation planning and policy decision-making. In activity-based travel models, a regional population of citizens is modeled as a set of individual agents, with discrete choice models representing the decisions they face as they carry out their daily travel plans. ABS

models simulate the decisions and resulting actions at an individual level. The physical dynamics of vehicles within the transportation network are typically modeled using directed graph representations of road networks that determine the physics of vehicle movement. *Traffic assignment* methods are used to determine the path taken by each vehicle given a specified origin and destination (OD), and departure time and any additional relevant information for each vehicle, producing estimates of the volumes of vehicles on all links in the network at any given time as well as the speed, fuel consumption, and trip-level travel times and costs for each vehicle throughout the simulation.

The integration of individual-level behavioral models with physical models of vehicle dynamics within a representative transportation network in an activity-based ABS enables policy-makers to gain insight into the spatio-temporal distribution of important social, economic, and environmental impacts at the individual, household, zonal, and regional levels [29]. More simplified methods used for travel models, such as the traditional four-step trip-based model and traffic assignment models with fixed demand generally suffer from aggregation biases due to the representation of demand sensitivity at much lower spatial resolution, if at all, making it more challenging for stakeholders to interpret the outputs from these models and robustly incorporate them into policies [151, 152].

BISTRO

BISTRO is an open source transportation planning decision support system that leverages the power of ABS, activity-based travel models, and machine learning to assist stakeholders in addressing the increasingly complex problems arising in transportation systems worldwide. BISTRO facilitates the algorithmic optimization of user-defined transportation system intervention strategies using an empirically-validated ABS of multimodal metropolitan transportation systems, called the *Behavior Energy Autonomy and Mobility* (BEAM) framework, as its core simulation engine. BISTRO's multi-objective optimization framework enables experimentation with mechanism design by parameterizing system interventions and developing appropriate KPIs that best align with policy and planning objectives. The following subsections describe the pertinent sub-components of BISTRO, including the scenario development pipeline, BEAM simulation engine, and the optimization framework consisting of user-defined inputs, KPIs, and objective function design. For more detail about the design and operation of BISTRO, the reader is referred to chapter 4.

BISTRO Scenario Development

A BISTRO scenario is defined by the configuration of the transportation network, land-use, synthetic population of agents, and the operation of transportation services in a particular geography of interest. The transportation network includes the road network, represented as a directed graph of nodes and links, each of which is characterized by its' length, capacity, speed limit, and mode access restrictions, if any. In addition, multiple public transit networks may be configured, defined by one or more transit agencies, routes, stations, and the

vehicle type that serves each route. The population configuration includes the distribution of population density, vehicle ownership, and socio-demographic attributes of the population at a zonal level (e.g., traffic analysis zone (TAZ) or census tract), which is used to produce a synthetic population of households of one or more agents with a specified home location. The agents' activity plans, consisting of two or more activity types (e.g., home, work, school), locations, and desired start and/or end times, are also determined by the scenario configuration and kept fixed throughout the study. This approach excludes the possibility of induced demand within a scenario since agent trips are exogenous to the optimization of interventions. Scenario analyses may be conducted by generating multiple sets of activity plans and/or altering the coefficients of the mode choice model to optimize interventions across variations in the quantity of travel demand and the sensitivity of modal demand with respect to travel time components, travel cost, other attributes of modes, and/or individual characteristics. Lastly, operational parameters of transportation services such as public transit, on-demand mobility services, and other forms of shared mobility may be configured, including fare structures, hours of service, vehicle fleet mixes, and vehicle repositioning algorithms.

BEAM

The core simulation engine in BISTRO is BEAM, the Modeling Framework for *Behavior Energy Autonomy and Mobility*. BEAM is an activity-based ABS developed at the *Lawrence Berkeley National Laboratory* (LBNL). It extends the *Multi-Agent Transportation Simulation* Framework (MATSim) to enable analysis of urban transportation systems using discrete choice models of within-day travel decisions. Within the BISTRO framework, scenario configuration parameters discussed in section 5.2 serve as fixed inputs to BEAM, which executes the second-by-second movement and decision-making of all agents (including travelers and drivers) throughout a day of travel.

At the start of the simulation, all travelers are endowed with a plan of activities (including location and start/end times) and are located at their first activity of the day (e.g., at home or work) and all transit and TNC drivers are positioned with their corresponding vehicles at a starting location for the day. As travelers end an activity, their choice of transportation mode with which to travel to their next activity (e.g., work) is determined using a *multinomial logit model* (MNL) given the estimated travel time and cost of all available modes at that time in the simulation. The agent may choose to walk, bike, drive alone in a personal vehicle, ride alone in a TNC, or take pubic transit in combination with any of the other modes that facilitate access/egress to/from public transit stations. All private and TNC vehicles are routed via the shortest path to their destination with respect to driving time. However, mode choices consider only the maximum utility path for each mode which, in the presence of road pricing, may not always coincide with the shortest-time path.

BISTRO Optimization Framework

The BISTRO optimization framework consists of three key components: 1) *user-defined inputs* (UDIs) representing the transportation system intervention of interest, 2) KPIs encoding the objectives of the intervention strategy, and 3) the scoring function that formulates how the KPIs will be translated for use in the optimization algorithm. The optimization algorithm must define a set of UDI files to submit to BISTRO and receive the score values evaluated by BISTRO following a simulation run in which the UDIs are applied.

BISTRO provides a library of UDIs, including investment (e.g., vehicle fleet mix, public transit routing, parking supply, vehicle charging infrastructure), incentive (e.g., monetary incentives differentiated by socio-demographic groups for the use of a specific mode or combination of modes), and operational (e.g., public transit scheduling, public transit fares, road pricing/tolls, parking pricing) strategies [30]. The UDIs may be constrained using the *business rules* feature of BISTRO, which defines validation parameters for the UDIs (e.g., range of possible values).

KPIs are the sub-components that make up the objective function that is to be optimized. KPIs may represent operational, environmental, or social goals that are pertinent to evaluating the optimality of the strategy of interest. Two general types of KPIs have been developed in BISTRO, including KPIs that measure: 1) the operational efficiency of the transportation system (e.g., VMT, VHD, operational costs) and 2) the experience of transportation users (e.g., bus crowding experienced, generalized travel cost, accessibility).

Finally, a scoring function must be designed in order to define the output to be used in the optimization. In this chapter and in previous work by the authors, a scoring function that aggregates multiple *score components* into a one-dimensional scalar score has been used. Each score component corresponds to one KPI. In order to normalize the relative improvement achieved by a particular UDI across all KPIs, each score component is computed as the normalized ratio of the value of the corresponding KPI in the given simulation run to the value of the same KPI in the business-as-usual (BAU) run, in which all UDIs are null. The improvement ratios are normalized using KPI values produced by a randomized sample of the UDI space, accounting for differences in variance across KPIs. The composite score is thus a function of the normalized improvements from the BAU to the candidate input in each metric, as follows:

$$F\left(\mathbf{d}, \mathbf{x}, \mathbf{x}_0, \mathbf{K}, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\beta}\right) = \sum_{i \in \mathbb{K}} \beta_i \frac{\frac{K_i(\mathbf{d},\mathbf{x})}{K_i(\emptyset,\mathbf{x}_0)} - \mu_i}{\sigma_i} \tag{5.2.5}$$

where $\mathbf{K}$ is the vector of all KPIs evaluated for a given set of inputs, $\mathbf{d}$, and the resulting outputs, $\mathbf{x}$, produced by the simulation. The simulation outputs from the BAU run in which no inputs are provided is the vector $\mathbf{x}_0$. The vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are normalization parameters for each KPI, and $\boldsymbol{\beta}$ is a vector of coefficients that may be applied as appropriate. When dealing with a set of objectives that vary with respect to the direction of desired improvement (e.g., maximize revenue and minimize VMT), the β parameters may be used to change the

sign of a score component. The β parameters may also be used to apply a weighting scheme to the score components, as is demonstrated in this chapter.

5.3 Methods

The present case study investigates congestion pricing optimization using a benchmark BISTRO model called Sioux Faux, which is based on the city of Sioux Falls, South Dakota[1]. In particular, this case study uses BISTRO to optimize two potential zonal congestion pricing policies: 1) a cordon fee, and 2) a cordoned mileage fee. This study examines the utility of BISTRO for facilitating interpretable machine learning for the congestion pricing optimization problem using a Bayesian optimization algorithm with a scalarized multi-objective function. The following subsections present the scenario configuration, problem formulation and optimization approach taken.

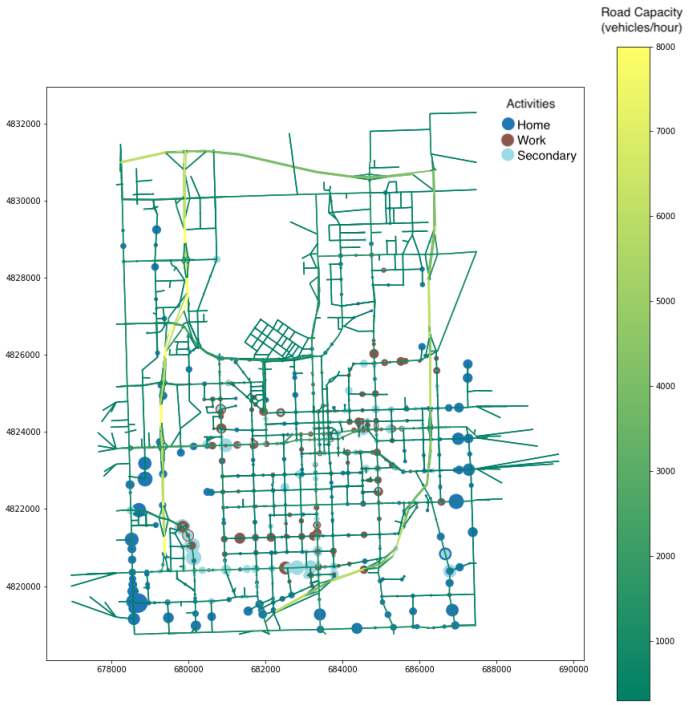


Figure 5.1: Sioux Faux road network and activity locations (activities are sized by quantity).

[1]Detailed information about the Sioux Faux benchmark scenario can found in Appendix A

Table 5.1: Sioux Faux mode choice model parameters

Variable	Description	Value
a_{drive}	Drive alone ASC	10.56
a_{walk}	Walk ASC	-11.0
a_{TNC}	TNC ASC	-0.124
$a_{drive-transit}$	Drive to/from transit ASC	0.0
$a_{walk-transit}$	Walk to/from transit ASC	-10.0
v	VoT	18 [$/hr]
r	Transfer coefficient	-0.6

Sioux Faux Scenario

The Sioux Faux scenario includes a road network of 2,350 nodes and 5,812 links with a synthetic population of 15,000 agents. Each agent is endowed with age, income, and gender attributes as well as a household and corresponding home location, some number of household vehicles, and the location of one primary activity they complete during the day, which is either work or a secondary activity (e.g., groceries, health care, etc.). The distribution of agent activities is shown overlaid on the road network of Sioux Faux in Figure 5.1. Agents may choose from the following available modes: drive alone, walk, ride alone in a TNC, and drive, walk, or ride alone in a TNC to/from a public bus. Agents' mode choices are determined by an MNL discrete choice model, in which the probability that an agent chooses a particular mode m for a particular trip i departing at time t is defined in equation 5.3.1 below:

$$P(x^i_{mode}[t] = m) = \frac{e^{a_m + x^i_{cost,m}[t] - v \cdot x^i_{dur,m}[t] + r \cdot x^i_{transfer,m}[t]}}{\sum_{m' \in modes} e^{a_{m'} + x^i_{cost,m'}[t] - v \cdot x^i_{dur,m'}[t] + r \cdot x^i_{transfer,m'}[t]}} \tag{5.3.1}$$

where $x^i_{mode}[t]$ is the mode chosen by agent i at time t and the variables $x^i_{cost,m}[t]$ and $x^i_{dur,m}[t]$ are the total estimated trip cost and duration of mode m for trip i at time t. The variable $x^i_{transfer,m}[t]$ denotes the number of transfers (i.e., between bus lines) involved in using mode m for the trip. The values of the alternative-specific constants (ASCs) a_m, value of time (VoT) v, and transfer coefficient are reported in Table 5.1.

For the purposes of this case study, the Sioux Faux scenario is configured with unlimited free parking in order to control for the potential confounding effects of parking supply on the optimization results. In addition, the location of each vehicle in the TNC fleet, the size of which is equal to 15% of the synthesized population, is randomly initialized within the road network. To control for additional VMT from TNC repositioning, the TNC vehicles simply wait at the location of their last drop off to be reassigned to the next ride request.

In the BAU simulation run in which no congestion pricing scheme is applied, the majority of agents drive, resulting in about 73% of all trips being completed by driving alone in personal cars. TNC trips account for 3% of all trips, while about 13% of trips use the bus

and about 12% of trips are completed by foot. Cars are used for the majority of commute trips, while the modal split for secondary trip purposes is more evenly split between car and public bus accessed by foot. The combination of personal car, TNC, and public bus use produces a total of 65,800 VMT, 30,000 kg of CO_2 equivalents, and an average vehicle hours of delay (VHD) of 0.225 hours (13.5 minutes) per passenger trip (see Table 5.2). The majority of congestion is produced by driving alone during peak commute hours. TNC trips, which tend to be between 0.5 and 3 miles long, are distributed throughout the day, with the greatest amount of TNC VMT generated from 9 AM to 4 PM. Finally, the average generalized cost burden is lower for work trips than for secondary trips in the BAU. This reflects the relative speed and low cost of predominantly car-based commutes.

Optimization Approach

An optimization problem of the form described in section 2.1.1 is formulated for each of two cordon-based congestion pricing policies: cordon tolls, and cordoned mileage fees. BISTRO is used to implement a Bayesian optimization of the two pricing policies in the Sioux Faux scenario with a scalarized objective function of six KPIs.

Problem Formulation

The congestion pricing policy is represented by a four-dimensional input vector, $\mathbf{d} \in \mathbb{R}^4$, consisting of the center (latitude and longitude) and radius (in meters) of a circular cordon, or charging zone, as well as a toll rate. In the cordon toll policy, the toll rate is the dollar amount charged upon crossing the boundary of the cordon (i.e., vehicles are charged upon entry and exit). In the cordoned mileage fee policy, the toll rate is the amount of dollars charged per mile driven on any road segment within the cordon. In both policies, all private passenger vehicle trips (drive alone and TNC) are charged the corresponding rate based on the route taken for the trip. The objective function is composed of six KPIs including total VMT, average VHD per passenger trip, total greenhouse gas (GHG) emissions, average generalized travel cost burden (CB) of work trips, average CB of secondary trips, and total toll revenue. The following constraints are applied: 1) the center of the cordon must be within the city limits (equations 5.3.4 - 5.3.5), 2) the diameter of the zone must be within a defined range (equation 5.3.6), and 3) the toll rate must be non-negative and less than $10 (equation 5.3.7). The resulting optimization problem is specified as follows:

$$\min_{\mathbf{d}\in\mathbb{R}^4} f(\mathbf{d}, \mathbf{x}; \mathbf{z}) \tag{5.3.2}$$

$$s.t. \quad \mathbf{x} = B(\mathbf{d}; \mathbf{z}) \tag{5.3.3}$$

$$d_1 \in [43.5007°N, 43.6195°N] \tag{5.3.4}$$

$$d_2 \in [-96.8112°W, -96.6498°W] \tag{5.3.5}$$

$$d_3 \in [0, 8.5] \tag{5.3.6}$$

$$d_4 \in [0, 10] \tag{5.3.7}$$

Table 5.2: BAU values and normalization parameters for each score component

		BAU	Cordon toll		Cordoned mileage fee	
i	Name	$K_i(\emptyset, \mathbf{x}_0)$	μ_i^c	σ_i^c	μ_i^m	σ_i^m
1	VMT	$83,000$	$35,000$	$1,700$	$52,000$	$21,600$
2	VHD	0.22	0.19	0.032	0.16	0.074
3	GHG	$4.9E9$	$4.3E9$	$1.7E8$	$3.5E9$	$9.8E8$
4	CB_{work}	0.00098	0.00065	0.000075	0.00089	0.00030
5	CB_{2ndary}	0.00081	0.0011	0.00020	0.0016	0.00011
6	TR	0	$\$12,000$	$\$14,000$	$\$45,000$	$\$38,000$

where the first two dimensions of the input vector $\mathbf{d}$ correspond to the latitude and longitude of the center of the charging zone while the remaining dimensions correspond to the radius of the zone and the toll rate, respectively.

Recall that the objective function $f(\mathbf{d}, \mathbf{x}; \mathbf{z})$ is the expected value of the stochastic performance of the simulation, defined by $F(\mathbf{d}, \mathbf{x}; \mathbf{z})$ in equation 5.2.5. The set of KPIs, $\mathbb{K} := \{K_i\}_{i=1}^6$ (see Table 5.2), are each a function of the fixed input vector $\mathbf{z}$ (i.e., person, vehicle, and network attributes) and the vector $\mathbf{x}$ that contains the intermediary output values that are used to calculate the KPI[2]. These include the VMT and GHG emissions of each vehicle movement during the simulation; the total travel time and passenger costs of all trips; the delay experienced during each passenger trip (i.e., the difference between the free-flow and the actual travel times of a trip). The generalized transportation CB for a trip is calculated as the ratio of the generalized transportation cost of the trip and the household income of the trip maker, where the generalized transportation cost is the sum of the costs incurred during the trip and the product of the population-average VoT ($18/hour) and the duration of the trip.

Scalarization

The examination of the role of scalarization in multi-objective transportation policy optimization is a key component of the analysis presented in this study. This process is initialized with a random search, for which 500 samples were drawn from a uniform distribution of the input space for each pricing policy. The normalization parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ for each KPI are calculated from these random samples (see Table 5.2). The distributions of the KPIs with respect to the policy parameters were examined to identify correlations between KPIs that informed the development of the scalarization scheme. The results of the random search and analysis of weighting schemes are presented in section 5.4.

[2]The formulation of each of the KPIs used in this case study can be found in the Sioux Faux general system specification in Appendix A

Bayesian Optimization

Bayesian optimization uses a probabilistic surrogate model to approximate an unknown objective using information from the prior distribution, along with an acquisition function that uses information from the posterior distribution to dictate how to explore the input space [153]. This study uses the *tree-structured Parzen estimator* (TPE) acquisition function [154]. Unlike Gaussian processes that model the posterior distribution $p(y|x)$ directly, the TPE models $p(x|y)$ and $p(y)$ separately. The algorithm iteratively optimizes $\mathbf{d}$ given some set of m observations $\mathcal{H} = \{(\mathbf{d}^i, y^i)\}_{i=1,\dots,m}$, where y^m is the score evaluated from the m^{th} simulation run using $\mathbf{d}^m$ as input: $y^m = F(\mathbf{d}^m, \mathbf{x}^m; \mathbf{z})$. After each sample evaluation (i.e., a BEAM simulation run), the TPE algorithm estimates $p(\mathbf{d}|y, \mathcal{H})$ by producing two densities as follows:

$$p(\mathbf{d}|y, \mathcal{H}) = \begin{cases} l(\mathbf{d}) \text{ if } \ y < y^* \\ g(\mathbf{d}) \text{ if } \ y \geq y^* \end{cases} \tag{5.3.8}$$

where $y*$ is the γ-quantile of the objective function value among the observations such that $\gamma = p(y < y*|\mathcal{H})$. Given a set of observations, $\mathcal{H}$, the next value of the UDI to be sampled, $\mathbf{d}^*$, is chosen using the maximum expected improvement, as follows:

$$\begin{aligned} \mathbf{d}^* = argmax_{\mathbf{d}\in\mathcal{D}} \int_{-\infty}^{y^*} (y^* - y) p(y|\mathbf{d}) dy \\ = argmax_{\mathbf{d}\in\mathcal{D}} \int_{-\infty}^{y^*} (y^* - y) \frac{p(\mathbf{d}|y, \mathcal{H}) p(y)}{p(\mathbf{d})} dy \end{aligned} \tag{5.3.9}$$

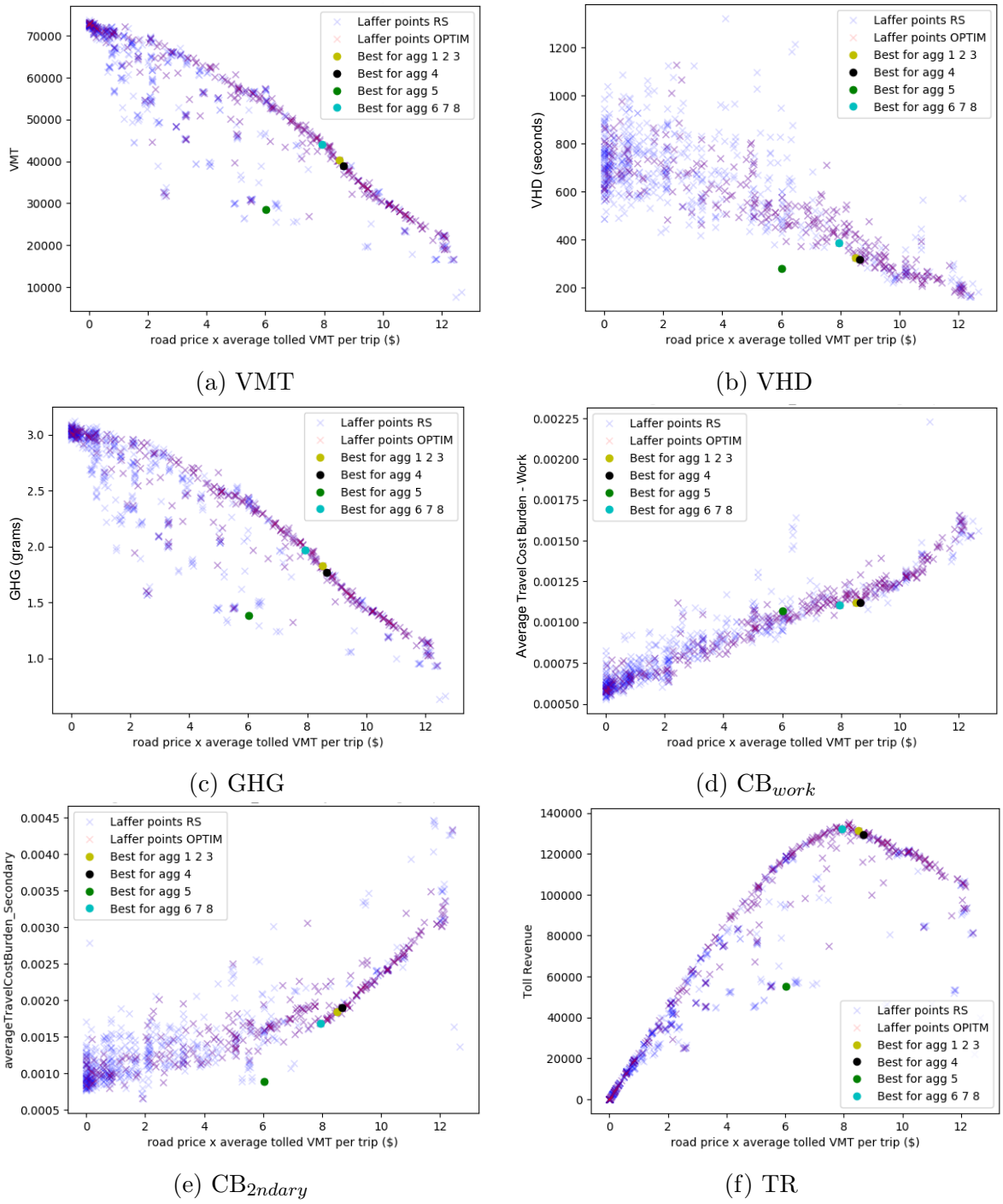
where, by construction, $p(\mathbf{d}) = \int_{\mathbb{R}} p(\mathbf{d}|y)p(y)dy = \gamma l(\mathbf{d}) + (1-\gamma)g(\mathbf{d})$. γ is set equal to 0.25 [154]; further investigation into the effect of tuning this hyper-parameter are under consideration in future work.

5.4 Results

This section presents the results of the random search and scalarization analysis, followed by an in-depth analysis of the optimization results for both pricing policies.

Scalarization

The random search results revealed strong correlations between two distinct groups of KPIs: 1) *Congestion KPIs* (i = 1,2,3), and 2) *Social KPIs* (i = 4,5). To demonstrate this, Figure 5.2 displays the distributions of each of the six KPIs with respect to the average toll paid per driving trip using all samples collected during the random search and optimization processes for the cordoned mileage fee, displayed as blue and red x's, respectively. The average toll is estimated as the product of the mileage fee and the average VMT within the cordon per driving trip. All of the congestion KPIs generally decrease with respect to the average toll,



(a) VMT

(b) VHD

(c) GHG

(d) CB_{work}

(e) CB_{2ndary}

(f) TR

Figure 5.2: Distribution of KPIs with respect to average toll per driving trip (samples from the random search (RS) and optimization (OPTIM) are shown in blue and red, respectively).

as displayed in figures 5.2a-5.2c. Conversely, the average CB increases with respect to the average toll, across both work and secondary trip types, as seen in figures 5.2d and 5.2e, respectively. Finally, in figure 5.2f we observe that the relationship between the total toll revenue and the average toll is approximately concave. Points that appear in the epigraph of figure 5.2f correspond to policies with relatively high mileage fees at a given average toll rate. The decreased total toll revenue of such policies compared to those with comparable average tolls reflect the behavioral response to higher fees which significantly reduces the number of driving trips within the cordon.

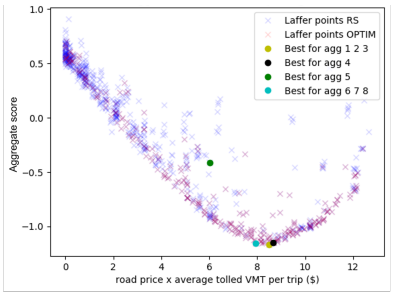
Due to the correlation across KPIs, we chose to constrain the scalarization parameters to three values as follows:

$$F\left(\mathbf{d},\mathbf{x},\mathbf{x}_0,\mathbf{K},\boldsymbol{\sigma},\boldsymbol{\mu},\boldsymbol{\beta}\right) = \frac{\beta_c}{3}\sum_{i=1}^{4}\frac{\frac{K_i(\mathbf{d},\mathbf{x})}{K_i(\emptyset,\mathbf{x}_0)}-\mu_i}{\sigma_i} + \frac{\beta_s}{2}\sum_{j=4}^{5}\frac{\frac{K_j(\mathbf{d},\mathbf{x})}{K_j(\emptyset,\mathbf{x}_0)}-\mu_i}{\sigma_i} - \beta_t\frac{K_6(\mathbf{d},\mathbf{x})-\mu_6}{\sigma_6} \quad (5.4.1)$$

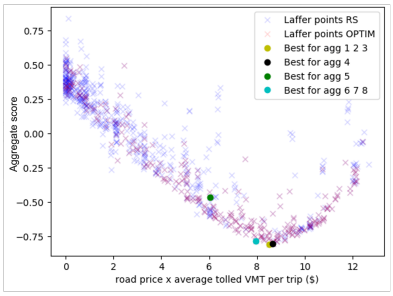
where β_c, β_s, and β_t are the weights for the congestion, social, and toll revenue score components, respectively, such that $\beta_c+\beta_s+\beta_t=1$. Note that the toll revenue KPI is not divided by the BAU value since there were no tolls in the BAU.

Eight different weighting schemes were tested. The distributions of equation 5.4.1 across all cordoned mileage fee samples using each weighting scheme are displayed in figure 5.3. The optimal samples corresponding to each weighting scheme are displayed in both figures 5.2 and 5.3. In the first five weighting schemes, the congestion and social score components were equally weighted, with successively decreasing weight applied to the toll revenue score component. The optimal policies for these weighting schemes (shown in light green, black, and dark green in figure 5.3). The remaining three weighting schemes, vary the weight distribution between the social and toll revenue score components while applying the least weight to the congestion score component.

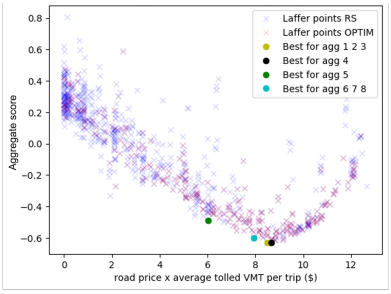
The optimal results for the various weighting schemes demonstrate the dominating effects of both the congestion and toll revenue score components. While the optimal policy under the last three weighting schemes (f-h) is close to the revenue-maximizing policy (see figure 5.2f), equal weighting between the congestion and social score components in the first four schemes (a-d) result in a slightly higher average toll rate that reduces congestion further while producing less total revenue. The optimal sample for the first three weighting schemes (a-c) was a \$3.50 mileage fee in a cordon that covered the entire city. In this sample, about 42% of trips shift from driving and TNCs to public transit and walking, resulting in reductions of about 50% in VMT and GHG, 40% in average VHD per passenger trip, and increases of about 17% and 111% in CB_{work} and CB_{2dary}, respectively. When the toll revenue weight is decreased to 0.2 in the fourth scheme (d), the optimal policy shifts to a higher average toll rate that further improves the congestion KPIs while worsening the social and toll revenue KPIs.



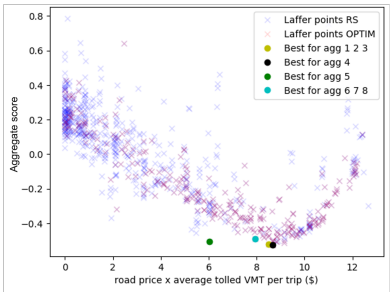
(a) (.25, .25, .5)



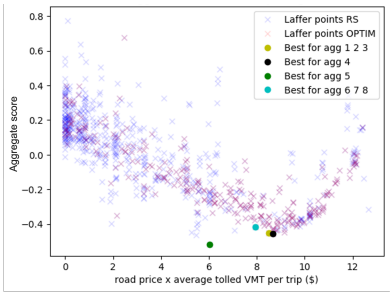
(b) (.33, .33, .33)



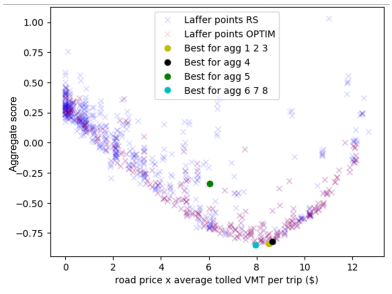
(c) (.375, .375, .25)



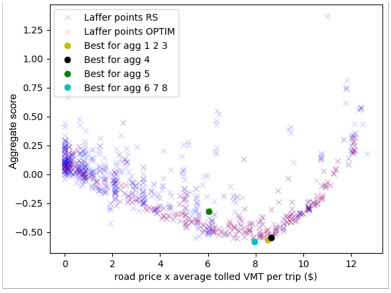
(d) (.4, .4, .2)



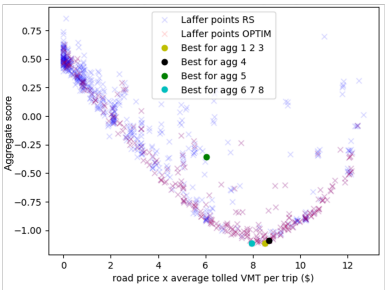
(e) (.415, .415, .17)



(f) (.2, .4, .4)



(g) (.2, .5, .3)



(h) (.2, .3, .5)

Figure 5.3: Distribution of the aggregate score with respect to average toll per driving trip under each weighting scheme (weights for the congestion, social, and toll revenue score components shown in parenthesis; RS samples are shown in blue; OPTIM samples are shown in red).

The outcomes of the optimal policy under the fifth weighting scheme (e) are highly unrealistic yet informative for future refinement of the problem constraints. Further reduction of the toll revenue weight results in much lower average toll rate, total toll revenue, congestion, and CB scores. The policy consists of a large mileage fee of $7.60 per mile within a 5.5-mile cordon centered in the south-western edge of the city that applies to most trips in the city, shifting all but about 15% of travelers to non-driving modes.

The optimal policy under the last three weighting schemes places the cordon around the entire city. However, the increased importance of the social score component relative to the congestion score component results in a mileage fee of just $3.20 per mile. We chose the sixth weighting scheme to use in the optimization in order to: 1) avoid over-emphasizing congestion reduction, which results in shifting an unnecessarily large portion of car-based trips to public transit and walking, and 2) remain agnostic in the valuation of CB versus toll revenue, since the latter has many potentially beneficial uses that can also be applied to reduce the social costs of a congestion pricing scheme (e.g., expansion or improvement of public transit, modal incentives, rebates).

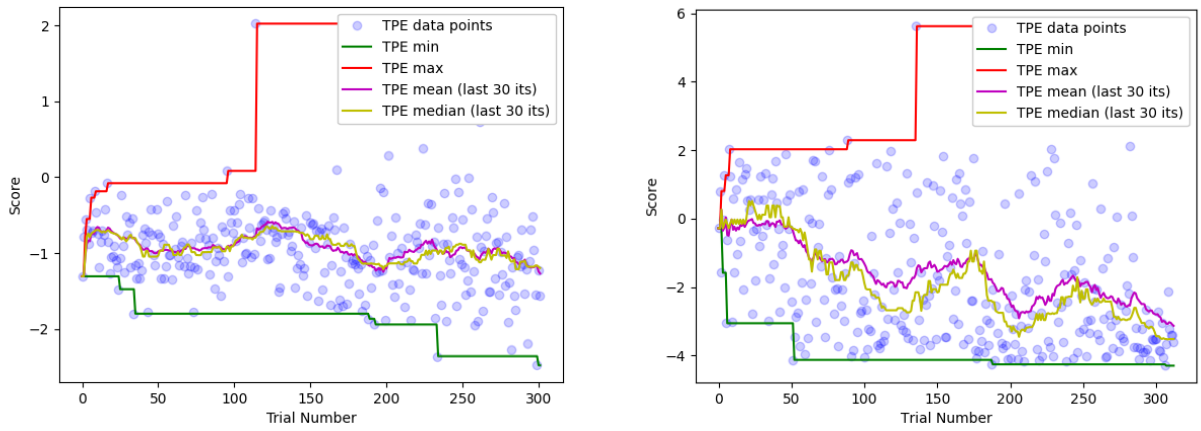
Optimization

This section presents the results of the optimization of the cordon toll and cordoned mileage fee policies in the Sioux Faux scenario using the sixth weighting scheme (see figure 5.3f). The convergence of the TPE optimization algorithm for the cordon toll and cordoned mileage fee are displayed in figures 5.4a and 5.4b, respectively. The mean and median scores of the samples in each optimization generally follow a downward trajectory, although a near optimal policy is sampled sooner in the optimization process for the mileage fee than for the cordon toll.

The optimal cordon toll is a charge of $6.90 for all car trips that cross a cordon with a radius of about 2.7 miles centered near the middle of the city (see Figure 5.5a). The toll, which applies to approximately 50% of all trips, including 52% of work trips and 47% of secondary trips, generates about $67,000 in total toll revenue. Only about 40% of all drive alone, TNC, or drive to/from public transit trips were tolled under the optimal cordon toll. In comparison, the optimal cordoned mileage fee is a charge of $3.20 per mile driven in a cordon with a radius of 7.9 miles that covers the entire city. Thus the mileage fee applies to all trips within the city and generates about $132,000 in total toll revenue.

The KPI values resulting from the optimal policies are presented in Table 5.3. The optimal cordon toll reduced the congestion-related KPI's by about 35 to 60%, while decreasing the average CB by 16% for work trips and increasing the CB for secondary trips by 21%. In comparison, the optimal cordoned mileage fee achieved greater reductions in total VMT and GHG emissions with a slightly smaller reduction in average VHD. It also increased the average CB for both work and secondary trips.

Both optimized congestion pricing policies resulted in a significant reduction in driving alone, as reflected by the improvements in the congestion-related KPIs. The optimal cordon toll and cordoned mileage fee resulted in 9% and 16% fewer drive alone trips, respectively.



(a) Cordon toll

(b) Cordoned mileage fee

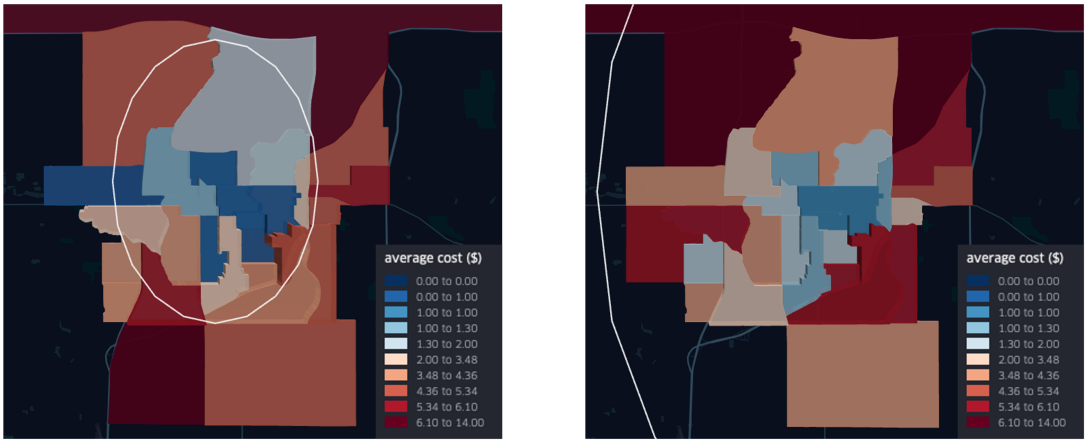
Figure 5.4: Evolution of the objective score during optimization.

Table 5.3: Sioux Faux congestion pricing optimization results

	Cordon Toll		Cordoned Mileage Fee	
i	$K_i^*(\mathbf{d}^*, \mathbf{x}^*)$	% change	$K_i^*(\mathbf{d}^*, \mathbf{x}^*)$	% change
VMT	54,000	−35%	39,000	−53%
VHD	0.092	−58%	0.11	−50%
GHG	$2.7E9$	−45%	$2.0E9$	−59%
CB_{work}	0.00082	−16%	0.0011	+12%
CB_{2ndary}	0.00098	+21%	0.0017	+110%
TR	\$67,000,	-	\$132,000	-

The optimal cordoned mileage fee had a greater impact on TNC use (-2.8%) than the optimal cordon toll, which only slightly reduced TNC mode share (-0.7%). Under both policies, the majority of trips that shifted from private auto modes resulted in walking trips, while the increases in public transit use were modest: +1% and +2% under the optimal cordon toll and cordoned mileage fee, respectively. It is important to note that about 66% and 70% of *attempted* transit trips were replanned in the optimal cordon toll and cordoned mileage fee scenarios, respectively. In BEAM, an attempted mode choice may be replanned in the event that the necessary resource (e.g., capacity on a public transit vehicle or a TNC vehicle with a reasonable wait time) is unavailable. Almost all replanned public transit trips were affected by buses operating at full capacity.

The two optimized pricing policies produced similar reductions in private auto use among trips less than two miles long (about 36% of trips), reducing the mode share of driving alone



(a) Cordon toll (b) Cordoned mileage fee

Figure 5.5: Average cost per AM peak driving trip by origin TAZ in each of the optimal policies.

and TNCs for trips of this length from about 85% in the BAU to about 76% and 75% under the optimal cordon toll and cordoned mileage fee, respectively. Greater differences across the two pricing policies are observed in the mode shifts among longer trips, with the optimal cordoned mileage fee resulting in much greater reductions in private auto use than the optimal cordon toll. Among trips two to five miles long (about 53% of trips), the optimal cordoned mileage fee reduced private car mode share from 72% in the BAU to 49% (compared to 64% in the optimal cordon toll). Finally, among the tenth of trips that were longer than 5 miles, the cordoned mileage fee reduced private auto mode share from 61% to 14% (compared to 57% in the optimal cordon toll).

Both pricing policies increase the monetary cost of driving, as shown in Figure 5.5. Since trips near the center of Sioux Faux tend to be shorter, the average cost of driving trips originating in this area were lower than in other parts of the city. Under the optimal cordon toll, trips close to the center of the city generally did not cross the cordon and thus were not tolled. Across both policies, the average cost per driving trip generally increased for trips originating further from the center of the city, as these trips tended to be longer in distance and thus incurred greater tolls.

The increased costs were partially offset with reduced travel times resulting from the overall reductions in VMT and average VHD under each policy. The average driving speed increased by about 35% and 25% under the optimal cordon toll and cordoned mileage fee, respectively. In addition, the optimal cordoned mileage fee produced mode shifts that reduced the average distances of drive alone and TNC trips by about 20% and 30%, respectively, while the optimal cordon toll reduced drive alone and TNC trip distances by only about 5% and 10%, respectively. Ultimately, the average generalized cost of driving alone (the sum of the monetary cost and the monetary value of the trip duration) only increased by

1% from the BAU under the optimal cordon toll, compared to a 40% increase under the optimal cordoned mileage fee. Furthermore, the average generalized travel costs of TNC trips decreased by about 20% under the optimal cordon toll (due to the increased speed and a slight reduction in distances) while under the optimal cordoned mileage fee it decreased only slightly (-3%).

Nonetheless, the average generalized cost *burden*, which depends on the distribution of generalized travel costs by household incomes, is heavily affected by the mode shifts from driving and TNCs to comparatively slower modes, namely walking and public transit. The overcrowding of public transit vehicles resulted in large increases in walking mode shares across both optimal pricing policies. On average, walking trips across all scenarios incurred about double the generalized travel cost burden of driving alone.

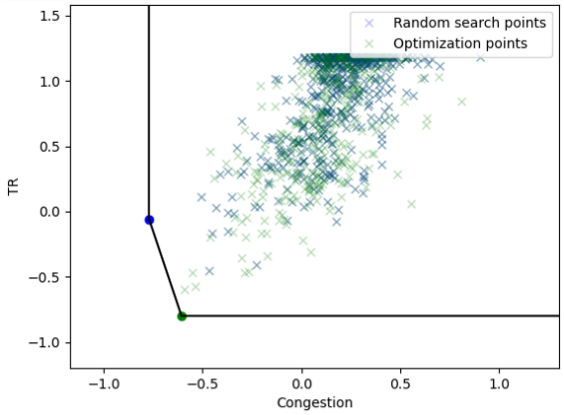
5.5 Discussion

This study of congestion pricing optimization demonstrates the complexity of policy-driven multi-objective optimization problems, particularly those with interdependent physical and behavioral dynamics and conflicting optimization objectives. The optimization of transportation systems with respect to both congestion and social equity objectives necessitates an iterative process to evaluate and refine the problem formulation, including the specification of input parameters, constraints, KPIs, and an overall objective function. While the Sioux Faux policy optimization scenario presented here is not intended to produce directly implementable policy recommendations, the study demonstrates the power of simulation-based optimization to generate granular insights about the individual- and system-level impacts of a multi-objective congestion pricing policy.

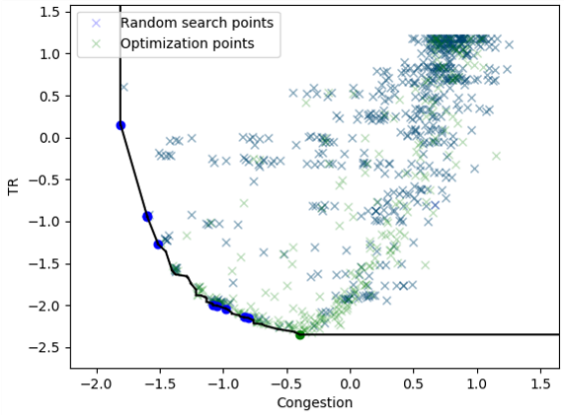
The use of ABS with an activity-based travel model for congestion pricing optimization enables the evaluation of the trade-offs between social and congestion KPIs. BISTRO facilitates the exploration of the policy input space and its relationship to KPIs as well as the relationship of objective function parameters with optimization results. The concavity of the relationship of total toll revenue with the average toll per driving trip was imperative for a feasible optimization within the context of this case study. Moreover, the investigation of scalarization parameters demonstrated the potential risk of overemphasizing congestion reduction in the optimization, which results in policies that produce unnecessarily large and potentially inequitable mode shifts away from driving.

Figure 5.6 displays the distributions of scores across all random search and optimization runs for each of the pricing policies with respect to pairs of score components[3]. The Pareto frontiers of each pair of score components are drawn and the Pareto points are highlighted. Across all three Pareto frontiers visualized, a comparison of the two pricing schemes reveals that the cordon toll constrains the range of the optimization problem to a subset of the

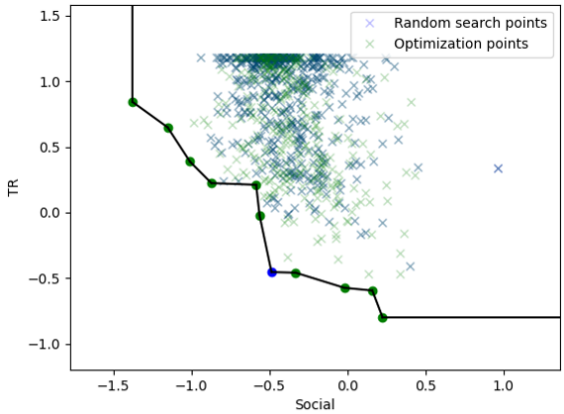
[3]The score components are standardized using the mean and standard deviation across all samples from both policies.



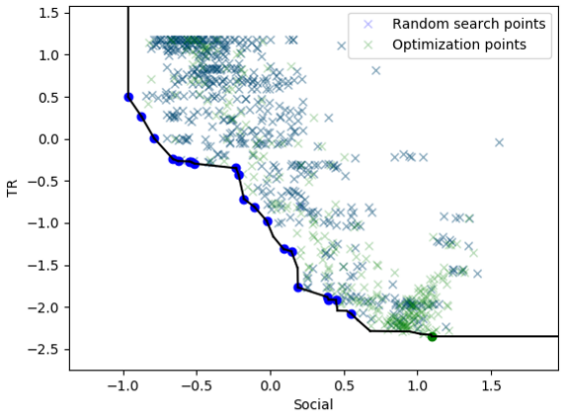
(a) Congestion-TR: Cordon toll



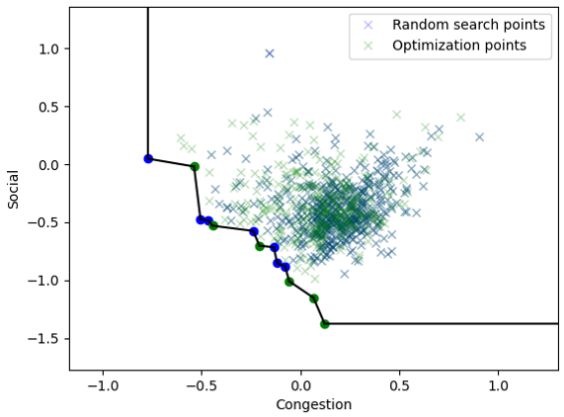
(b) Congestion-TR: Mileage fee



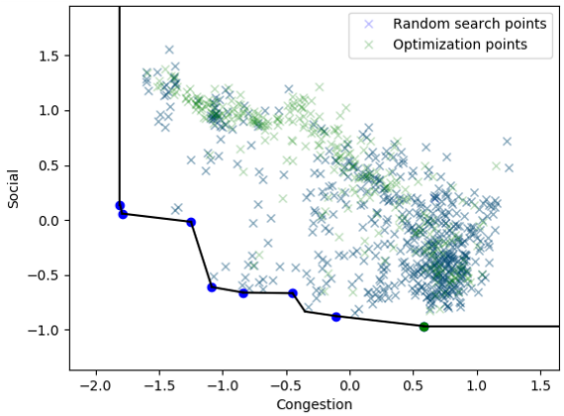
(c) Social-TR: Cordon toll



(d) Social-TR: Mileage fee

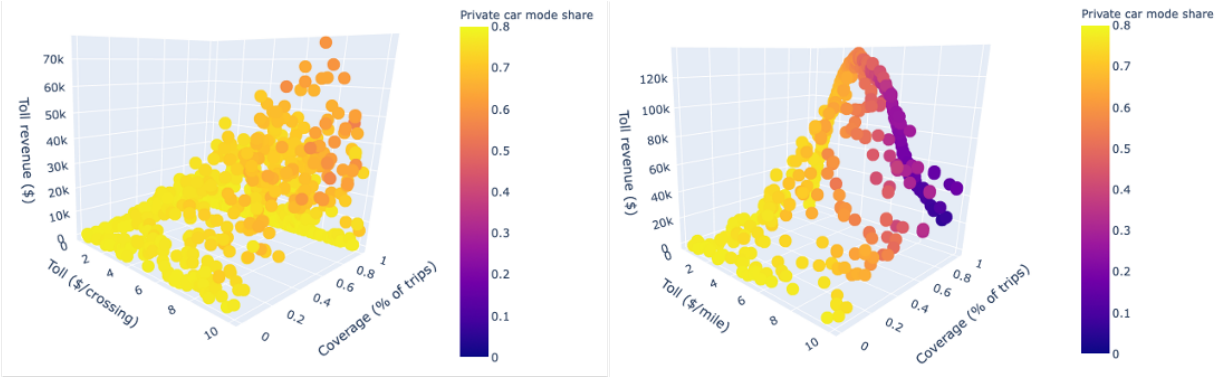


(e) Congestion-Social: Cordon toll



(f) Congestion-Social: Mileage fee

Figure 5.6: Pareto fronts of pairs of score components. Samples from the random search and optimization are shown in blue and green, respectively.



(a) Cordon toll

(b) cordoned mileage fee

Figure 5.7: Total toll revenue with respect to toll rate and cordon coverage.

range of the cordoned mileage fee optimization. This is consistent with the congestion pricing optimization literature, which has found that distance-based and time-varying schemes outperform simple cordon tolls with respect to social welfare [39, 54, 55].

This study also investigates behavioral responses to congestion pricing design. The cordoned mileage fee is found to be more effective in eliciting mode shifts among longer distance auto trips compared to the cordon toll, for which the majority of mode shifts are among trips less than 2 miles in distance. This results in greater VMT reductions under the cordoned mileage fee than the cordon toll, although the shift to slower modes (i.e., walking, public transit) results in greater average CB, across both work and secondary trips. Figure 5.7 displays the distributions of the total toll revenue from both random search and optimization samples for each pricing policy with respect to the toll rate and the geographic coverage of the cordons, calculated as the percent of all trips in Sioux Faux for which the toll applies. The color of the samples corresponds to the private car mode share (including driving alone and TNCs). Both plots reflect the behavioral response to pricing policies with higher tolls and greater coverage: the private car mode share decreases with respect to the toll rate and likelihood of being tolled. Both of the optimal pricing policies produced by the chosen weighting scheme in this case study actually produce less than the maximum toll revenue for each policy type. The maximum point shown in Figure 5.7a produces about $\$9,000$ more toll revenue than the optimum discussed previously, with a smaller cordon that charges about \$0.5 more per crossing and results in 60% private car mode share (compared to 65% in the optimum). The difference is smaller for the mileage fee, where the maximum toll revenue is just $2,800$ greater with a rate that is just \$0.1 greater per mile and just 0.5% less private car mode share compared to the optimum discussed previously.

Further refinement of the KPI formulation and scenario configuration are recommended for future studies. In particular, the prevalence of replanned public transit trips should be addressed for more realistic applications of BISTRO. Careful calibration of the capacity of the public transit fleet is recommended to ensure more accurate representation of pubic

transit capacity. Furthermore, the KPIs may be altered to better reflect the inconvenience of over-crowded public transit resulting in unrealistic replanned trips. The BISTRO optimization library includes a public transit crowding KPI, which measures the average time spent by public transit passengers on vehicles filled over seating capacity. Such a KPI can serve to disincentivize policies that result in overly crowded transit vehicles. In addition, the average generalized cost burden KPIs may be adjusted to provide a stronger negative signal in response to lengthy public transit and walking trips by increasing the VoT multiplier applied to walking and waiting time. Such an adjustment is supported by the travel behavior literature, which has found that travelers are significantly more averse to waiting time than they are to walking time, which in turn is more undesirable than in-vehicle time [19]. Nascent literature on automated vehicles also suggests that an adjustment of the VoT for in-vehicle time in a TNC would be appropriate, as the ability to be productive and/or access entertainment while riding in a vehicle as a passenger may provide more value than in-vehicle time spent from driving.

In addition, there are several improvements to the travel demand models used in this study that should be considered for future work. The exogenous nature of agents' travel plans precludes BISTRO's ability to optimize system interventions with respect to the impact that they may have on the quantity, time, and location of travel demand. Research on the impacts of roadway expansion have found that congestion reductions from increased road capacity induce additional travel demand by individuals able and willing to take advantage of the faster travel speeds to make trips they may not have made previously [155, 156]. Further research is needed to understand the extent to which congestion pricing may also induce travel demand. Since congestion pricing both reduces travel times and increases the cost of driving, the magnitude and distribution of demand that may be induced is not clear. Additional car trips may be induced among drivers or TNC users with high willingness to pay the congestion charges to access reduced car travel times (either due to income or circumstance). Congestion pricing may also induce additional public transit and active transportation trips due to the reduced volume of cars making public transit faster and active transportation safer.

Thus, by keeping travel demand fixed, the optimization results produced by BISTRO may a) overestimate the long-term reductions in congestion and emissions and b) underestimate the social welfare benefits from increased mobility. The simple travel model used in this study may be improved by including secondary trips in all agents' travel plans, with a 'no-travel' option included in the mode choice model. This would enable employed travelers to choose to work from home, chain together a secondary trip with their commute tour by making a stop on their way to/from work or during the workday (e.g., going out for lunch or running an errand), or even take another trip before or after the work day (e.g., going out for dinner, running errands, etc.). It would also enable non-employed travelers to travel at various other times of the day for various reasons. The resulting congestion pricing optimization would be more robust to short-run sensitivities in travel demand. More complex model development is necessary to integrate a feedback loop between the simulation outputs and the travel plan generation model to reflect long-run sensitivities, including geographic and temporal changes

in home, work, and other activities.

5.6 Conclusion

As congestion pricing policies are increasingly considered for adoption across the globe, the optimal design of congestion charging zones and pricing structures will play a crucial role in determining the success and acceptability of these congestion mitigation strategies. The use of ABS and activity-based travel models to evaluate a broad spectrum of design parameters offers many advantages over traditional theoretical and scenario-based analyses. The methodology presented in this study demonstrates the capability to explore a variety of policy design parameters and discrete outcomes, including the distribution of behavioral responses, congestion effects, and revenue generation. The presented framework can aid in the process of designing and evaluating precise objectives and weights to be applied to the optimization of public policies in partnership with various stakeholders.

Chapter 6

Congestion Pricing Optimization Case Study II: San Francisco Bay Area

6.1 Introduction

This chapter builds upon the previous chapter with a refined methodological approach for the multi-objective optimization of a cordoned mileage fee and multi-modal incentive scheme in a calibrated model of the San Francisco Bay Area. It was written in collaboration with Jarvis Yuan, Susan Shaheen, and Alexandre Bayen and is in preparation for submission for publication. I led the study design, including the development of the methodological and analytic approaches, oversaw the implementation and execution of the study, and conducted all analysis presented in this chapter.

Of the three pillars of sustainability, the social implications of congestion pricing have historically taken a backseat while the economic and, more recently, environmental implications have driven technological, operational, and political developments in the field. The London Congestion Charge was introduced in 2003 to reduce traffic congestion, which was 'seen as directly contributing to increasing economic prosperity, making London liveable, increasing London's accessibility and reducing environmental degradation' [157]. Similarly, the Stockholm congestion charging trial was launched in 2006 to 'test whether the efficiency of the traffic system could be enhanced by congestion charges' [158].

In recent years, the stated objectives of congestion pricing strategies have gradually expanded to include considerations for transportation *equity*, which focus on the distribution of costs and benefits realized by the transportation system across various dimensions of the population. For example, the Los Angeles Traffic Reduction Study, which is assessing potential options for a congestion charging zone in the Los Angeles area, aims to 'develop an Equity Strategy that considers reinvesting congestion pricing revenue as a key source of funds to minimize economic impacts to low-income drivers' [159]. The Downtown Congestion Pricing Study being conducted by the San Francisco County Transportation Authority (SFCTA) also explicitly states that 'advancing equity by improving health and transporta-

tion for historically underinvested communities' is among its key goals [160]. SFCTA also asserts that discounts and incentives are essential to implementing a fair congestion pricing system.

The emergence of on-demand ride services, including transportation network company (TNC, also known as ridesourcing and ridehailing)[1] (e.g., Lyft, Uber) and microtransit services[2] (e.g., Via, Chariot, Bridj), have also driven recent interest in pricing strategies to mitigate for the congestion produced by these services [19]. Although on-demand ride services have the potential to reduce the negative impacts of road transportation while improving accessibility for underserved communities by offering flexible, affordable, and convenient alternatives to auto ownership, TNC services have been found to contribute to increasing vehicle mileage, traffic congestion and greenhouse gas (GHG) emissions, in part due to the relative portion of miles driven without passengers, known as *deadheading* [6, 8, 9, 10, 11, 19, 161].

In addition to compelling travelers to internalize the costs of the externalities of driving, congestion pricing is expected to encourage the use of *pooling*, in which multiple travelers making similar journeys share a ride in the same vehicle. With the growth of app-based carpooling, microtransit, and on-demand pooled ride services, there are more service options than ever to facilitate dynamic pooling in addition to traditional forms of pooling such as public transit (PT), casual carpooling (also known as slugging), and vanpooling [19, 162]. However, even prior to the suspension of many existing pooled ride services during the COVID-19 pandemic, the rates of pooled ride requests among on-demand ride service users were relatively low, with pooled ride requests making up less than 40% of TNC rides in 2017 [17].

While flat fee congestion pricing (e.g., cordon pricing) can increase the financial incentive to pool, mileage-based pricing can create an even larger incentive for microtransit, which reduces the vehicle miles travelled (VMT) of any particular trip by requiring riders to walk some distance to/from a pickup/dropoff location in addition to distributing the cost across a larger number of riders. Moreover, by reducing the number of vehicles on the road, congestion pricing is expected to improve the level of PT service and generate revenue that can support the expansion of service and/or further subsidization of PT services [163].

There are several potential negative equity implications of congestion pricing strategies to be considered. While many of the positive environmental and economic benefits of fleet-based ride services depend on the rate of pooling, the social benefits depend on the continued accessibility and affordability of the services for otherwise underserved communities. The congestion pricing optimization literature has found that pricing policies are only progressive (i.e., the costs of the policy scale up with individuals' ability to pay) in so far as the revenue generated is redistributed in a manner that disproportionately benefits travelers most likely to be dissuaded from driving. In the context of on-demand ride services, research has found

[1]TNC services provide travelers with pre-arranged and/or on-demand access to a ride using a digitally-enabled application or platform (e.g., smartphone apps) to match riders to drivers for a fee.

[2]Microtransit refers to technology-enabled transit services that typically use shuttles or vans to provide pooled on-demand transportation with dynamic routing.

that lower-income people of color (POC) without access to personal vehicles are among the most frequent users of TNC services [19]. While frequent users of TNC services are more likely than other travelers to have multimodal travel profiles wherein they use a number of transportation modes, including PT, they are not necessarily using TNCs to access PT [19, 161]. In other words, TNCs are often used as an alternative to PT, suggesting that there are times, places, and trip contexts for which the relative speed, convenience, and other factors of TNC service outweighs the higher cost in comparison to PT services. This is particularly the case for historically underserved communities who are using TNCs to travel to essential activities such as work, groceries, and other errands [19].

This chapter explores the potential to design congestion pricing strategies that achieve the intended objectives of reducing congestion and pollution while maintaining or even improving equitable access to mobility in a metropolitan region with both ride alone and pooled on-demand ride services available. A simulation-based optimization approach is applied to a case study of the San Francisco Bay Area, where a congestion pricing scheme for the City of San Francisco is currently under development. The study is inspired by the SFCTA's Downtown Congestion Pricing Study, which has been assessing the potential features and feasibility of a congestion pricing scheme in downtown San Francisco since as early as 2008 and has most recently outlined two potential pricing zones (see Figure 6.1) along with an income-based cordon tolling structure (where drivers are charged upon entry to the zone during weekday rush hours[3]) that applies to all private vehicles, including personal vehicles and TNCs [160]. The pricing scheme is expected to reduce travel times, increase safety for all road users, reduce pollution, and advance equity by "improving health and transportation for historically underinvested communities" [160]. In addition to income-based discounts and free access for very low-income drivers, a 50% discount is proposed for drivers with disabilities and discounts are under consideration for residents living within the charging zone, drivers already paying a bridge toll to enter the city, and low income PT users.

The *Berkeley Integrated System for TRansportation Optimization* (BISTRO) is used to conduct the simulation-based optimization of a multi-cordoned congestion pricing scheme with income-specific subsidies using a calibrated activity-based travel model of commuter behavior in the San Francisco Bay Area. BISTRO facilitates the algorithmic optimization of transportation system intervention strategies (e.g., PT scheduling, changing vehicle fleet mixes, and altering rate structures for various modes) using the Behavior Energy Automation Mobility (BEAM) agent-based simulation framework to simulate and measure key performance indicators (KPIs) of alternative scenarios and intervention strategies. BISTRO enables the exploration of a larger domain of multi-dimensional transportation system interventions than is possible under traditional scenario-based travel demand forecasting approaches, which this study leverages to generate a more complete understanding of the range of potential economic, environmental, and social outcomes of congestion pricing at both the individual and regional levels. I implemented a Pareto-based Bayesian optimization algorithm to determine the optimal locations and charging rates of a circular charging zone in

[3]The proposed charging periods are from 6 am to 9 am and from 3:30 pm to 6:30 pm.

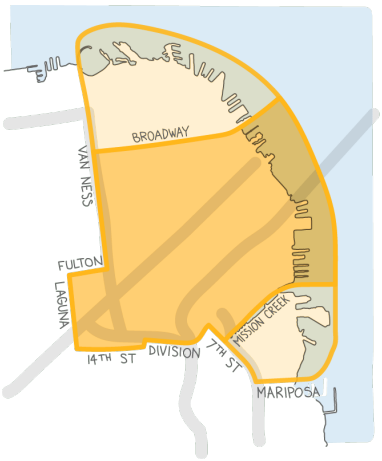


Figure 6.1: SFCTA draft congestion pricing zone(s) [160]

the City of San Francisco as well as income-specific subsidies for pooled modes (including pooled on-demand rides and PT). The Pareto-based algorithm seeks to maximize the Pareto frontier - the set of optimal trade-offs - between KPIs of the operational (total VMT, average vehicle delay (VD)), environmental (total GHG emissions, total $PM_{2.5}$ emissions), economic (net public revenue(NPR)), and social (average travel cost burden (CB), average PT vehicle crowding (TC) experienced) outcomes estimated via simulation. Analysis of the key relationships between the design parameters of the congestion pricing and incentive scheme with the KPIs reveals the significant impact that each parameter has on the congestion reduction, equity, and financial sustainability of the strategy. Incentives are found to contribute positively to all objectives except for NPR and the examination of Pareto-optimal strategies demonstrates the challenge of optimizing the design of a multi-faceted TDM strategy with respect to multiple competing objectives, generating key insights for future research.

The following section provides background on the congestion pricing optimization literature. Next, section 6.3 details the methodology of the case study, including an overview of BISTRO, the travel model development and calibration, and the optimization approach applied. Results are presented next, followed by discussions of study limitations, policy implications, and future research directions.

6.2 Background

Congestion pricing is a road pricing strategy devised to elicit a reduction in the volume of vehicles in congested areas and/or time periods in order to improve travel time reliability and reduce environmental pollution. The secondary goals of congestion pricing include the generation of public revenue, the improvement of PT service, and increased safety within the charging zone [157, 160, 163]. Zonal congestion pricing, including cordon- and area-

based pricing schemes, charges vehicles for entry into or movement within a designated area, respectively. Area-based pricing includes both flat rate (e.g., the London Congestion Charge) and distance-based pricing schemes (e.g., Singapore's Electronic Road Pricing 2.0). The following subsections provide an overview of the academic literature on congestion pricing optimization, including considerations for and lessons learned regarding demand sensitivity, pricing structure and cordon design, optimization objectives, optimization approaches, and equity analysis.

Congestion Pricing Optimization

The optimization of congestion pricing schemes focuses on determining the optimal rate to charge under various assumptions regarding the behavioral and physical dynamics of the transportation system. Congestion pricing optimization studies have gained complexity over the years, from purely analytical models of vehicle dynamics on a single road or in a simplified road network to regional network-scale studies optimizing dynamic link-based tolls using simulation-based approaches. The following subsections provide an overview of the various aspects of the congestion pricing optimization problem as studied in the literature, including the specification and analysis of various dimensions of demand sensitivity, pricing structure and cordon design, optimization objectives, optimization approaches, and equity analysis.

Demand sensitivity

The sensitivity of travel demand to pricing is essential to the analysis and optimization of congestion pricing schemes. The congestion pricing literature primarily focuses on the sensitivity of facility demand (i.e., the flow of vehicles on individual road links) and, less prominently, the sensitivity of modal demand (i.e., the flow of individuals using each travel mode). The travel demand profile defining the quantity of individuals traveling between each origin-destination (OD) pair in a road network at any given time is often considered a fixed input. Together with a model of traffic dynamics in the network, road users' route and mode choices are typically determined according to assumptions of utility-maximizing (i.e., cost-minimizing) decision-making based on the estimated costs of each alternative. In the simplest cases, travel time and cost are traded off using a value of time (VoT) parameter which is assumed to be uniform across the population. However, several studies have demonstrated that ignoring heterogeneity in VoT can result in drastic underestimation of the of the welfare benefits of optimal tolling policies [34, 35, 36, 37]. In addition to being a key determinant in an equilibrium distribution of routing choices, heterogeneity is central in determining the elasticity of demand with respect to travel time (e.g., departure time, arrival time), mode (e.g., PT, walking, biking, choosing not to travel), and even longer-term decisions such as home or work location choices that may also be sensitive to the distribution of costs in a transportation network [35, 38, 39, 40, 41, 42, 43].

The inclusion of multiple travel modes and mode choice behavior in congestion pricing optimization research has demonstrated that road pricing can induce a "virtuous cycle" of

mode shifts, PT service level improvements, and operating cost reductions for both users and PT operators that can result in considerably greater improvements to consumer surplus than is projected by models of the impacts to private vehicles alone [35, 38, 40, 43, 44]. However, the distribution of benefits relies heavily on the portion of travel demand that is dependent on PT as well as the potential of the congestion pricing scheme to significantly reduce PT costs through reduced congestion [40].

Traditional carpooling, as incentivized by high occupancy vehicle (HOV) and high occupancy toll (HOT) lanes has been examined in the congestion pricing optimization literature, typically in comparison to first-best pricing which tolls all lanes on a highway. In a highway setting, price differentiation between single-occupant vehicles (SOV) and HOVs (e.g., by allowing HOVs to accessed a tolled lane for free) can achieve most of the gains of first-best pricing while lessening some of the undesirable distributional effects of ubiquitous tolling [46, 164]. However, the magnitude of potential benefits from carpool exemptions depend heavily on the initial share of carpooling in the study area and in practice. HOV to HOT lane conversions have had negligible and even slight negative effects on carpooling and the share of HOVs on a facility is generally insensitive to travel time savings [46, 50, 51].

More recently, the adoption of TNCs and the development of automated vehicle (AV) technology has spurred a renewed interest in the potential impacts of pooling rates on overall transportation system efficiency. Ostrovsky and Schwarz (2019) postulates that the provision of pooled on-demand ride services by shared AV (SAV) fleets in the presence of ubiquitous tolling will produce the optimal conditions to achieve the economies of scale in pooling demand necessary to support high match rates [18]. However, it remains to be seen whether pricing mechanisms and automation alone will engender sufficient demand for pooling, particularly given several key socio-economic issues including the willingness to share a ride, curb management, social equity, personal safety, and labor considerations [52].

Finally, consideration for the sensitivities of departure time and the frequency of travel across trip purposes and destinations are also important in congestion pricing optimization in order to reflect the relative impacts of pricing schemes on shifts in preferred travel times, modes, and/or purposes in response to the changes in the travel times and costs of each alternative over time [38, 165]. In a post-hoc analysis of the model forecasts for the Stockholm congestion charge, [165] found that the actual impacts on off-peak travel were underestimated by five to ten percentage points due to the larger than predicted effect of the charge on leisure trips. Even though the travel model used for the design and forecasting of the Stoccckholm congestion charge included a nested mode and destination choice model providing a feedback loop between the simulated travel times and costs of the congestion charge and the travel demand profile, the omission of trip-chaining behavior in the travel model may have contributed to the underestimation of leisure trips that were foregone due to the shift away from driving for work trips (i.e., leisure trips that used to be easier to make by car on the way to/from work that were no longer desirable). In addition, about 80% of the affected leisure trips were shifted to different destinations outside of the cordon, which also contributed to a greater than expected reduction in congestion during off-peak hours.

Pricing structure and cordon design

The optimization of congestion pricing schemes may also entail the design of the charging zone itself. Network-scale studies predominantly formulate link-based tolling schemes (i.e., an optimal toll level is determined for each link in the network). While it is possible that the optimization of link-based tolls may result in schemes with discernible tolled and untolled areas, it is unlikely that such a result would be produced without the application of explicit constraints on the optimization problem. Examination of spatially and/or geometrically constrained tolling schemes is valuable for understanding the trade-offs in infrastructure investment decisions for pricing schemes that are more technically and practically feasible than unconstrained link-based schemes [42, 44, 53].

In comparison to traditional cordon tolls in which drivers are charged upon entry to a charging zone, area-based, multi-cordon, and distance-based schemes have been found to produce greater social welfare gains by increasing the coverage of the tolling zone and achieving a greater correlation with the marginal congestion contributed by each tolled vehicle [39, 42, 45, 54]. Generally, higher toll levels produce greater reductions in driving which lead to greater social benefits in terms of reduced travel time [42], as does greater coverage of the charging zone [37, 42, 45]. Distance-based and time-varying schemes also outperform simple cordon tolls with respect to social welfare [39, 54, 55]. However, greater complexity and higher toll rates generally lead to lower public acceptance rates [42]. Qualitative design objectives for politically feasible pricing schemes should emphasize simplicity and ease of understanding in addition to aiming for a charge level that is acceptable and perceived as fair to the public [42].

Equity analysis of congestion pricing

Many studies have assessed the equity of outcomes from congestion pricing schemes, both in theory and in practice. Distributional analyses of the travel times and/or tolls paid by heterogeneous road users under optimized congestion pricing schemes find that lower income road users are generally worse off than higher income users as a result of lower and higher values of time, respectively [37, 45, 48, 56]. The location of priced roads relative to the spatial distribution of travelers also plays a significant role in the equity outcomes of road pricing, particularly in geographies with significant levels of spatial segregation [45, 56, 48]. Yang and Zhang (2002) propose constraints to bound the relative percentage changes in generalized travel costs across OD pairs and income groups according to a specified level of inequity. However, across all examples shown, the optimal tolls still produce the most negative effects on the lowest income groups while high income groups are impacted the least, and in some cases, receive benefits [48]. The simulation-based optimization of cordon tolls and cordoned mileage fees in chapter 5 finds that the average generalized travel cost burden in a benchmark travel model generally increases with respect to congestion reduction, particularly in the case of a cordoned mileage fee which levies a greater average toll rate on driving trips than does a simple cordon toll.

Optimization objectives

The design of optimal tolls is inherently dependent on the definition of objectives used in the optimization. A vast majority of the literature defines optimality with respect to social welfare, typically defined as the total user benefit (e.g., the total generalized travel cost) minus the total social cost (e.g., the total travel time plus the total toll revenue) [34, 36, 45, 46, 56]. Few studies consider other objectives for toll optimization. Pricing optimization studies of managed lanes tend to consider throughput and/or revenue maximization as primary objectives, which follow from assumptions of the purpose of such facilities [37, 57]. The correlation of efficiency objectives with environmental objectives depend strongly on the modeling assumptions made. Some studies find that minimizing congestion results in emissions reductions due to the reduction in VMT achieved by mode shifts [42, 58]. The congestion pricing optimization study presented in chapter 5, finds that GHG emissions and total VMT are directly correlated. Yet other studies conclude that the two objectives are at odds, although optimal tolling schemes with respect to specified trade-offs between congestion and emissions are feasible [166].

Few congestion pricing optimization studies have considered transportation equity objectives. Jalota et al. (2021) establish the existence of optimal tolls and revenue refunding schemes with respect to a wealth inequality measure that satisfy a user favorability condition [167]. Yin and Yang (2004) formulate two equity objective functions for congestion pricing optimization including the minimax of individual utilities for the lowest VoT road user and minimization of a Box-cox transformation of utilities that minimizes the gaps in utilities across road users. However, neither of these two options are explored empirically. Lastly, the congestion pricing optimization study in chapter 5 includes an average cost burden KPI as a measure of social equity in a weighted multi-objective function, finding that this metric is directly at odds with congestion-related KPIs.

Optimization approaches

Low-dimensional models of pricing optimization are often employed for the purposes of simplicity and interpretability. Such models can be solved numerically using mathematical approaches. However, in higher dimensional networks, computing the optimal toll and traffic assignments using numerical methods becomes infeasible. A multitude of methods are used to overcome the challenge of simultaneously optimizing road user choices and toll optimization in order to compute an equilibrium, including heuristic algorithms [37, 47, 53], genetic algorithms [61, 168], and macroscopic fundamental diagram (MFD) models [39, 62, 169].

A few studies have employed agent-based simulation (ABS) engines for this purpose, which explicitly model the decisions and movements of individual road users throughout discrete and/or continuous time horizons. May and Milne (2000) demonstrated the importance of considering rerouting effects in congestion pricing design by utilizing a static traffic assignment and simulation framework [54]. De Palma, et al. (2005) used the METROPOLIS simulator, which incorporated departure time, mode, and route choice models with uniform

VoTs [55]. A combination of heuristics and response surfaces were utilized to iteratively update and simulate the tolling schemes until converging to optimal solutions for six variations of link-tolling schemes. Chapter 5 uses BISTRO to conduct a simulation-based optimization with the *Behavior Energy Automation Mobility (BEAM)* ABS framework, which incorporates within-day decision-making including departure time, mode, and routing choices as well centrally managed on-demand ride services. The Tree-Structured Parzen Estimator (TPE) algorithm, a Bayesian optimization algorithm, was used to optimize the location and size of circular cordons as well as the charging rate. This type of approach is considered a surrogate-based technique, in which the optimization is approached as a black-box problem with stochastic input-to-output relationships estimated via simulation. Other surrogate-based optimization approaches have been applied to optimize tolls using dynamic traffic assignment focusing on highway tolls [63, 64].

6.3 Methods

BISTRO

The *Berkeley Integrated System for TRansportation Optimization (BISTRO)* is an open-source transportation planning decision support system that leverages the power of ABS, activity-based travel models, and machine learning to assist stakeholders in addressing the increasingly complex problems arising in transportation systems worldwide [30]. Given a set of discrete policy inputs, BISTRO alters the "business as usual" (BAU) configuration of a BEAM scenario and post-processes the resulting simulation outputs into KPIs used in a user-defined objective function for optimization. BISTRO also includes a dashboard of visual analytic tools to analyze and compare the outcomes of discrete strategy alternatives.

BEAM is a multi-modal activity-based travel demand and ABS framework developed at the Lawrence Berkeley National Laboratory (LBNL) [170]. BEAM simulates the mobility of hundreds of thousands of individual agents representing a synthetic population of urban travelers. Each individual traveler is endowed with socio-demographic characteristics (e.g., gender, age, income, vehicle ownership, etc.) and a daily travel plan denoting the locations and desired start times of various activities (e.g., "Home", "Work", "School", etc.). Agents select a travel mode (e.g., drive, walk, bus, etc.) with which to complete their desired trips that are then executed in a discrete-event mobility simulation. At the end of the simulation period, travelers' executed plans are scored based on the duration and cost of travel. In an iterative process, a subset of agents are randomly selected to undergo replanning, or mutation, of their travel plans that may alter the activity start times, mode choices, and/or route choices based on the realized travel times from the prior iteration. The altered plans are then applied in another iteration of the day-long simulation. The process of replanning, simulation, and scoring is executed until the simulation converges to a stochastic equilibrium in which agents are no longer able to improve their scores through replanning.

BEAM also includes *within-day* replanning dynamics by including an online discrete

choice model of agents' mode choice decisions that depends on the travel times and costs of each available mode estimated *during* the simulation. This enables agents to make unplanned time-sensitive utility-maximizing decisions that can be affected by both the network and market resource dynamics of the transportation system. Mode and route choices are thus responsive to network and resource market dynamics which are central to determining the level of service of on-demand fleet-based transportation services such as TNCs. On-demand ride services are rebalanced, reserved, and routed in real-time by a centralized fleet management algorithm throughout the simulation.

In addition, BEAM includes multi-modal trip planning in which agents explicitly choose whether to walk, drive, or use an on-demand ride service to access PT. Moreover, the within-day re-planning dynamics enable responsiveness to the real-time resource capacity constraints of fleet-based services. In the case that an agent encounters an overfull PT vehicle or a longer than expected wait time for the PT or on-demand ride service they planned to use, the agent undergoes another mode choice event to re-plan how to complete the trip already in progress. For example, if an agent misses the bus or is unable to board a full vehicle, it may decide to reserve an on-demand ride to their destination rather than simply wait for the next bus to arrive, depending on the estimated wait time, in-vehicle time, and costs of each option at that time in the simulation.

San Francisco Bay Area scenario

For the purposes of this case study, I adapted an existing activity-based travel model representing a single day of commute trips in the nine-county San Francisco Bay Area - including San Francisco, San Mateo, Santa Clara, Alameda, Contra Costa, Solano, Napa, Sonoma, and Marin Counties. The model includes a simplified road network of about 70,000 nodes and 180,000 links wherein local roads are aggregated in order to reduce the computational burden of the simulation. PT services across the region are simulated using open-sourced General Transit Feed Specification (GTFS) data from each PT agency in the region. TNC service is simulated using a single fleet management operator with a demand-following repositioning algorithm. Furthermore, the model consists of a synthesized population of 3,157,000 commuters. The population synthesis, including the generation of individuals, households, home and work locations, and preferred activity start times, was conducted using UrbanSim, a regional land use and transportation model that simulates housing and employment choices [104]. The distribution of employment density per census tract as reported in the 2019 American Community Survey (ACS) [171] and modeled in BEAM are displayed in figures B.1a and B.1b, respectively. As shown, the distribution of the synthesized population in the BEAM model is generally similar to that of the actual population, with the greatest density in the City of San Francisco and in the urban centers of the East and South Bay Area.

Each household in the synthesized population was endowed with two cars and two bicycles. While this approach overestimates personal car and bicycle access in the San Francisco

Bay Area[4], it enables, in a simple manner, flexibility in mode choices that may reflect the potential impacts of pricing strategies on vehicle purchasing decisions. In order to enable analysis of the impacts of congestion pricing strategies on fuel consumption and emissions from vehicle use, data from the 2017 National Household Travel Survey (NHTS) and the California Air Resources Board (CARB) were used to configure the personal and TNC vehicle fleet mixes, respectively [9, 172]. The fuel type of each household car was randomly generated from the distribution of fuel types corresponding to the household's annual income, which was estimated from the NHTS (see Figure 6.2). The TNC vehicle fleet mix was configured more simply, by generating the fuel types of the TNC vehicles from the distribution of vehicle fuel types estimated from a 2018 CARB TNC driver survey in the state of CA: 67% gas-powered, 26% hybrid, and 7% plug-in hybrid (PHEV) vehicles. This distribution is similar to that found by another survey of TNC drivers in the San Francisco Bay Area that was conducted in 2016, which estimated a higher portion of conventional gas-powered vehicles and lower portion of PHEVs [161]. The fuel type and consumption rates of each vehicle type in the model are documented in table B.1 in the appendix. In addition, all vehicles in the TNC fleet were programmed to have a seating capacity of 7 passengers in order to enable analysis of the potential for generating demand for higher-occupancy pooled on-demand rides than are currently offered by TNCs, such as microtransit services.

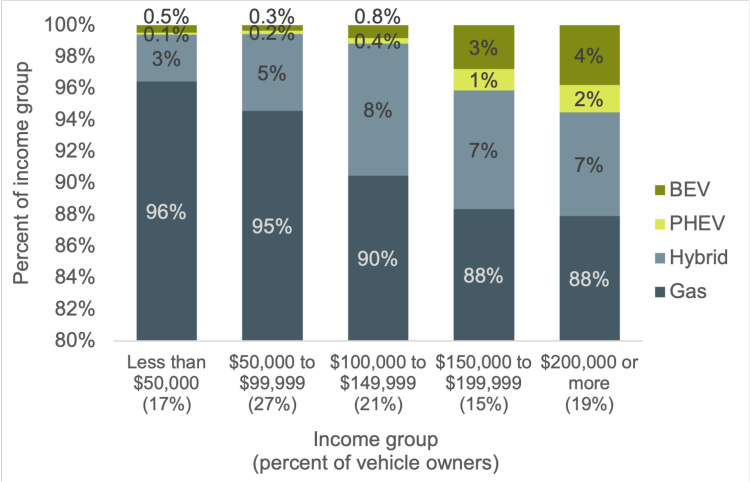


Figure 6.2: Personal vehicle fleet mix in the San Francisco Bay Area by income group [172]. Note, BEV stands for plug-in battery electric vehicles and PHEV stands for plug-in hybrid electric vehicles.

The simulation of the full SF Bay Area model takes about 8 hours to run on a virtual machine using 128GB of memory. In order to achieve a tractable run time for optimization, which requires running the simulation hundreds of times with varying inputs, we constrained

[4]As per the 2017 National Household Travel Survey (NHTS), there are 1.90 cars per household in the San Francisco Bay Area; comparable data on bicycle ownership is not available [172].

the focus of this study to the most congested area of the region: the City and County of San Francisco. To do this, we randomly generated a sub-sample of 50,000 agents that commute to, from, or within San Francisco[5]. In this way, we also restricted the provision of incentives to travelers that live and/or work in San Francisco. While this limits the accuracy and granularity with which the simulation can predict the secondary effects of the congestion pricing scheme on commuters that do not travel within San Francisco, the aggregate changes in in- or out-bound vehicle flows to the City may still be analyzed for this purpose. The sub-sampled model of 50,000 commuters takes about 3.5 hours to run using a high performance computer with 64GB of memory.

The default mode choice model in BEAM was augmented for the purposes of this study to include various individual- and alternative-specific parameters, including: income-based VoT coefficients and individual-specific coefficients corresponding to age, income, trip origin, and destination. In BEAM, mode choices are simulated using a multinomial logit (MNL) model, defined by equations 6.3.1 and 6.3.2 below, where $U_{i,t,m}[n]$ is the utility of mode m for trip t by traveler i at time step n, including both systematic and random components, $V_{i,t,m}[n]$ and $\epsilon_{i,t,m}$, respectively, and $P_{i,t}(m|x[n])[n]$ is the probability that traveler i chooses one of the available travel modes $m \in C_{i,t}[n]$ for a particular trip t at time step n.

$$U_{i,t,m}[n] = V_{i,t,m}[n] + \epsilon_{i,t,m} \tag{6.3.1}$$

$$Prob(U_{i,t,m}[n] \geq U_{i,t,m'}[n] \ \forall m' \in C_{i,t}[n]) = \frac{\exp^{V_{i,t,m}[n]}}{\sum_{m' \in C_{i,t}[n]} \exp^{V_{i,t,m'}[n]}} \tag{6.3.2}$$

The systematic component of the utility is a weighted linear combination of the state of the simulation at that time step $x[n]$, the attributes of the trip $y_{i,t}$ and the individual characteristics of the traveler z_i. The random component is assumed to be independently, identically distributed extreme value. The systematic utility equation used in this study is specified in equation 6.3.3 and the parameters are defined in table 6.1 below.

$$V_{i,t,m}[n] = \beta^{asc}_m + x^{cost}_{i,t,m}[n] + \beta^{vot}(\beta^{mult}_{i,m} x^{in-veh}_{i,t,m}[n] + \beta^{wait} x^{wait}_{t,m}[n]) + \beta^{transfer} x^{transfer}_{m,t}[n] + \beta^{dest}_{i,m} y^{dest}_{i,t} + \beta^{age}_{i,m} z^{age}_i + \beta^{inc}_{i,m} z^{inc}_i + \beta^{veh}_{i,m} z^{veh}_i \tag{6.3.3}$$

The coefficients for the model were estimated using stated preference (SP) data from a general population survey of four metropolitan regions in CA (including the San Francisco Bay Area) that was distributed in fall 2018 [19, 173]. For detailed descriptions of the survey methodology and estimation of the mode choice model used in this study, the reader is referred to [19] and [173], respectively.

[5]This sub-sample is approximately 6.3% of the corresponding synthesized population of the full San Francisco Bay Area.

Description	**Coefficient**	**Parameter**	**Type**
Alternative-Specific Constant (ASC)	β_m^{asc}	1	Constant
Value of time (VoT)	β^{vot}		Multiplier
Estimated in-vehicle time (minutes)	$\beta_{i,m}^{mult}$	$x_{i,t,m}^{in-veh}[n]$	Continuous
Estimated wait time (minutes)	β^{wait}	$x_{t,m}^{wait}[n]$	Continuous
Estimated net trip cost ($)	1	$x_{i,t,m}^{cost}[n]$	Continuous
Transfers	$\beta^{transfer}$	$x_{m,t}^{transfer}[n]$	Scalar
Trip destination (home, work, transit station)	$\beta_{i,m}^{dest}$	$y_{i,t}^{dest}$	Categorical
Age (Under 30, 30-50, 50-70, over 70)	$\beta_{i,m}^{age}$	z_i^{age}	Categorical
Income (under $100k, $100k and above)	$\beta_{i,m}^{inc}$	z_i^{inc}	Categorical
Car ownership (none, 1 or more)	$\beta_{i,m}^{veh}$	z_i^{veh}	Categorical

Table 6.1: Mode choice model parameters for the San Francisco Bay Area scenario

The resulting activity-based model including 50,000 commuters living and/or working in San Francisco traveling on a PT network and simplified road network spanning the full San Francisco Bay Area is hereby referred to as SFSim. The following subsection details the methodology for calibrating the SFSim model. Following the calibration, which altered the parameters of the mode choice model, network, and TNC fleet configuration, all aspects of the scenario configuration were kept fixed throughout the remainder of the study.

Calibration

The calibration of the SFSim model included four phases, each of which focused on a particular set of configuration parameters and target metrics, as outlined in table 6.2 below. To the extent possible, all reference data is representative of travel conditions in 2018. In the first phase of calibration, the overall mode splits were calibrated by adjusting the intercepts (commonly referred to as alternative-specific constants (ASCs)) of the mode choice model. Mode split refers to the percent of trips that use each mode (i.e., the mode split of drive alone (DA) is the percent of all trips that are made by DA); overall mode split is the mode split across all trips while income-specific mode split is the mode split across travelers from a particular income group. In the overall mode split calibration, the intercepts β_m^{int}, of the utility equations for each mode were adjusted until the mean absolute error (MAE) with respect to mode splits estimated from the 2019 SFMTA Travel Decision Survey (TDS) was less than 3% (see Appendix B.1 for a description of the corresponding data and the estimation procedure) [174].

The next phase of calibration focused on adjusting the TNC supply so that the baseline ratio of deadheading miles (i.e., miles driven without a passenger in the vehicle) to total TNC VMT was close to the level that CARB estimated for the state of California in 2018: about 38.5% [9]. To do this, the TNC vehicle fleet size was adjusted until the deadheading ratio was within 15% of the target value.

Phase	Calibration parameters	Performance metrics	Threshold [Data source]	Business as usual (BAU) mean (sd)
1	Mode choice model intercepts	MAE of overall mode splits	$\leq 3\%$ [174]	0.63% (0.06%)
2	TNC fleet size	AE of TNC deadheading ratio	$\leq 15\%$ [9]	17% (0.0076%)
3	Link flow-capacity factor	Mean Z-score of peak average travel times by OD	≤ 1.96 [175]	1.72 (0.12)
4	Mode choice model intercepts and multipliers	MAE of overall mode splits MAE of income-specific mode splits	$\leq 1\%$ [174] $\leq 3\%$ [174]	0.63% (0.06%) 1.54% (0.17%)

Table 6.2: Overview of calibration methodology: Four-phase approach

Next, the travel speeds on the road network were calibrated by adjusting the flow-capacity factor $z^{flow-capacity}$ in the link travel time function, as shown in equation 6.3.4, which serves to scale the capacity of the network. The network speeds were calibrated with respect to average weekday peak period (7 to 10 AM and 4 to 7 PM) travel times for each OD pair of traffic analysis zones (TAZ) in the San Francisco Bay Area during Q3 2018, which was queried from Uber Movement [175][6]. In order to measure the performance of this calibration phase, we estimated the average z-score of the difference in mean OD travel times, with a target of 1.96 signifying that, on average, the simulated peak OD travel times were within the 95% confidence interval of the reference data.

$$x_l^{in-veh}[n] = z_l^{free-flow}\Big(1 + z_l^{\alpha}\big(\frac{x_l^{veh}[n]}{z^{flow-capacity} z_l^{capacity}}\big)^{z_l^{\beta}}\Big) \tag{6.3.4}$$

In the final stage of calibration, the overall and income-specific mode splits were calibrated by adjusting the mode choice model intercepts β_m^{asc}, income-specific VoT multipliers $\beta_{i,m}^{mult}$, and the mode-specific income coefficients $\beta_{i,m}^{inc}$ in the utility equation. The target mode splits of each of three income groups (less than $\$35,000$, $\$35,000$ to $\$100,000$, and $\$100,000$ or more) were estimated from the 2019 SFMTA TDS [174].

The simulation outputs of the calibrated model are hereby referred to as the business as usual (BAU) results; this serves as a baseline of comparison for the study. In order to account for the stochasticity of the model in the characterization of the BAU, we ran the calibrated model 10 times and estimated the mean and standard deviation of each metric that is used in the study, including the calibration metrics. The remainder of this subsection provides an overview of the BAU results and the key discrepancies in comparison to the calibration targets.

The distributions of the mean overall and income-specific mode splits across the ten runs of the final calibrated model are presented in figure 6.3 alongside the target distributions.

[6]The SFSim model includes a total of 1,332 TAZs.

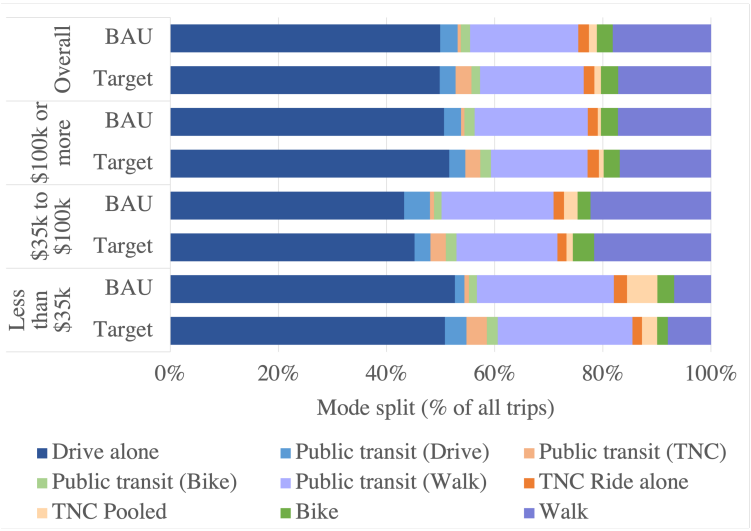


Figure 6.3: Comparison of target mode splits to BAU mode splits.

The mean MAE of the overall mode splits is 0.63% with a standard deviation (sd) of 0.06%. The biggest discrepancy is in the mode split of PT accessed/egressed via TNC, which is about 2.34% lower than the calibration target. The calibration process revealed that the mode choices in the SFSim model were relatively insensitive to the adjustment of the ASC corresponding to using PT with TNCs. This suggests that, in the BAU, the option to use TNCs to access/egress PT was significantly more undesirable than other alternatives, likely due to comparatively long travel times and high travel costs. Moreover, since all households have access to personal vehicles, it's likely that the option to drive to/from PT dominated that of using TNCs to access/egress PT in a majority of cases.

The mean MAE's of the income-specific mode splits were about 1.52% (sd of 0.10%), 1.39% (sd of 0.12%), and 1.73% (sd of 0.06%) for the low-, middle-, and high-income groups, respectively, amounting to a total MAE of 1.54% (sd of 0.17%) across all modes for all three groups. The biggest discrepancies between the BAU and the target mode splits for the lowest income group (about 11% of the population) were excesses of about 2.8% and 1.6% of trips using pooled TNCs (PTNC) and DA, respectively, and deficits of about 3.0% and 2.2% of trips using TNCs and DA to/from PT, respectively. For the middle income group (about 35% of the population), there was also a deficit in trips using TNCs to/from PT, though only by about 2.2% of trips. The mode splits of DA and biking were also short of the targets by about 2.7% and 1.6%, respectively, while those of driving and walking to/from PT were over the targets by about 1.8% and 1.7%, respectively. Finally, the biggest discrepancies in the mode splits of the highest income group (about 54% of the population) with respect to the targets were an excess of about 2.9% of trips walking to/from PT, and deficits of about 2.2% and 1.4% of trips using TNCs to/from PT and DA, respectively.

On average, about 4.0% (sd of 0.07%) of trips in the BAU were made by TNC (about

1.9% ride alone, 1.5% pooled, and 0.6% to/from PT). The calibrated TNC vehicle fleet size was about 10% of the population size and produced an average of about 20 VMT per vehicle. Extrapolating the total VMT by TNCs, the BAU produces about 865,000 TNC VMT. For comparison, the SFCTA estimated that TNC VMT within the city of San Francisco during November to December 2016 was about 570,000 miles, with an average of about 100 VMT per vehicle operating *within the city* [176]. The average ratio of deadheading VMT to total TNC VMT in the BAU results was about 0.56 (sd of 0.0076), which is about .17 above the target. The SFSim model likely overestimates TNC deadheading due to the exclusion of non-commute travel. Non-commute travel in the region likely contributes to better economies of scale of TNC service than in the SFSim model by enabling shorter distances between successive TNC ride requests serviced by any given vehicle, particularly outside of the City of San Francisco. The match rate for PTNC ride requests across the BAU runs was higher than estimated in the literature (about 22-23%), with about half of pooled ride requests resulting in a successful match between requests for similar ODs. The overall average number of passengers per TNC trip was about 1.13, which is lower than the estimate of 1.55 passengers per vehicle reported by CARB for the state of California in 2018 [9]. This is likely due to a few reasons: 1) the SFSim model focuses on commute trips, which have been estimated to make up 16.5% of TNC trips to, from, or within San Francisco [177], and 2) other trip purposes are more likely to involve travel by companions (e.g., two people requesting a private or PTNC ride together) or be in areas or during periods with a higher density of TNC activity (i.e., to/from commercial areas to go to a restaurant, bar, or other recreational activity).

Congestion pricing optimization problem specification

Next, we present the methodological approach for congestion pricing optimization using BISTRO and the calibrated SFSim model. We'll begin by defining the problem, including the policy input parameters and KPIs, followed by an explanation of the multi-objective evolutionary optimization algorithm implemented in BISTRO for this study.

Policy parameters

The congestion pricing scheme considered in this study includes a circular cordon with a mileage fee for driving or using TNCs within the cordon as well as a two-tiered incentive scheme for PTNCs and PT. The optimization algorithm is designed to determine the location and size of the cordon, the amount to be charged, and the income thresholds for eligibility for each of two levels of incentives for each of two mode categories: 1) PTNCs and 2) PT accessed/egressed by any mode. The two incentive levels for each of the two mode categories are also determined by the optimization.

Several constraints were applied to the input space to ensure that the pricing scheme was within the scope of the study and generally complied with realistic design principles:

1. The search space for the cordon center was constrained to a circular area including the most congested links in San Francisco, as estimated from the BAU results. The congestion level per link in the BAU was estimated by dividing the simulated vehicle flow by the capacity of each link within the San Francisco city boundaries in order to run a k-means clustering analysis on the top 5% of most congested links. Five clusters were classified using the coordinates of the link nodes; the three clusters with the greatest average congestion, displayed in Figure 6.4, were considered for the optimization search space. An algorithm was applied to fit a circle around the convex hulls of all links in each cluster, and iteratively adjust the circle radii to achieve non-overlapping search spaces for each cluster (see Appendix B.1 for a detailed explanation). We constrained the search space for the cordon center of this study to the most congested of these clusters (shown with a blue dot at its center in Figure 6.4), which was located near the cardinal center of the city, spanning from the Civic Center to the western edge of the Golden Gate Park and from Noe Valley to Japantown. This area includes the heavily trafficked Market Street, which carries traffic to and from the central business district (CBD) of the city, as well as the portion of the US 101 Highway that flows North/South through the Eastern part of the city. The second-most congested cluster (shown with an orange dot at its center in Figure 6.4) circumscribed the CBD of the city. While this is the area that has been the focus of SFCTA's congestion pricing exploration studies, the search space defined by the most congested cluster enables the consideration of cordons that would cover the CBD and/or other congested areas of the city. A cordon placed at the edges of this search space with a maximum allowable radius (see the next constraint) would almost cover the entirety of San Francisco. Constraining the cordon center to this area thus serves to reduce the size of the search space without compromising the ability to explore extreme solutions in which a mileage fee is applied to all or most of the city. In future work, we plan to utilize the other search spaces identified by the clustering analysis to optimize multiple cordons simultaneously.

2. The radius of the cordon was constrained between 500 m and 6 km.

3. The toll rate was constrained between \$0.25/mile to \$5/mile [7].

4. The first income threshold was constrained to annual household incomes between \$35,000 to \$100,000 and the second was constrained between \$50,000 to \$200,000. In addition, the second income threshold was constrained to be greater than or equal to the first.

5. The PTNC incentives were constrained between \$0 and \$20 per trip and the PT incentives were constrained between \$0 and \$10 per trip[8]. Finally, the second level of

[7] The search range for the toll rate was determined based on the findings of chapter 5.

[8] We note that, in the BAU, the average trip costs of PT and PTNCs across the whole study area were \$5 and \$19, respectively. Within San Francisco, the average costs of PT and PTNCs in the BAU were \$3 and \$11, respectively, while the average costs of trips to/from the City were \$6 and \$36 for PT and PTNCs, respectively.

the incentives for each mode were constrained to be lower than the first, ensuring a progressive incentive scheme (i.e., travelers earning below the lowest income threshold are eligible for greater incentives than those eligible foor the second level of incentives).

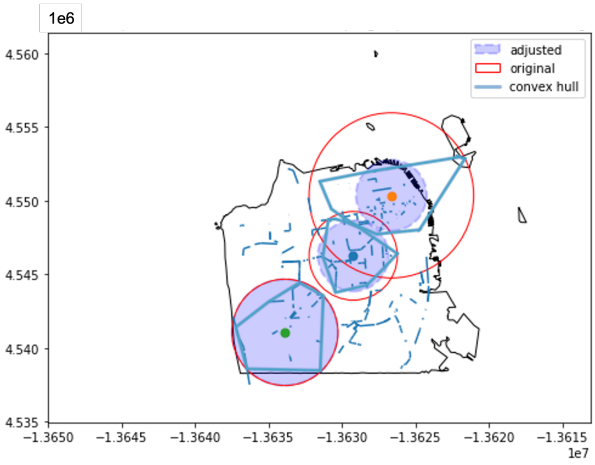


Figure 6.4: Three most congested clusters of links in the BAU simulation run.

Optimization objectives

Seven KPIs were considered for the optimization objective, representing various operational, environmental, social, and financial objectives of congestion pricing strategies, including average vehicle minutes of delay (VD) per passenger trip, total VMT, total GHG emissions (tank to wheel emissions), total particulate matter ($PM_{2.5}$) emissions, average minutes of PT crowding (TC) experienced per PT passenger, average generalized travel cost burden (CB) of travel, and the net public revenue (NPR) from PT fares, tolls paid, and incentives used. The specifications for each of these KPIs can be found in detail in Appendix A; a brief overview is provided here for general comprehension.

The average vehicle minutes of delay per passenger trip, a metric of congestion experienced by the average road user, is calculated as the average difference in the simulated travel times of passenger trips (i.e., DA and TNC trips) with the free-flow travel times (i.e., the estimated travel time on an empty road network) for those trips. The total VMT, GHG emissions, and $PM_{2.5}$ emissions are calculated using the total miles driven by each vehicle fuel type and the corresponding fuel consumption and emission rates (see Appendix B.1). The average TC experienced per PT passenger is a PT user-focused congestion metric designed to reflect the potential negative impacts of overcrowding the PT system. It is calculated as the average perceived VoT per trip on-board PT vehicles using VoT multipliers that scale with the degree to which a PT vehicle is filled over seating capacity. Since adjustments to PT supply are not in the scope of this study, the average TC KPI enables analysis of the

trade-offs between the immediate benefits that may be achieved by levying a pricing strategy and the need to expand PT service in order to maintain a desirable quality of service. Next, the average generalized CB of travel is a transportation equity metric, which is calculated as the weighted average of the generalized travel cost (i.e., the sum of the net cost of travel and the product of the population-average VoT and the duration of travel) for each trip, with the weights being the inverse of the annual household income of the trip-maker. Thus, the generalized travel costs of lower-income travelers weigh more heavily in this metric than do those of higher-income travelers and strategies. Finally, NPR of the pricing strategy is computed as the sum of total PT fare revenue and toll revenue subtracted by the total amount of incentives used.

Problem formulation

To summarize, the optimization of the congestion pricing and incentive scheme described in the previous two subsections is formulated in the following stochastic optimization problem:

$$\min_{\mathbf{d}\in\mathbb{R}^{10}} f(\mathbf{d},\mathbf{x};\mathbf{z}) = \mathbb{E}[\mathbf{F}(\mathbf{d},\mathbf{x};\mathbf{z})] \tag{6.3.5}$$

such that

$$\mathbf{x} = B(\mathbf{d};\mathbf{z}) \tag{6.3.6}$$

$$(37.76477^\circ N - d_1)^2 + (-122.43350^\circ W - d_2)^2 \leq (2400)^2 \tag{6.3.7}$$

$$d_5 \leq d_6 \tag{6.3.8}$$

$$d_8 \leq d_7 \tag{6.3.9}$$

$$d_{10} \leq d_9 \tag{6.3.10}$$

$$d_3 \in [500, 6000] \tag{6.3.11}$$

$$d_4 \in [0.25, 5.25] \tag{6.3.12}$$

$$d_5 \in [30000, 100000] \tag{6.3.13}$$

$$d_6 \in [50000, 200000] \tag{6.3.14}$$

$$d_7, d_8 \in [0, 20] \tag{6.3.15}$$

$$d_9, d_{10} \in [0, 15] \tag{6.3.16}$$

where the objective $f(\mathbf{d},\mathbf{x};\mathbf{z}) : \mathbb{R}^{10} \to \mathbb{R}^k$ maps the decision vector $\mathbf{d}$, simulation output variables $\mathbf{x}$ (e.g., trip characteristics, vehicle movements, etc.), and fixed configuration data $\mathbf{z}$ (e.g., fuel consumption rates, network characteristics, PT schedules, etc.) to a k-dimensional vector of the scores corresponding to each KPI. In equation 6.3.6, the function B represents the simulation given a particular decision vector $\mathbf{d}$ and configuration inputs $\mathbf{z}$, which results in the output data, $\mathbf{x}$. The various constraints on the input space that were discussed previously are encoded in constraints 6.3.7 through 6.3.16.

Table 6.3: Score component parameters: mean and standard deviation of BAU and random search KPI values and score component multipliers

		KPI	BAU		Random search		
i	Name	Units	$K_i(\emptyset, \mathbf{x}_0)$	σ_i^0	μ_i	σ_i	β_i
1	VD	minutes	5.92	0.16	0.94	0.076	1
2	VMT	1M miles	21.58	0.24	0.96	0.04	1
3	GHG	1k tons	24.06	0.077	0.99	0.01	1
4	$PM_{2.5}$	kg	54.37	0.58	0.97	0.02	1
5	TC	minutes	14.32	0.18	0.99	0.02	1
6	CB	$ per $1,000	0.89	0.015	0.99	0.09	1
7	NPR	$100,000	17.68	0.65	1.06	0.62	−1

Following the BISTRO scoring methodology, the scores for each of the KPI's are calculated as follows:

$$F_i\left(\mathbf{d}, \mathbf{x}; \mathbf{z}, \mathbf{x}_0, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\beta}\right) = \beta_i \frac{\frac{K_i(\mathbf{d},\mathbf{x};\mathbf{z})}{K_i(\emptyset,\mathbf{x}_0;\mathbf{z})} - \mu_i}{\sigma_i} \; \forall i = 1, \ldots, k \tag{6.3.17}$$

where $\mathbf{x}_0$ is the BAU simulation output (i.e., the calibrated simulation run with a null decision vector) and $\frac{K_i(\mathbf{d},\mathbf{x};\mathbf{z})}{K_i(\emptyset,\mathbf{x}_0;\mathbf{z})}$ is the *score component* for the ith KPI, and μ_i and σ_i are the mean and standard deviation of the ith score component from a random search of the input space. The random search included 95 samples that were randomly generated from uniform distributions of each input dimension and 70 samples that were generated with only the cordon parameters and the toll rate randomized while the incentive parameters were set to 0. The latter sample was generated in order to provide a basis of comparison for an exploratory analysis of the effects of incentives on the KPIs. Finally, β_i is a multiplier used to ensure that all score components decrease with respect to the direction of improvement. In particular, it is equal to negative one for the NPR score component and equal to one for all other score components. The values of $K_i(\emptyset, \mathbf{x}_0; \mathbf{z})$, $\mu_i, \sigma_i,$ and β_i are presented in Table 6.3, as well as the sd of each KPI value in the BAU results, σ_i^0. As seen in the table, the random search of the input space produces slight improvements in all of the KPIs (from 1 to 6%), on average. Furthermore, we note the large sd for the NPR KPI, which reflects the significant effect of incentives on this particular KPI, as discussed further in section 6.4.

The random search results confirmed strong colinearity between the the VMT, GHG emissions, and $PM_{2.5}$ emissions. Thus, only the VMT was included in the final objective function. However, the two emissions metrics were evaluated for all solution samples throughout the study and are included in the analysis of optimization results.

Objective function

This study employs a vector-valued objective function of the form $\mathbf{f} : y \in \mathcal{D} \rightarrow \mathbf{f}(y) = (\mathbf{f}_1(y), \ldots, \mathbf{f}_k(y)) \in \mathbb{R}^k$ that maps an input vector (i.e., the decision variables) from the search space $\mathcal{D} \subset \mathbb{R}^{10}$ defined by the problem constraints (equations 6.3.7 to 6.3.16) to the k-dimensional objective space. The optimization of this function is formulated with respect to a set of solutions that represent the optimal set of tradeoffs between each of the k objectives. This set is called the *Pareto set*, and is defined using the concept of *Pareto dominance* wherein a solution $\mathbf{d} \in \mathcal{D}$ *weakly Pareto dominates* another solution $\mathbf{d}' \in \mathcal{D}$ (written $\mathbf{d} \preceq \mathbf{d}'$) if and only if $\mathbf{f}_i(\mathbf{d}) \leq \mathbf{f}_i(\mathbf{d}')$ for all $i \in \{1, \ldots, k\}$. If $\mathbf{f}_i(\mathbf{d}) < \mathbf{f}_i(\mathbf{d}')$ for all $i \in \{1, \ldots, k\}$, then $\mathbf{d} \in \mathcal{D}$ *Pareto dominates* $\mathbf{d}' \in \mathcal{D}$ (written $\mathbf{d} \prec \mathbf{d}'$). The set of non-dominated solutions is the Pareto set and its image under $\mathbf{f}$ is called the *Pareto front*.

The *hypervolume* of a set of solutions $\mathcal{H}$ with respect to a reference point $\mathbf{r} \in \mathbb{R}^k$ is the Lebesgue measure of the Pareto front of $\mathcal{H}$. It is defined as $HV_{\mathbf{r}}(\mathcal{H}) = \lambda_k(\{\mathbf{z} \in \mathbb{R}^k : \exists \mathbf{d} \in \mathcal{H}, \mathbf{f}(\mathbf{d}) \prec \mathbf{z} \prec \mathbf{r}\})$ where λ_k is the Lebesgue measure. The hypervolume is a strictly monotone indicator of the quality of a set of solutions to a multi-objective optimization problem that essentially measures the size of the subset of the search space that is Pareto dominated by a given set of solutions. Maximizing the hypervolume of a set of solutions for a multi-objective optimization problem produces a set of solutions that maximize the optimality of the trade-offs between the conflicting objectives of the problem. In the context of this study, this translates to finding multiple possible congestion pricing and incentive schemes to recommend for consideration by policy-makers and other stakeholders. The particular strategy chosen depends on the priorities of decision-makers. The following hypervolume maximization approach entails that the objective of the optimization problem (equation 6.3.5) is replaced by the following set-valued objective:

$$\max_{D \subseteq \mathcal{D}, |D| \leq p} HV_{\mathbf{r}}(D) \tag{6.3.18}$$

where the set D is the *p-distribution*, a set of p or fewer solutions to problem 6.3.5 - 6.3.16.

Optimization algorithm

Several approaches exist to optimize the hypervolume of a multi-objective optimization problem. This study applies an evolutionary algorithm called the Comma-Selection Multiobjective Covariance Matrix Adaptation Evolutionary Strategy (COMO-CMA-ES)[178]. The COMO-CMA-ES is a multi-objective evolutionary optimization algorithm that uses several instantiations of the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) to iteratively optimize the p-distribution of a given problem. As with other evolutionary algorithms, CMA-ES samples *generations* of solutions for an optimization problem from a multivariate normal search distribution with parameters (i.e., the mean and covariance of the distribution) that are iteratively updated based on the best-performing candidate solutions from one or more prior generations [179]. The basic equation for sampling the $\lambda \geq 2$

solutions of generation $n+1$ is as follows:

$$\mathbf{d}_i^{(n+1)} \sim \mathcal{N}\big(\mathbf{m}^{(n)}, (\sigma^{(n)})^2 \mathbf{C}^{(n)}\big) \; \forall i = 1, \ldots, \lambda$$

where $\mathbf{m}^{(n)}$ is the weighted mean of $\mu \leq \lambda$ selected solutions from the previous generation of samples $(\mathbf{d}_1^{(n)}, \ldots, \mathbf{d}_\lambda^{(n)})$, $\mathbf{C}^{(n)}$ is the covariance matrix at generation n, and $\sigma^{(n)}$ is the 'overall' sd (also called the step size) at generation n. All three of these parameters are updated after each iteration. For the specification of the corresponding update rules and the CMA-ES algorithm, the interested reader is referred to [179].

The COMO-CMA-ES algorithm used in this study involves p instantiations of CMA-ES algorithms each optimizing the *uncrowded hypervolume indicator* (UHVI) of the solutions they sample [178]. The uncrowded hypervolume indicator, $\text{UHVI}(\mathbf{d}, \mathcal{H})$ is a measure of the distance of the sample $\mathbf{d}$ to the empirical Pareto Front of the history of samples $\mathcal{H}$. In particular, the UHVI ensures that the search direction of the optimization algorithm points from non-dominated samples toward the non-dominated search space and from dominated points toward the uncrowded space between non-dominated samples, with the following definition:

$$\text{UHVI}_{\mathbf{r}}(\mathbf{d}, \mathcal{H}) = \begin{cases} \text{HV}_{\mathbf{r}}(\mathcal{H} \cup \{\mathbf{d}\}) - \text{HV}_{\mathbf{r}}(\mathcal{H}) & \text{if } \text{EPF}_{\mathcal{H},\mathbf{r}} \not\prec \mathbf{f}(\mathbf{d}) \\ -d_{\mathbf{r}}(\mathbf{d}, \mathcal{H}) & \text{if } \text{EPF}_{\mathcal{H},\mathbf{r}} \prec \mathbf{f}(\mathbf{d}) \end{cases}$$

where $\text{EPF}_{\mathcal{H},\mathbf{r}} = \{\mathbf{z} \prec \mathbf{r} : \forall \mathbf{d} \in \mathcal{H}, \mathbf{f}(\mathbf{d}) \not\prec \mathbf{z}\}$ is the *empirical* Pareto front, the boundary of the objective space that dominates the reference vector $\mathbf{r}$ and is not dominated by the samples in the search history $\mathcal{H}$ [178]. For further specification of the COMO-CMA-ES algorithm, the interested reader is referred to [178].

The COMO-CMA-ES was implemented using the pycomocma python package. The hyper-parameter settings for the algorithm were determined using the random search results and the guidelines provided by the package developer. In particular, the algorithm was instantiated with 6 CMA-ES solvers, an initial generation size of 11 samples per solver, and five of the seven dimensions of the objective. KPIs 3 and 4, corresponding to GHG and $PM_{2.5}$ emissions were excluded from the algorithm due to their strong colinearity with VMT. The number of solvers was chosen with the expectation of 6 interesting two-way trade-offs between the KPIs: VMT (and VD, GHG, $PM_{2.5}$) - TC, VMT - CB, VMT - NPR, TC - CB, TC - NPR, and CB - NPR. The reference point was set to double the maximum of each score component produced by the random search results ($\mathbf{r} = (3, 3, \emptyset, \emptyset, 3, 3, 4)$). In addition, the input space was normalized to a range of $[0, 10]$ and the initial sigma of the problem was set to $\sqrt{10}$. All 165 samples from the random search were sorted in descending order of Euclidian distance from the reference point and the top 66 samples were used to instantiate the pycomocma solver (hereby referred to as iteration 0). Ten iterations of the algorithm were run, producing a total of 714 samples in addition to the initial 165 random search samples. 54 of these samples comprised the Pareto Front of the full sample history.

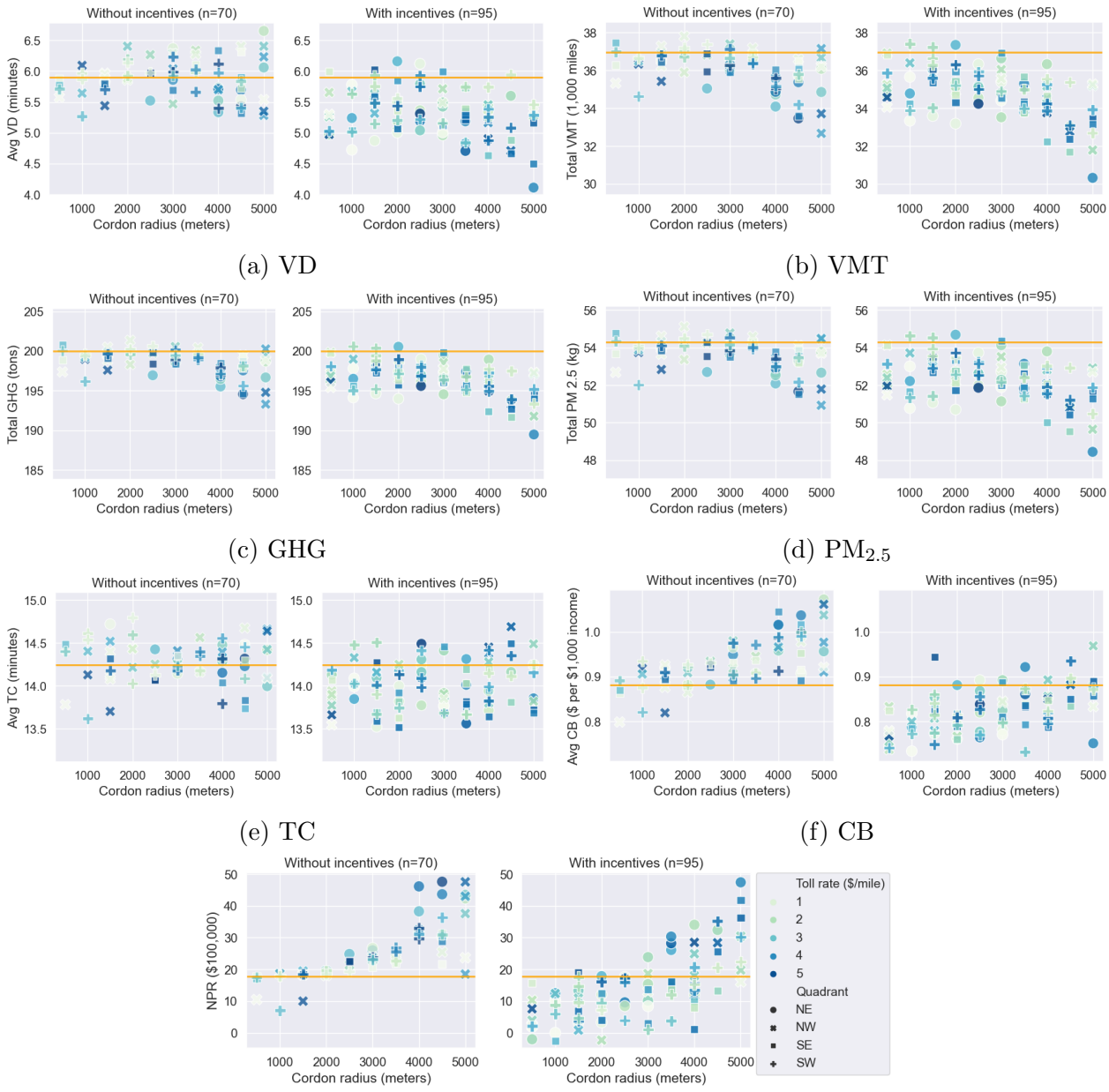
6.4 Results

This section begins with a brief analysis of the random search results followed by the presentation of the optimization results, including the performance of the evolution of the sample history across iterations of the optimization algorithm, and analyses of the key trends observed in the relationships between the input parameters and the KPIs. Finally, we present an in-depth analysis of a selection of Pareto-optimal congestion pricing and incentive strategies.

Random search

The random search results provide insight regarding a) the general relationships between the congestion pricing and incentive scheme design parameters and the KPIs and b) the relative impact of incentives on each of the KPIs. The distributions of the extrapolated KPI values with respect to cordon size (i.e., the radius in meters) evaluated from the random search samples both without (left) and with (right) incentives are displayed in Figure 6.5. The BAU KPI values are shown with an orange line across each plot and the color scale represents the toll rate (in dollars per mile) of each sample. Finally, the points are displayed with different shapes representing the quadrant of the search space in which the cordon is placed. Across all of the congestion-related KPIs, including VMT, VD, GHG, and $PM_{2.5}$, improvements (i.e., decreases) in the KPI values generally correlate with the size and toll rate of the cordons, particularly for cordons with radii greater than 2 km (see figures 6.5b-6.5d). Moreover, the inclusion of incentives results in significant improvements across all four of these KPIs. The NPR also clearly increases with respect to the size of the cordons and the toll rate (see figure 6.5g), with a majority of samples without incentives producing greater NPR than the BAU across all cordon sizes, locations, and toll rates. The inclusion of incentives generally worsens the NPR, particular for samples with cordon radii smaller than 4.5 km. The relationships of the incentive parameters with each KPI are investigated further using the results across all samples from both the random search and optimization processes.

The average CB generally worsens (i.e., increases) with respect to the size and toll rate of the cordons, with most of the random search samples without incentives resulting in worse average CB than in the BAU. This suggests that the increases in the travel cost across the population outweigh the travel time reductions induced by the congestion pricing strategies without incentives, particularly for lower income travelers. Moreover, congestion pricing without incentives is likely to increase the travel times of lower income individuals who choose not to pay the tolls and shift to slower travel mode alternatives (e.g., PT). Incentives serve to significantly reduce the average CB, with most of the samples including incentives resulting in better average CB than in the BAU. Lastly, the average TC does not vary significantly with respect to the cordon parameters either with or without incentives. However, the samples that include incentives generally produce lower average TC than those without incentives. The cause for this trend is explored further in the following subsections



(a) VD

(b) VMT

(c) GHG

(d) $PM_{2.5}$

(e) TC

(f) CB

(g) NPR

Figure 6.5: Distributions of KPIs in the random search sample with (right) and without (left) incentives by cordon radius (x-axis), toll rate (color), and location (shape). The mean BAU KPI value is shown as an orange line across all graphs.

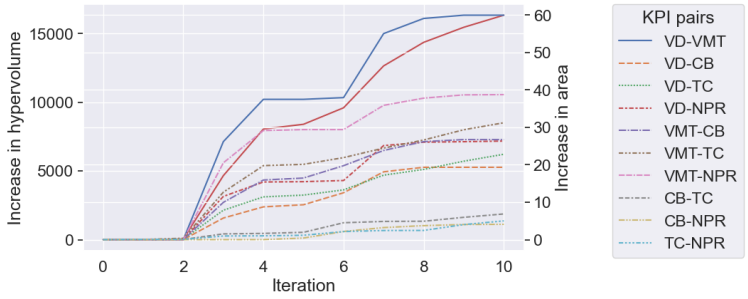


Figure 6.6: Evolution of the overall (red) and pairwise hypervolumes of the objective space across iterations of the COMO-CMA-ES optimization algorithm.

using the full history of samples generated from both the random search and optimization algorithm.

Optimization

The COMO-CMA-ES algorithm was run for ten iterations, with 11 samples per optimizer in the first iteration and 12 samples per optimizer in the subsequent iterations. A total of 714 samples were evaluated during the optimization process, amounting to a total of 879 samples across both the optimization and random search processes. Over the course of the COMO-CMA-ES algorithm, the Pareto front of the sample history grew from 15 samples from the random search to 54 samples after the tenth iteration. The resulting evolution of the hypervolume of the dominated objective space across the ten iterations is displayed in Figure 6.6, with the random search shown as iteration 0. The red solid line shows the increase in the overall hypervolume of the dominated objective space of the sample history at each iteration while the remaining lines show the increases in the area of dominated objective spaces for each pair of KPIs. In the first two iterations, the hypervolume of the dominated objective space was expanded by a total of four samples, with small improvements corresponding to the average CB and TC. In the subsequent iterations, improvements were made across all dimensions of the objective space, with the greatest increases achieved by samples with successively larger improvements in the average VD and total VMT KPIs.

Figure 6.7 illustrates the evolution of the Pareto front across the sample history by displaying the pairwise distributions of the KPI values for samples on the Pareto front at each iteration of the optimization algorithm. The points shown in the lightest color (iteration 0) are the samples on the Pareto front of the random search while those shown in the darkest color are the samples that were added to the Pareto front following the tenth iteration. The samples highlighted with a green circle comprise the Pareto front of the full sample history. The last column of plots in Figure 6.7 shows how the distribution of each of the KPI values on the Pareto front evolved across the random search and the ten iterations of the optimization process. Across all KPIs except for TC, there is a trend of improving KPI values among the

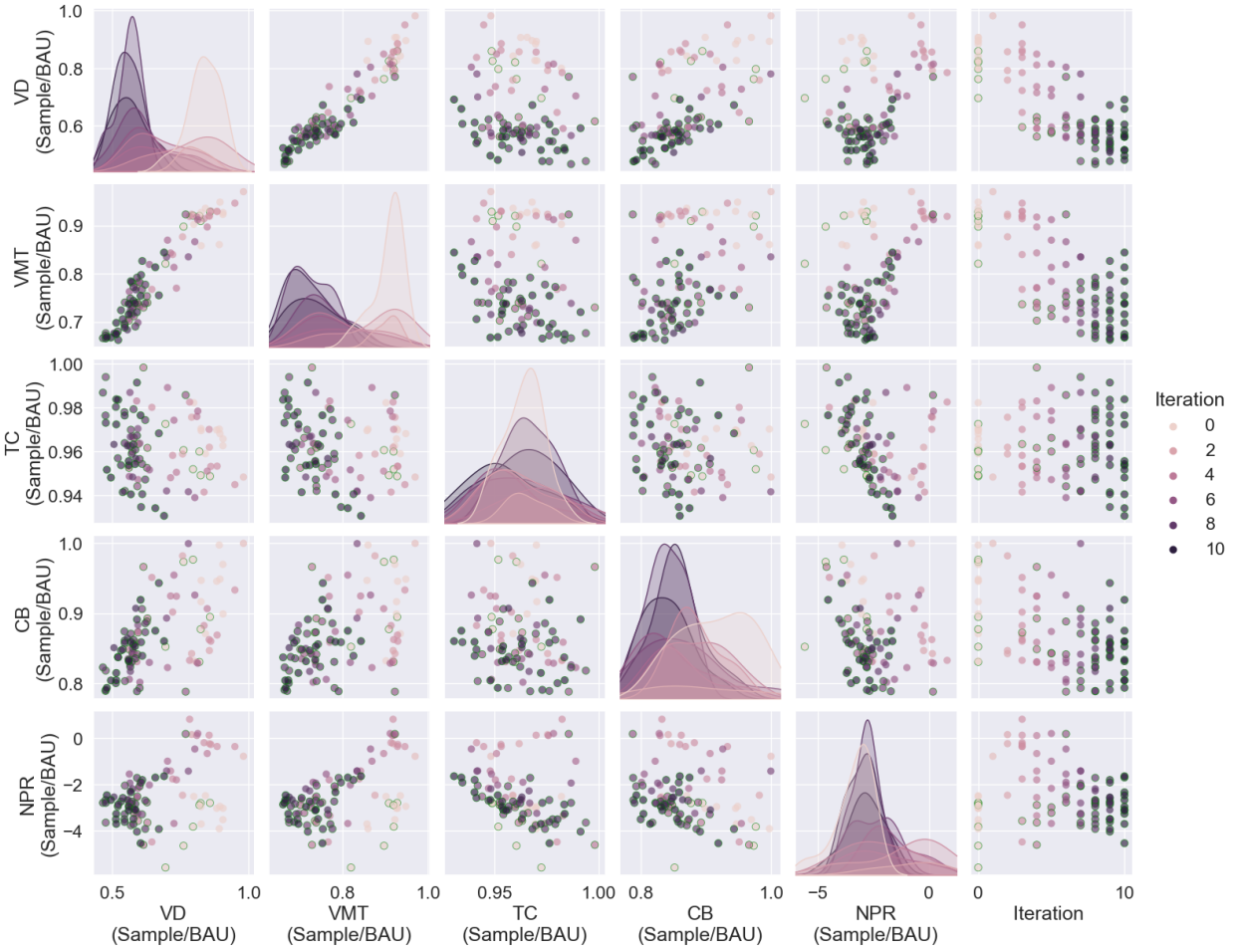


Figure 6.7: Evolution of the Pareto front for each pair of KPIs by iteration of the optimization algorithm. The Pareto front of the entire sample history is highlighted in green. Note, the KPI values are shown as the ratio of the sample KPI value to the mean BAU KPI value; iteration 0 refers to the random search samples.

Pareto front samples across iterations, with a slight increase in variance in the last two to three iterations. There is also a slight downward (i.e., improving) trend in the TC values across iterations, though with greater variance than in the other KPIs.

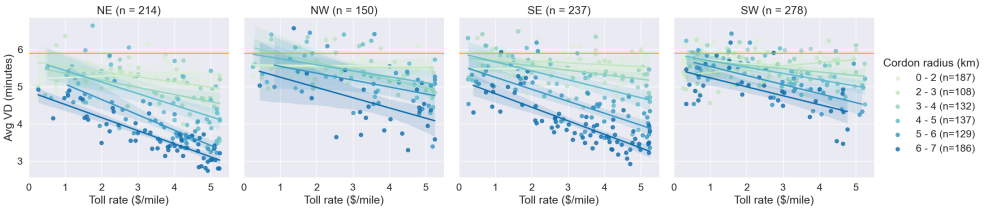
The visualization of the pairwise distributions in figure 6.7 helps to understand the relationships between KPIs and the significance of the tradeoffs between two KPIs at a time. Although higher-dimensional relationships are difficult to visualize, one can glean insight regarding these relationships by observing the trends in the corresponding pairwise Pareto fronts. The Pareto front with respect to VD and VMT is composed of just three samples due to a strong correlation between these two KPIs. This suggests that the objectives of VD and VMT reduction are complementary; GHG and $PM_{2.5}$ emissions reduction are also

complementary to these objectives. This strong complementary relationship is likely to have biased the search history of the optimization, particularly since neither VD nor VMT were constrained, thus rewarding the optimization algorithm for exploring strategies with greater impacts on congestion. In contrast, the Pareto fronts of NPR with the other KPIs appear almost linear and downward sloping, suggesting direct substitution between the objective of raising NPR and the objectives of reducing congestion, CB, and TC. Finally, the Pareto fronts of CB and TC with each other and with each of VD and VMT are approximately convex, suggesting more nuanced trade-offs between these KPIs.

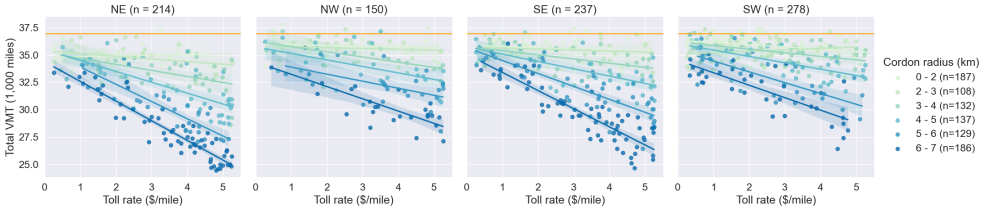
Effects of congestion pricing parameters on policy outcomes

Next, we analyze the relationships between the optimization input parameters and the KPIs. A correlation matrix of all input parameters and KPIs as well as other output metrics is presented in Table B.3. Figure 6.8 displays the distributions of the KPI values with respect to the toll rate by quadrant of the cordon search space. In addition, a linear regression was fit to each of six subsets of the data binned by cordon radii, as shown by the solid lines and shaded 95% confidence intervals on each of the plots. Many of the trends that were observed from the random search samples with incentives are reflected in figure 6.8, with some additional insights gleaned from the full history of 879 samples. In particular, the cordon size and toll rate are negatively correlated with the reduction of congestion-related KPIs, including average VD, total VMT, and total GHG and $PM_{2.5}$ emissions. The Pearson correlation coefficients (hereby referred to as the r value) of the cordon size and the VD, VMT, GHG, and $PM_{2.5}$ KPIs are -0.60, -0.73, -0.73, and -0.71, respectively, while those of the toll rate with each KPI are -0.47, -0.51, -0.51, and -0.49, respectively. The total GHG and $PM_{2.5}$ emissions are perfectly correlated with the total VMT and exhibit the same trends as are seen in figure 6.8b. In addition, placement of the cordon in the Northeast quadrant of the cordon search space resulted in the greatest reductions in the VMT, VD, GHG, and $PM_{2.5}$ KPIs, across cordon sizes and toll rates. Cordons with radii of 5 km or greater placed in the Southeast quadrant also performed slightly better in comparison to the two Western quadrants, across all four congestion-related KPIs.

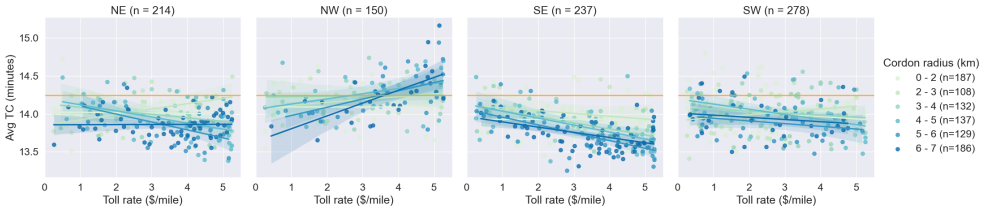
The relationships between the cordon parameters and the TC, CB, and NPR KPIs are less strong, likely due to confounding effects produced by the incentives. As observed in figures 6.5e - 6.5g and confirmed by the correlation coefficients of the incentive parameters (see Table B.3), incentives reduce the TC, CB, and NPR. Average TC decreases slightly with increasing toll rates across all quadrants except in the Northwest, where there is a slight upward trend in TC as the toll rate increases, particularly for cordon radii of 4km or greater. The average CB also decreases slightly with increasing toll rates in the Eastern quadrants but appears to increase slightly as the toll rate increases for the largest interval of cordon radii in the Western quadrants. This may reflect a disparate distribution of impacts on travel times and cost from cordons placed in the Western half of the search space. Further analysis is needed to discern the effects of the cordon placement on the distribution of travel time and cost changes across the population and the resulting impact on cost burden. Lastly, figure



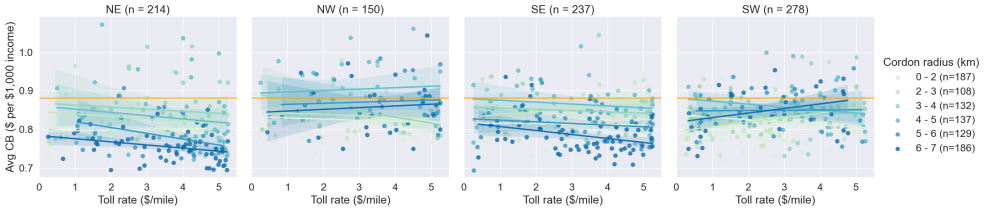
(a) VD



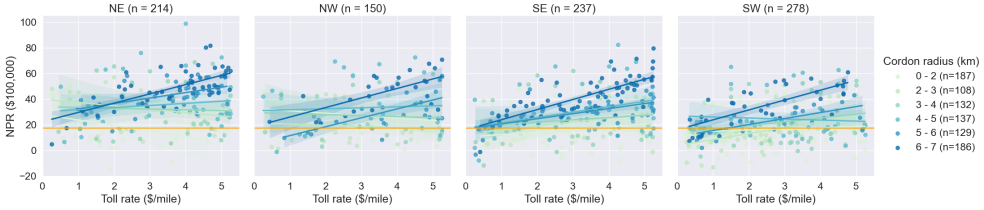
(b) VMT



(c) TC



(d) CB



(e) NPR

Figure 6.8: Distributions of KPIs across all samples by cordon radius (x-axis), toll rate (color), and location (from left to right: NE, NW, SE, and SW). The mean BAU KPI value is shown as an orange line across all graphs.

6.8e reflects an increasing trend in NPR with respect to increasing toll rates, particularly for cordons with 5km or greater radii[9].

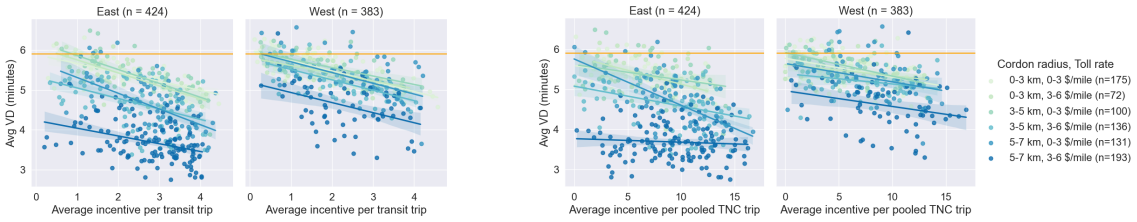
Next, we investigate the effects of the incentives on the KPIs. To this end, we estimated the average incentives received per PT and PTNC trip for each sample. This metric summarizes the four input parameters that determine the incentive structure for each mode: the two income thresholds and corresponding incentive levels[10]. Figure 6.9 displays the distributions of the KPI values with respect to the average incentives received per PT and PTNC trips. The best fit and 95% confidence interval of a linear regression for each of six bins of cordon radii and roll rates overlay each of the plots.

The average VD is negatively correlated with all six of the incentive parameters, as reflected by the negative trends displayed across figures 6.9a and 6.9b, with a stronger negative correlation to the PT incentives than the PTNC incentives. The average PT incentives are similarly correlated with VMT as they are with VD (r values of -0.57 and -0.59, respectively) while the correlation between average PTNC incentives and VMT is weaker than with VD (r values of -0.45 and -0.51, respectively). This may be due to the positive correlation between PTNC incentives and TNC VMT (r value of 0.65) which, despite the almost equally strong negative correlation with the TNC deadheading ratio (r value of -0.59), may counterbalance the reduction in VMT from reduced driving alone. Thus, while increased PTNC adoption increases the overall efficiency of TNC service and reduces VD, the increased TNC VMT necessary to fulfill the increase in demand may dampen the ultimate effect on overall VMT in the region. An exception to these trends is seen in figures 6.9b and 6.9d where the average VD and total VMT for cordons with radii of five km or greater and toll rates of $3/mile or greater placed in either the NE or SE quadrants of the search space do not appear to be affected by the average incentive for PTNC. This is likely do to the relatively high impact of this level of toll rates with such a large cordon on TNC trips to the most densely visited areas of the city.

In figure 6.9e, we observe that the average TC decreases as the average incentive per PT trip increases. This trend contradicts the logic guiding the design of the TC KPI, which was intended to reflect the potential degradation of the quality of PT service from the overcrowding of vehicles. The correlation matrix in Appendix B confirms that the PT mode split increases with respect to increases in the PT incentive parameters as well as with respect to the average incentive per PT trip. Due to the geographic diversity of the SF Bay Area, the rate of increase of PT passenger-time on un-crowded routes may surpasses that of passenger-time on crowded routes. Thus, the overall increase in PT use merely dilutes the effect of crowded passenger-time in the average TC metric and results in a decreasing relationship between the average TC and the quantity of PT users.

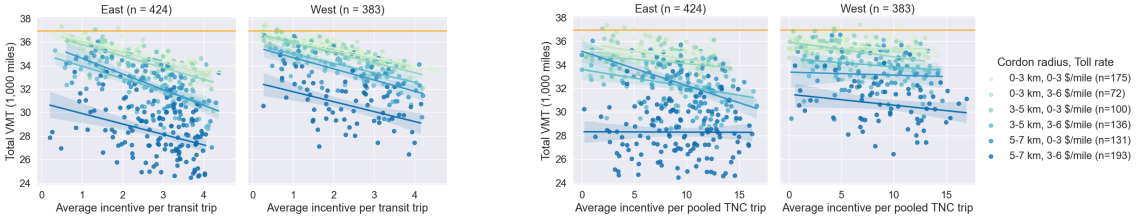
[9]The r values of the cordon radius and toll rates and the NPR are -0.48 and -0.34, respectively.

[10]The intuition behind the use of the average incentive received per trip is as follows: as either of the thresholds are increased, the number of users of the corresponding mode generally increases thus increasing the portion of users receiving incentives and the resulting average incentive received per trip by that mode. Moreover, as the incentive levels increase, so do the number of users of the corresponding mode as well as the amount of incentive received by those users thus increasing the average incentive received per trip.



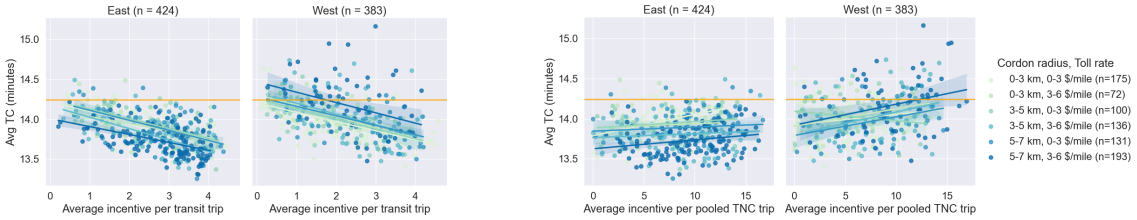
(a) VD - PT

(b) VD - PTNC



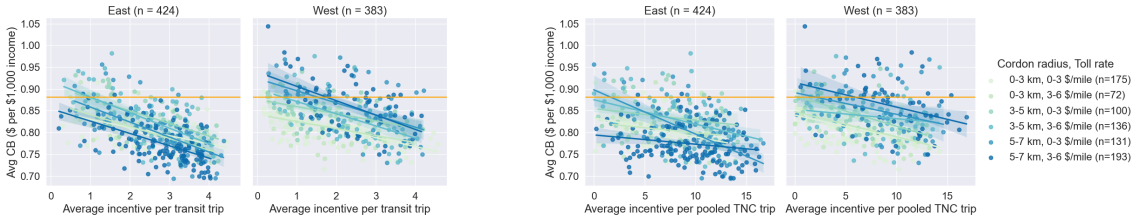
(c) VMT - PT

(d) VMT - PTNC



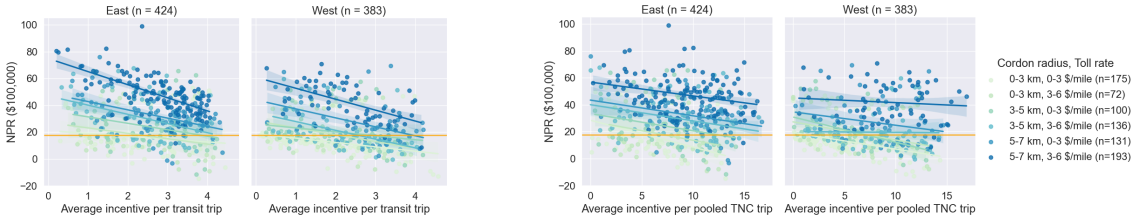
(e) TC - PT

(f) TC - PTNC



(g) CB - PT

(h) CB - PTNC



(i) NPR - PT

(j) NPR - PTNC

Figure 6.9: Distributions of KPIs across all samples by the average PT and PTNC incentives received per trip (x-axis), cordon radius and toll rate (color), and location (from left to right: East, West). The mean BAU KPI value is shown as an orange line across all graphs.

Figures 6.9g and 6.9h show that the average CB decreases with respect to increasing incentives due to the reduced trip costs for eligible travelers. However, as observed in the relationships between VD and VMT with the average PTNC incentive, the CB for cordons in the East half of the search space with radii greater than 5km and toll rates greater than \$3/mile is relatively insensitive to the average incentive for pooled trips.

Lastly, we observe that, across both incentive types and all cordon parameters, the NPR decreases with respect to both incentives. Although total PT revenue is weakly correlated with the toll rate (r value of 0.20) and location of the cordon (r values of 0.17 and -0.22 between PT revenue and placement in the NE and SW quadrants, respectively), it is not correlated with the PT incentives (r value of 0.05). This suggests that the subsidization of PT fares via the incentives merely increases PT ridership without generating additional PT revenue. Thus, the balance of the congestion pricing scheme's impact on NPR depends primarily on the relative amount of toll revenues and incentives distributed.

Analysis of Pareto-optimal congestion pricing strategies

Next, we present an in-depth analysis of six samples on the Pareto front of the sample history with respect to the VD, VMT, CB, and NPR KPIs as well as an alternative TC metric measuring the percent of passenger hours on PT that were on a crowded vehicle. In addition, the minimum and maximum as well as the BAU values of each metric are displayed. The samples were selected based on their locations on the Pareto fronts of each of the pairs and triples of these five KPIs in order to examine the similarities and differences in the strategies that equally balanced different sets of KPIs. This was done by first normalizing the KPIs so that the ratios of the KPI to BAU scores all ranged from 0 to 1. Next, the Pareto fronts of each pair and triple of KPIs were identified and the sample closest to the unit vector (e.g., x = y for the pairs and x = y = z for the triples) were chosen for further investigation. In this way, nine samples were selected from the ten pairwise Pareto fronts, one of which corresponded to two different pairs of KPIs (sample number 626 corresponded to both the VHD-VMT and VMT-CB Pareto Fronts). In addition, 12 samples corresponded to the three-dimensional Pareto fronts. The remainder of this subsection focuses on five of these samples. Table 6.4 presents the policy input parameters, KPIs, aggregate mode splits, and other key metrics for these samples, which are referred to by the number corresponding to the order in which they were sampled. The corresponding Pareto fronts for each sample are noted in the table as well.

First, we compare samples 626 and 702. Sample 626 is the sample on the Pareto fronts of VD and VMT and of VMT and CB that balances each of these pairs of KPIs as well their triple. Sample 702 is on the Pareto fronts of VD, CB and NPR as well as of VD, VMT, and TC, balancing each of these triples. The cordons of the two samples are relatively similar, with that of 626 placed about 1 mile northwest of that of sample 702 with a 500m longer radius and 20 cent per mile greater toll rate. Moreover, the incentive thresholds for the two strategies are almost identical, including about 25% and 50% of commuters at the first and second threshold, respectively. However, strategy 626 provides significantly greater PT and

				Samples				
	Metric	Units	BAU	#251	#389	#470	#626	#702
Cordon parameters	Quadrant		-	SW	SE	NE	NE	NE
	Latitude	degrees	-	37.7613	37.7510	37.7684	37.7744	37.7745
	Longitude	degrees	-	-122.4393	-122.4160	-122.4123	-122.4251	-122.4123
	Radius	km	-	5.59	5.96	4.85	6.87	6.35
	Toll rate	$/mile	-	5.1	0.27	4.3	4.86	4.66
Income thresholds	1st level, 2nd level	$1,000	-	65.6, 138.8	71.3, 147.5	89.0, 148.7	64.1, 196.1	61.4, 197.8
% Eligible	1st level, 2nd level	% of pop.	-	28, 33	32,33	41,24	27,53	25,55
PT incentives	1st level, 2nd level	$	-	8.7, 8.7	5.47, 5.47	7.54, 7.54	14.64, 8.57	9.84, 9.84
PTNC incentives	1st level, 2nd level	$	-	19.59, 3.57	19.86, 17.14	12.65, 12.65	17.49, 8.09	5.15, 5.15
KPIs	VD	minutes	5.92	4.26	4.55	3.54	2.82	2.84
	VMT	1M miles	21.58	16.99	19.87	17.23	14.33	14.39
	GHG	1000 tons CO_2	24.06	22.57	23.49	22.63	21.72	21.73
	PM2.5	kg	54.37	47.49	51.44	47.7	43.37	43.38
	TC	minutes	14.32	13.87	14.04	13.46	13.86	13.56
	TC ratio	% of pax-hours	9.4	12.9	7.9	13.2	18.5	16.7
	CB	$ / $1,000	0.89	0.82	0.69	0.72	0.70	0.70
	NPR	$100,000	18	26	-3	32	50	44
	Pareto fronts		-	VMT-TC	CB-TC	VD-TC	VD-VMT	VD-CB-NPR
			-		CB-TC-NPR		VMT-CB	VD-VMT-TC
			-				VD-VMT-CB	
Mode splits	DA	% of all trips	50	27	40	26	17	19
	PT		25	43	28	46	55	55
	RA TNC		1.9	0.5	0.9	0.5	0.4	0.4
	PTNC		1.5	3.2	7.9	3.6	3.1	1.4
	Active		21	26	24	24	25	24
Revenue	Total toll revenue	$100,000	0.0	23.6	6.9	28.3	49.6	39.6
breakdown	Total PT revenue		17.7	31.2	20.4	33.6	40.4	40.4
	Total PT incentives		0.0	21.4	13.6	24.0	33.9	35.0
	Total PTNC incentives		0.0	7.6	17.0	6.2	6.6	1.2
TNC	TNC VMT/total VMT	%	7	10	12	10	12	10
operations	TNC deadheading ratio		56	51	40	55	58	64
	PTNC match rate		52	60	66	63	63	47
	% TNC rides $\geq$4 pax		7	15.8	27.6	17.8	14.4	4.4
Other statistics	Mean VMT/auto trip	miles/trip	24	28	23	28	30	30
	Mean VMT/DA trip		26	32	29	32	36	33
	Mean VMT/TNC trip		25	24	16	23	25	32

Table 6.4: Summary of five Pareto-optimal congestion pricing strategies

PTNC incentives at the first level that are $4.80 and $12.34 greater than those of strategy 702, respectively. In addition, the second level PT incentive in strategy 626 is $1.27 lower than in strategy 702 and the second level PTNC incentive is $2.94 higher. All of the resulting KPI values are almost identical with the average VD and total VMT being slightly lower in strategy 626, the TC ratio being greater by about 1.8% and the NPR being greater by $600k. The key differences in the mode splits reflect both the higher toll rate and higher PTNC incentives in strategy 626, with 2.0% fewer driving trips and 1.7% greater PTNC trips in strategy 626 compared to 702.Despite the lower DA mode split, strategy 626 produces $1M more in toll revenue (likely due to the larger cordon and higher toll rate) and by offering lower second level PT incentives, it spends $110k less on PT incentives. On the other hand, the greater PTNC incentives result in spending $540k more on PTNC incentives. Although the insignificant differences across the congestion-related and CB KPIs begs the question of whether that increased spending is effective, we see that the increased number of PTNC trips increases the PTNC match rate by a third and reduces the TNC deadheading ratio

and average VMT per TNC trip by about 9 and 22%, respectively. Moreover, about 14% of PTNC trips included four or more passengers.

Next, we examine samples 251 and 470, the samples on the Pareto fronts of VMT and TC and of VD an TC, respectively. In strategy 251, the cordon is placed in the Castro neighborhood with almost 6 km in radius and a toll rate of \$5.1 per mile. The incentive scheme provides \$8.7 per PT trip for 61% of commuters eligible at either threshold, and \$19.59 and \$3.57 per PTNC trip for the 28% and 33% of commuters eligible at each threshold. In comparison, strategy 470 has a cordon almost 1 km smaller placed in the northern portion of the Mission district with a toll rate of \$4.3 per mile. The incentive structure includes 13% more commuters at the lower level but provides a similar PT incentive that is equal across the two levels and PTNC incentives of \$12.56 for both levels. Interestingly, the mode splits of the two samples are almost identical, with strategy 251 producing slightly greater DA and active mode splits and strategy 470 producing slightly greater PT, and TNC mode splits. While this results in similar reductions in VMT, GHG, and $PM_{2.5}$ compared to the BAU (about 20%, 6%, and 12%, respectively), the average VD in strategy 251 is about 20% higher than in strategy 470. Moreover, the NPR in strategy 470 is about \$6M greater than in strategy 251, due to greater toll and PT revenues amounting to \$7M greater revenue despite greater expense of on PT incentives (about \$2.5M greater than in strategy 251). Lastly, the CB in both of these strategies are worse than in the other strategies examined, with that of strategy 251 being about 15% worse than that of strategy 470.

In contrast to the four strategies examined thus far, sample 389 - the strategy balancing the average CB and TC ratio as well as the CB, TC ratio, and NPR - produces only slight improvements to the BAU across all of the KPIs except for the NPR, which is worsened by about \$2.1M from the BAU. In this strategy, a cordon of almost 6km in radius is placed on the southern edge of the Mission District with a \$0.27/mile toll rate. The cordon covers almost all of the City of San Francisco with the exception of a small portion of the northern edge of the city and the area to the West of the CA 1 highway. The income thresholds for the incentives in this strategy include 32 and 33% of commuters; the PT incentives are both \$5.47 while the PTNC incentives are at the high end of the search space, at \$19.86 and \$17.14, successively. This strategy results in a reduction of the DA and ride alone TNC mode splits of 10% and 1%, respectively, while increasing the PT, PTNC, and walk mode splits by 3%, 6.4%, and 3%, respectively. Strategy 389 raises total PT revenue by about \$2.74M. However, the total incentives distributed for both PT and PTNC and a meager \$600k toll revenue result in the net loss of \$300k in public revenue. As a result, the VMT, GHG, and $PM_{2.5}$ are each reduced by less than 10% from the BAU and average VD is reduced by almost a quarter, from about 5.9 to 4.6 minutes per passenger vehicle trip. In addition, the efficiency of TNC operations increases dramatically, with reduction of the TNC deadheading ratio to 40% of TNC VMT, an increase in the PTNC match rate to 66%, and about 28% of PTNC rides resulting in pools of four passengers or more.

To better understand the policy implications of the high PTNC incentives in strategy 389, we analyzed the mode splits and trip characteristics at the regional level. Figure 6.10 displays the average travel times per trip taken during the AM peak period (i.e., between 7

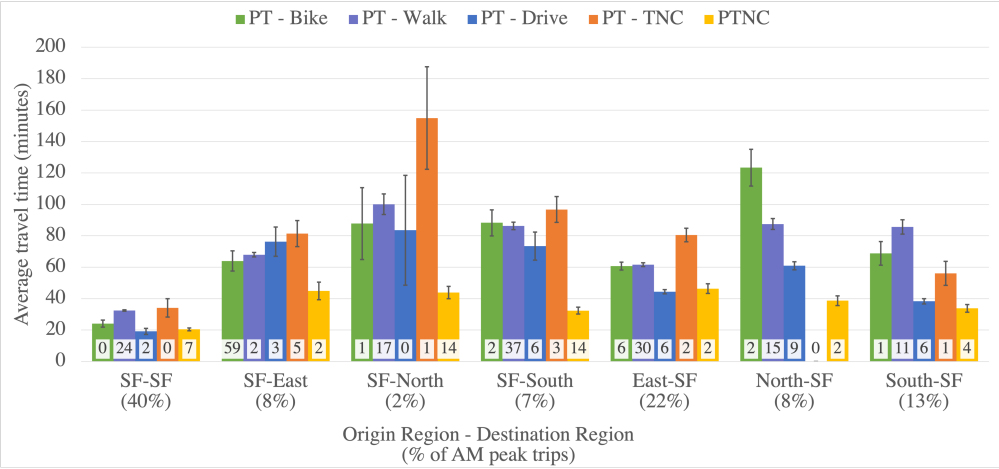


Figure 6.10: Comparison of average AM peak travel times per trip by trip origin and destination regions and travel mode. The 95th% confidence intervals (brackets) and the mode split of each mode in each origin-destination pair (white labels) are overlaid on each bar.

and 10 am) by travel mode segmented by the regions of the trip origins and destinations. It is important to note that the following analysis is based on the average travel times of the trips that were taken and is not representative of the overall average travel time available in any of the region ODs. The data are skewed by the revealed preferences of the agents in the model given the congestion pricing and tolling scheme applied in strategy 389. For example, we note that while the average travel time of PT combined with TNC should be similar to that of PT combined with driving, the average travel time of the trips that used this mode is likely skewed due to the relatively high cost of this option resulting in a greater preference for other access/egress modes to PT whenever they are viable.

On average, PTNC users making the 'reverse commute' from San Francisco to other parts of the Bay Area during the AM peak experienced significant travel time savings over PT service due to the relatively low levels of congestion on routes flowing out of the city during this time. The average occupancy of these PTNC trips were about 2.4, 1.7, and 1.4 passengers per trip going from San Francisco to the East, North, and South Bay Area, respectively. For trips within San Francisco, the average PTNC user had a similar travel time as commuters that drove to/from PT, with an average vehicle occupancy of 2.9 passengers per PTNC trip. PTNC users traveling into the city from the East and South Bay Area, had a similar travel time as driving to/from PT, while for those traveling from the North Bay Area into the city, PTNC was significantly faster, on average. This is likely due to the availability of heavy and light rail services connecting both the East and South Bay Area to the City of San Francisco which, when combined with a driving access trip, results in a comparable travel time to PTNC service. In contrast, PT service from the North Bay Area is more sparse, consisting of relatively few fixed route bus and ferry services. PTNC trips from the East and South Bay Area into San Francisco had average occupancies of

2.5 and 2.3 passengers per trip while those from the North Bay Area into the city had an average occupancy of 1.6 passengers per trip. Across all OD pairs of regions, the average PTNC occupancies likely reflect the density of trip requests, with SF-SF trips having the highest density and thus the greatest average occupancy, SF-East and East-SF having the next highest density, and so on.

6.5 Discussion and Conclusions

This case study demonstrated the application of simulation-based multi-objective optimization using BISTRO for the study of congestion pricing and incentive schemes in the San Francisco Bay Area. The methodological approach of this study offers insights to further the development of algorithmic policy decision support systems for the design and analysis of multi-faceted TDM strategies. This section first discusses the limitations of the study and potential improvements for the design and calibration of the SFSim model, specification of the congestion pricing and incentive optimization problem, and the application of multi-objective simulation-based optimization methods for transportation policy optimization. Next, this section reviews the key findings of the study and discusses relevant policy implications and future research directions for the study of congestion pricing optimization.

Study limitations

Several limitations of the SFSim model and the problem specification should be taken into consideration when interpreting the results of this study. The SFSim model was calibrated using data from 2018 to ensure that the commute mode splits, region-wide road network travel speeds, and TNC fleet operations in the BAU were representative of the mobility patterns in the San Francisco Bay Area during that time. Nonetheless, this model merely reflects the commute travel patterns of the region at a snapshot in time - much has and will continue to evolve in the travel behavior of the region. In addition to ever-changing land-use patterns throughout the San Francisco Bay Area and the introduction of new shared micromobility services such as dockless shared electric bicycles and scooters that launched in San Francisco in 2018 [180], the COVID-19 pandemic has had a significant impact on the quantity and distribution of travel demand throughout the region. From 2019 to 2021, the estimated number of daily trips in the City of San Francisco decreased from 4.5 million to 4.4 million and the PT mode split fell by 11% [181]. The share of inter- and intra-city trips that were made for commute purposes also fell by nine and 11 percentage points, respectively [181]. While the model used in this study does not reflect these major developments, the comparison of the impacts of the various congestion pricing strategies evaluated throughout the study to the BAU nonetheless provides insight into the direction and relative magnitude of the impacts of the policy parameters on the KPIs and other metrics explored. The SFSim model also excludes non-work travel activity and trips that do not interact with the City of San Francisco, thus underestimating the impacts of congestion pricing and incentive

schemes on non-work travel and travel in other parts of the Bay Area. Conversely, the study does not account for the impact that the excluded travel has on commuters under the various congestion pricing strategies. Future iterations of this research should include a more complete representation of travel across trip purposes and geographies in the study region.

In addition, since the travel plans of all agents were kept fixed across all simulation runs, this case study did not account for the potential effects of induced demand from congestion pricing and incentive strategies. The reduction of congestion achieved by congestion pricing and incentives are likely to induce additional travel demand by individuals able and willing to take advantage of the faster travel speeds to make trips they may not have made previously. The induced travel demand may include latent demand by higher income individuals willing to pay the congestion charges to make trips that were previously too slow to be worthwhile, which could further exacerbate the inequities of congestion pricing. On the other hand, incentives for PTNC and PT may also serve latent demand by eligible travelers who are induced to make additional trips that were previously too slow or too costly to be worthwhile. Congestion pricing and incentive schemes may also affect the residential, workplace, and other activity location choices of the population, resulting in induced demand across the region. By keeping travel demand fixed, the optimization results may a) overestimate the long-term reductions in congestion and emissions and b) underestimate the social welfare benefits from improved accessibility.

In addition to including a more representative set of trip purposes in the travel models used by future BISTRO studies, the realism of the travel demand modeled in BISTRO can be improved by including a no-travel option in the mode choice model to represent latent demand for travel. This may reflect the choice to work from home, substitute an errand trip with an online purchase, or simply to not travel for a particular purpose. However, this approach is still limited by an exogenous travel plan generation model which determines the OD pairs and travel times of each potential trip in an agent's travel plan. More complex model development is needed to integrate a multi-step trip generation and mode choice model framework to represent the sensitivity of travel demand more fully. This could be done by running a trip generation model between iterations of BEAM that mutates the initial set of exogenously determined travel plans for each agent based on an aggregate representation of the state of the previous iteration (i.e., the hourly average travel time and cost of travel for each OD TAZ pair and travel mode).

Refinements to the vehicle ownership and vehicle fleet mix configuration would also provide meaningful improvements to the SFSim model and its use for congestion pricing optimization. Although the distributions of vehicle types across the private and TNC vehicle fleets were representative of the region with respect to income groups and on aggregate, respectively, they did not reflect heterogeneity in vehicle ownership and vehicle type preferences with respect to other individual characteristics. Moreover, the SFSim model was configured with ubiquitous ownership of two cars and bikes per household, enabling all commuters to use either option. In reality, socio-demographics and travel patterns are significant predictors of the number and type of vehicles an individual or household owns and thus influence the

distributions of fuel consumption and emissions across vehicle trips in a region. Congestion pricing and incentive strategies have varied impacts on the travel behavior of residents of a region that depend on their willingness to pay and the spatio-temporal characteristics of their trips, including the length and time of the trip as well as access to alternatives to driving of comparable quality. Therefore, a more realistic configuration of the private vehicle fleet mix would more accurately reflect the impact that TDM strategies may have on emissions from fuel consumption across the region. Furthermore, the inclusion of a vehicle ownership model to determine the number and type of vehicles owned by each household as a function of factors in the simulation state (i.e., the travel time and cost of travel by each mode in the previous iteration) would further improve the ability of the model to reflect emergent behavior in response to the policies simulated, such as vehicle purchasing or shedding.

In addition, refinement to the distributions of fuel consumption rates and emissions factors would greatly improve the value of the ABS-based optimization approach for assessing the environmental and energy consumption impacts of congestion pricing applied in this study. While the GHG and $PM_{2.5}$ emissions KPIs were perfectly correlated with the VMT KPI across the sample history, the average GHG and $PM_{2.5}$ emissions per mile driven varied slightly across the samples. I plan to explore these variations and their influencing factors further in future work using a greater variation of vehicle types with varying fuel consumption and emissions rates. In addition, the value of the $PM_{2.5}$ emissions KPI may be improved by refining the specification to measure the impacts of localized emissions instead of the global total emissions that were calculated in this study. Localized estimates of air quality impacts of transportation, which account for the flow of emissions from the point of release to the point of exposure to residents in the surrounding area, are important for addressing environmental justice concerns that arise due to the prevalence of high-capacity roadways in areas where the lowest income and otherwise most disadvantaged populations reside [182]. Since the BEAM ABS model simulates the movements of vehicles at the link-level, estimates of $PM_{2.5}$ emitted by each vehicle on each link in the network are available for any time in the simulation. This data can serve as input to vehicle emissions-intake models that estimate the intake of emissions (i.e., in grams per time period) at a zonal level, which can be aggregated into one or more KPIs. For example, the average $PM_{2.5}$ intake per resident of a neighborhood or demographic group deemed vulnerable to the impacts of air quality would serve as a KPI of the environmental justice impacts of congestion pricing.

Several aspects of the TNC fleet operations in the SFSim model warrant consideration. The fleet management was configured to reposition vehicles using a demand-following algorithm that sought to minimize deadheading between passenger trips. While this is among the most realistic algorithms available for TNC fleet management in the BEAM framework at the time of this study, the fact that the vehicle fleet size remained fixed across all samples may have resulted in excess TNC VMT, particularly in samples where there was an excess of TNC vehicle supply in comparison with the TNC demand. In reality, TNC fleet size would vary in response to demand and other labor market forces. While the TNC fleet size configuration is merely a calibration issue when using ABS for discrete scenario analyses, it becomes a more critical issue when using a fixed travel model and ABS-based optimiza-

tion where optimization of the fleet size is not relevant to the problem at hand. Since the congestion pricing optimization problem in this study is posed as a design problem for a public decision-maker, the same optimization objectives cannot be used to simultaneously optimize the TNC fleet size or other aspects of its operations. The TNC vehicle fleet size adjustment should be incorporated into the vehicle fleet manager in order to appropriately scale the TNC fleet operations according to the revenue-maximizing and other objectives of the service.

On a related note, the PT operations were also kept fixed throughout this study. The inclusion of PT parameters such as fares, route design and scheduling, and the number and type of vehicles serving various routes in the congestion pricing optimization problem studied here would enable an investigation of the potential benefits of reinvestment of toll revenues into PT. The simultaneous optimization of tolls, incentives, and PT service would also provide a better understanding of the tradeoffs in the costs of adjusting PT service versus merely providing incentives. In addition, the inclusion of indirect PTNC service models such as microtransit would enable a more rich exploration of the differences in outcomes from investment in different types of pooled services.

Policy implications and future research directions

The extensive sample history of 879 congestion pricing strategies, 809 of which included a bi-level incentive scheme for PT and PTNC, produced rich insights regarding the influence of the design parameters of this multi-faceted pricing strategy on key policy objectives. In addition to the cordon size and toll rate, the placement of the cordon was shown to have a significant impact on the KPIs. Future work will incorporate the infrastructural and operational costs of each cordon sampled to better reflect the impacts that the location and size of the cordon have on the financial and political feasibility of the congestion charging scheme. When controlling for the cordon design parameters, congestion-related objectives were found to be negatively correlated with PT and PTNC incentives, though less so with the latter. Both incentives also correlated negatively with average CB, demonstrating their potential to alleviate the inequitable impacts of congestion pricing by reducing the cost burden of driving alternatives for those who are the most likely to be significantly affected by the increased cost of driving.

This is the first study to employ a Pareto-based multi-objective optimization algorithm using ABS to study the optimization of congestion pricing. Rather than impose a weighting scheme to translate the objective space into a single dimension, the COMO-CMA-ES algorithm was used to optimize the hypervolume of the Pareto front of a five-dimensional objective function. The COMO-CMA-ES algorithm implemented using BISTRO successfully generated policy designs that iteratively improved the hypervolume of the sample history, thus enriching the exploration of trade offs in the policy objectives. However, it is unlikely that the algorithm was near convergence after the ten iterations that were run. With 879 samples each taking four to five hours to run on a high-performance computer, the efficiency of the algorithm in exploiting the sample history to hone in on a fruitful search direction

left much to be desired. In addition to resource and time constraints that prevented running more iterations, refinements to the problem specification and algorithm hyperparameter settings are needed to improve the performance of the algorithm before initializing another optimization routine for this case study.

Nevertheless, the results of the Pareto-based optimization approach enabled a characterization of the competition and complementarity between policy objectives, confirming that, in the valuation of congestion pricing schemes, congestion-related KPIs are almost perfect complements with one another. On the other hand, the congestion-related KPIs were found to be in direct competition with NPR. The relationship of average CB with other KPIs was more nuanced, demonstrating the difficulty of designing congestion pricing strategies that balance congestion-related objectives with equity considerations while maintaining financial sustainability.

Five Pareto-optimal congestion pricing and incentive strategies were examined in depth. The analysis elucidated how competing objectives can influence the outcomes of the optimization of a multi-faceted pricing policy. In particular, the magnitude of impacts on certain KPIs fluctuated dramatically when comparing strategies that were Pareto optimal for different subsets of KPIs. There is more work to be done investigating the implications of the trade offs between the various outcomes explored in this study. The samples analyzed represented the strategies balancing various pairs and triples of KPIs. Further investigation is needed to understand the influence of constraining the range of outputs investigated and/or tipping the scale toward certain objectives on the design and outcomes of congestion pricing strategies. While the congestion pricing strategies designed by the Pareto-based optimization approach taken in this study produced reductions in VD and VMT of up to 55% compared to the baseline, such extreme mitigations are likely unnecessary and politically infeasible. Future iterations of this study will refine the KPI specifications, input parameter, and output constraints in order to ensure that each KPI is accurately reflecting the network effects on the intended objective and that the range of KPI values explored is within a reasonable and realistic range (e.g., by imposing a lower bound on the congestion-related KPIs). This would help to limit the exploration of design options to those that produce politically feasible outcomes and improve the value of the insights gleaned from analyzing the Pareto-optimal results.

This study found that incentives encourage greater adoption of pooled modes, with both PT and PTNC adoption being strongly positively correlated with average PT and PTNC incentives per trip, respectively. This results in additional reductions in VD, which benefit all road users. In addition to reduced travel time, the reduced costs of travel for the incentive recipients lowers the average CB of travel. PT incentives offer improvements across all KPIs barring NPR, which decreases with respect to the average PT and PTNC incentives per trip. Further research is needed to better understand the trade offs in the use of congestion pricing revenues for incentivization versus investment in the expansion or improvement of PT services. The latter is likely to provide more sustainable mode shifts by tangibly improving the level of service of PT rather than simply reducing the cost. In regions where the quality of PT service relative to driving is variable, as in the San Francisco Bay Area, the targeted

incentivization of PTNC services can facilitate greater adoption of pooled modes of travel by individuals who have a particularly high need or willingness to pay for faster and more reliable service.

Continued research is needed to monitor trends in the willingness of travelers to pool across all types of pooled service options and trip purposes. The results of this study are dependent on estimates of the sensitivity of demand for pooling from a representative survey of Californians in 2018. The lasting effects of the COVID-19 pandemic on pooling are yet to be seen, with PT ridership still lagging behind pre-pandemic levels in many urban areas across the United States [183]. Moreover, newer pooled ride services such as microtransit and shared automated vehicle (SAVs) pilot projects are only recently being deployed, thus the understanding of users' preferences regarding these services is still nascent. This study supports prior research in finding that greater PTNC adoption across a region can produce the economies of scale necessary to achieve higher match rates and higher occupancy trips [173]. As seen in the analysis of the Pareto-optimal strategies, when the PTNC mode split increased to three or even seven percent, about 15 to 28% of those trips pooled four or more passengers per vehicle. Further analysis is needed to understand the characteristics of these high occupancy PTNC trips and their implications for the potential success of investment in microtransit or express bus services to support the roll out of congestion pricing by supplementing existing PT services.

The levels of incentives that accomplished such high PTNC adoption were rather large, on the scale of 15 to 20$ per trip. While only about a third of the population was eligible to receive such incentives, this level of subsidization is likely too high to be feasible in reality. In addition to being economically inefficient, such high incentives run the risk of inducing demand for PTNCs, which would be counterproductive to the goals of the congestion pricing scheme. The model used in this study restricted PTNC service to full-length trips (i.e., PTNC could not be combined with PT). There are a variety of alternative pooling options to consider that were not modeled in this case study that may prove more cost efficient. The SFSim model was limited to door-to-door PTNC service, which picks up and drops off passengers directly from their desired trip start and end locations and did not include the option of taking a PTNC to/from PT. Indirect pooled on-demand ride services that require users to walk some distance to or from pickup/dropoff locations may provide greater economies of scale for pooled rides by reducing both the excess in-vehicle time and VMT from deviating from the shortest common route between passengers. Indirect pooled ride options may serve both medium distance intra-city trips that fill in gaps in the connectivity of fixed-route PT services, or longer distance inter-city trips that provide 'express' service without stops along the way. In addition, targeted incentives for either door-to-door or indirect pooled ride services that connect to PT would help to increase adoption of PT while reducing overall travel time and costs.

The examination of the PTNC trip characteristics in strategy 389 demonstrated the viability of each of these use cases. Since the PTNC trips within San Francisco were, on average, about the same duration as driving to/from PT and only about five to ten minutes faster than PT paired with an active mode, the unrestricted subsidization of door-to-door

PTNC services is not likely to be worth the cost. Instead, incentives for the use of PTNC as a first/last mile connection to PT services combined with strategic investment in express bus services is likely to provide more value. In particular, examination of the distribution of the quality of PT services accessible across the City is recommended to prioritize investment in service improvements and supplementary PTNC incentives that target communities that may be disparately affected by the confluence of poor PT service quality and increased driving costs produced by congestion pricing. Furthermore, the inter-city PTNC trip characteristics in strategy 389 demonstrate the comparative speed of driving versus PT service connecting the City of San Francisco to the rest of the Bay Area. For commutes into the City from the East and South Bay Area, where rail services offer relatively fast travel times but access/egress by active transportation is not widely accessible, investment in PTNC service to/from PT is recommended. Although not modeled in this study, shared micromobility services (e.g., bikesharing, scootersharing) can also expand the catchment area for non-auto access and egress to PT in these regions. A limitation of the BEAM model version used in this study is the inability to combine anything other than walking with another access/egress mode to PT (e.g., a PT-Drive trip that starts with a DA access leg can only be completed by a walking egress leg; the agent does not have the option to request a TNC for the egress leg). Thus, it is likely that PT demand was suppressed for trips that may have been made by using different access and egress modes. Nevertheless, the results suggest that a holistic regional congestion pricing and incentive scheme would benefit from the consideration of investment in or incentivization of services outside of the congestion pricing zone. In the context of the San Francisco Bay Area, the City of San Francisco may achieve additional congestion reduction and equity benefits by making toll revenues available to bolster the use of shared micromobility and pooled ride services to access/egress PT in other cities of the region.

There are numerous directions for refining and expanding the scope of the policy optimization considered in this study. Variations to the pricing scheme, including vehicle-specific (e.g., discount for EV), income-specific, and temporally variable or dynamic pricing structures are all of relevance for further research. Many other dimensions could be considered for price differentiation, including age, race/ethnicity, medical condition, neighborhood classifications, etc. In addition, there are many other secondary effects of the policy that could be considered, including the impacts on reductions in fuel consumption and the resulting revenues, decreased health risks from local emissions, induced demand, and increased traffic safety, to name a few.

The impacts of the congestion pricing and incentive scheme on non-work travel and other travel activity in the region not connected to San Francisco are also of great interest for future work on this case study. In 2019, about 78% and 70% of all intra- and inter-city trips in San Francisco were made for non-work purposes [174]. Since the VoT for non-work travel tends to be lower than for commute trips, travelers are likely to be more sensitive to both the congestion charges and incentives when making other trips not modeled in this study. The implications of the resulting impacts on travel demand, modal demand, and economic activity are of great importance for the evaluation of congestion pricing schemes.

In addition to incorporating a more complete set of trip purposes for trips to/from and in San Francisco, future iterations of this work should include representation of the other travel in the region, at least in an abstracted form. The simulation of all trips across the full Bay Area would drastically increase the resource intensity of an optimization-based study of congestion pricing and thus is not recommended. A potential approach to incorporating other trips in the region may involve estimating the aggregate flow of vehicles on main arterial roads and highways resulting from a prediction of the travel demand and modal split for aggregate pairs of ODs (e.g., the aggregate flow of drivers from City A to City B). This would enable the optimization to account for the feedback loop of impacts between the fluctuations in travel demand to and from San Francisco and those of travel demand outside of the city. A more complete analysis of the aggregate impacts of congestion pricing and incentive schemes at the city level would also be facilitated by this approach.

This study contributes to the development of machine learning-based approaches for the design, optimization, and analysis of data-driven transportation policies. It is the first congestion pricing optimization study to apply a Pareto-based optimization algorithm to the task of congestion pricing optimization in addition to being the first to include incentives in the multi-objective optimization of a congestion pricing scheme. In reality, the policy development process includes so much more than data-driven decision-making. Truly fair and equitable policy making relies on the participation of stakeholders, including outreach to and input from community members. Ideal processes for thoroughly translating community values and needs into policy objectives are not addressed here, but are acknowledged as fundamental to developing equitable transportation policies. This work lays a foundation for further development of algorithmic decision support systems to support policy makers and stakeholders in the complex transition toward truly smart, sustainable, and equitable civil infrastructure.

Chapter 7

Conclusions and Research Directions

Amid the rapid evolution of transportation systems worldwide, there is a pressing need for dynamic data-driven decision-making tools to design and analyze multi-faceted transportation demand management (TDM) strategies. Numerous competing objectives, including the environmental, economic, and social sustainability of the transportation system, pose significant challenges for purely algorithmic strategies - exemplifying the need for human-in-the loop tools that aim to enhance understanding of the inherent trade-offs in policy design while enabling stakeholders to hone in on 'optimal' strategies that use as many levers as feasible. To this end, this dissertation develops a methodological framework for the multi-objective optimization of systemic transportation policies using agent-based simulation and demonstrates the feasibility of optimizing both operational and pricing-based strategies with respect to a handful of policy objectives while producing rich insights on both the individual- and system-level impacts of policy design. A review of the main contributions and key findings of this dissertation including its limitations and directions for future research are provided in this chapter.

7.1 Mechanism Design for Optimal Transportation Pricing Policies

Chapter 3 of this dissertation contributes to the theory of transportation pricing optimization by proving the existence of optimal multi-modal pricing schemes including both tolls and monetary incentives that are optimized with respect to social objectives defined on the basis of the individual characteristics of groups of travelers. Whereas much of the transportation pricing optimization literature adheres to a Utilitarian philosophy, with a focus on minimizing system-wide costs of travel, the model formulated in this chapter seeks to incorporate a broader range of social objectives. By incorporating monetary incentives that can be targeted to specific modes and groups of travelers in a transportation pricing scheme and proving that such schemes can optimize functions of their group-level outcomes, I build a foundation on which to explore the implications of various equity-based objectives on trans-

portation pricing strategy design. The dimensions by which the incentives are specified may include socio-demographic characteristics (e.g., age, income, and physical impairments), occupational characteristics (e.g., essential workers), or spatial attributes of a trip (e.g., origin or destination).

Continuation of this work includes the sensitivity analysis and application of the alternative social choice functions presented in this chapter. Further work is also under way to develop a solution approach to solve the optimal social traffic assignment and incentive problem on road networks at scale. In addition, the accuracy with which the model estimates and therefore optimizes travel times can be improved by incorporating a *dynamic* traffic assignment approach into the model formulation. This would also enable the study of dynamic tolls using this model. Ultimately, I aim to develop a methodology to apply the model developed in this chapter as a tool to 'warm-start' simulation-based optimization studies by generating an initial set of sample solutions that provide a coarse representation of the input-to-output relationships between pricing policy design parameters and objectives.

7.2 BISTRO

The fourth chapter of this dissertation presents BISTRO, an open-source transportation planning and policy decision support tool that uses ABS to conduct multiobjective simulation-based optimization of transportation systems. I designed the system specification, including the specification of input parameters and KPIs. BISTRO was piloted in a machine learning competition at Uber, Inc, which proved the feasibility of using BISTRO to conduct multiobjective simulation-based optimization of multi-dimensional transportation system interventions. The pilot also produced important lessons regarding objective function design and the challenge of producing explainable and interpretable results from simulation-based optimization of transportation policies, particularly when optimizing a diverse set of inputs over a large number of competing objectives.

7.3 Congestion pricing optimization case studies

Chapters five and six of this dissertation build upon the lessons learned from the initial pilot study of BISTRO to develop a methodological approach for multiobjective simulation-based optimization of congestion pricing policies. I develop this methodology across two case studies using two different types of black-box optimization algorithms (Bayesian and evolutionary algorithms) and two different approaches to multi-objective optimization (scalarization and pareto-based). In the first case study, using the benchmark Sioux Faux travel model, I investigated the optimization of both cordon tolls and cordoned mileage fees with respect to a scalarized objective function with various weighting schemes. This study contributes to the field of congestion pricing optimization by showing that mileage-based fees Pareto-dominate simple cordon tolls and by characterizing a three-dimensional Laffer curve of toll

revenue with respect to codon coverage and toll rates. Finally, this study demonstrates the importance of objective function design and elucidates a key challenge in the optimization of congestion pricing schemes in multimodal travel models: congestion-related KPIs are often overpowering and may lead the optimization to produce unrealistic policy designs with unnecessarily large reductions in driving.

The second case study focuses on a more realistic model of the San Francisco Bay Area and includes a multi-level incentive scheme for pooled TNCs and public transit as input parameters to the congestion pricing optimization problem. The addition of incentives and inclusion of public transit revenue in the net public revenue (NPR) KPI enabled a more nuanced analysis of the financial sustainability of congestion pricing schemes with and without incentives included. In contrast to the conclusions of the multi-objective optimization of cordon tolls and cordoned mileage-based fees in chapter 5 that found that the toll revenue and congestion objectives have a concave relationship, the case study in chapter 6 found that, in the optimization of cordoned mileage fees with incentives, NPR is at odds with congestion reduction. This is due to the added congestion mitigation effects of distributing the toll revenue as PT and PTNC incentives for low-income travelers. This study demonstrates the potential of incentives to reduce the inequities of congestion pricing schemes while contributing to the further reduction of congestion and emissions. The use of a Pareto-based optimization algorithm enabled an exploration of the set of Pareto-optimal congestion pricing and incentive strategies that optimally trade off various KPIs. While the analysis of five Pareto-optimal policies demonstrated the key differences across policies and the complexity of the relationships between various KPIs with behavioral and operation factors of the transportation system, more work is needed to cull out additional insights from this study.

7.4 Future work

This dissertation demonstrates that a robust policy optimization framework relies heavily on the design of the objective function. While the case studies in my dissertation applied just one of the three alternative social choice functions presented in chapter 3, I plan to continue developing a principled and equitable approach for defining and operationalizing the notion of social optimality with respect to both equity and sustainability. The level of service KPIs implemented in BISTRO, including measures of public transit vehicle crowding, transportation cost burden, and accessibility are a first step in this regard, and the research I have conducted thus far on the impacts of weighted and Pareto-based optimization have illuminated the challenges in automating public policy decision-making.

Additional technical challenges for the deployment of data-driven TDM strategies include stochasticity, resource constraints, and data privacy. BISTRO offers functionality for scenario-based optimization wherein the optimization may be conducted across multiple baselines (as opposed to a single BAU), each representing a potential variation in the distribution of land use, mobility preferences, or other factors that may affect the distribution of supply of or demand for transportation. Further model development is needed

to incorporate the sensitivity of travel demand in the BISTRO framework to enable the representation of induced demand in future studies. This is important to ensure that the optimization results produced by BISTRO are robust to potential changes in the quantity and characteristics of the travel demand in a region that may occur in response to system interventions. In addition, BISTRO may be used to develop algorithms for transfer learning of policies across varying scenarios and geographies. Such methodologies would characterize hyper-parameters representing factors in the transportation network, services, and demand profile that are significant across scenarios or geographies thereby increasing the efficiency and robustness of the optimization. In addition, BISTRO can be used to study competition in multi-agent systems such that may arise when the control of a mobility marketplace is decentralized across public and private operators. These issues are of immediate concern, as we have already seen the negative effects of independently operating on-demand mobility services (e.g., Uber and Lyft, competing micromobility services), which will likely amplify as additional operators and services enter the market. Another potential arena for disruption within the transportation system is that of energy management, with the impending rapid electrification of vehicle fleets across the globe likely to bring forth new challenges in managing the distribution of and access to energy for electric vehicle charging.

Furthermore, it is imperative that automated control of civil systems be capable of operating independently of personally identifiable or otherwise compromising data. While many of the KPIs that I have used in my research thus far have benefited from full access to individual-level data, I plan to improve the robustness of my approach by designing mechanisms that rely only on minimally sufficient data from the transportation system. I intend to explore the extent to which such a system can benefit society by preserving both privacy and resources. In particular, I am interested in quantifying the tradeoffs between the efficiency and efficacy of TDM mechanisms that seek to optimize transportation equity with limited data access. In the short to medium term, this work will aid in the transition from a heavy reliance on infrequent and expensive travel surveys to the use of passive data streams for transportation planning and policy.

Bibliography

[1] International Energy Agency (IEA). *Transport*. 2022. URL: https://www.iea.org/reports/transport.

[2] *Public Road Mileage, Lane-Miles, and VMT 1900 - 2019*. Tech. rep. Office of Highway Policy Information, Policy and Governmental Affairs, Federal Highway Administration, U.S. Department of Transportation, 2019. URL: https://www.fhwa.dot.gov/policyinformation/statistics/2019/vmt421c.cfm.

[3] Michael Burrows Charlynn Burd and Brian McKenzie. *Travel Tim ork in the United States: 2019*. Tech. rep. United States Census Bureau, 2021. URL: https://www.census.gov/content/dam/Census/library/publications/2021/acs/acs-47.pdf.

[4] *INRIX: Congestion Costs Each American Nearly 100 hours, $1,400 a Year*. Tech. rep. INRIX, INC., 2020. URL: https://inrix.com/press-releases/2019-traffic-scorecard-us/.

[5] Mi Diao, Hui Kong, and Jinhua Zhao. "Impacts of transportation network companies on urban mobility". In: *Nature Sustainability* 4.6 (Feb. 2021), pp. 494–500. DOI: 10.1038/s41893-020-00678-z.

[6] Sneha Roy et al. "Why is traffic congestion getting worse? A decomposition of the contributors to growing congestion in San Francisco-Determining the Role of TNCs". In: *Case Studies on Transport Policy* 8.4 (2020), pp. 1371–1382. ISSN: 2213-624X. DOI: https://doi.org/10.1016/j.cstp.2020.09.008.

[7] Bruce Schaller. *The New Automobility: Lyft, Uber and the Future of American Cities*. Report. Schaller Consulting, 2018. URL: http://www.schallerconsult.com/rideservices/automobility.pdf.

[8] Bruce Schaller. *Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City*. Report. Schaller Consulting, 2017. URL: http://www.schallerconsult.com/rideservices/unsustainable.pdf.

[9] California Air Resources Board (CARB). *2018 Base-year Emissions Inventory Report*. data retrieved from CARB, https://ww2.arb.ca.gov/resources/documents/2018-base-year-emissions-inventory-report, 2019.

[10] Judd Cramer and Alan B. Krueger. "Disruptive Change in the Taxi Business: The Case of Uber". In: *American Economic Review* 106.5 (May 2016), pp. 177–182. DOI: 10.1257/aer.p20161002.

[11] Alejandro Henao and Wesley E. Marshall. "The impact of ride-hailing on vehicle miles traveled". In: *Transportation* 46.6 (Sept. 2018), pp. 2173–2194. DOI: 10.1007/s11116-018-9923-2.

[12] Luis Martinez Jose Viegas. *Shared Mobility: Innovation for Liveable Cities*. Tech. rep. 2016. URL: https://www.itf-oecd.org/sites/default/files/docs/shared-mobility-liveable-cities.pdf.

[13] Boston Consulting Group (BCG) World Economic Forum (WEF). *Reshaping Urban Mobility with Autonomous Vehicles Lessons from the City of Boston*. Tech. rep. June 2018. URL: https://www3.weforum.org/docs/WEF_Reshaping_Urban_Mobility_with_Autonomous_Vehicles_2018.pdf.

[14] Gordon S. Bauer, Jeffery B. Greenblatt, and Brian F. Gerke. "Cost, Energy, and Environmental Impact of Automated Electric Taxi Fleets in Manhattan". In: *Environmental Science & Technology* 52.8 (2018). PMID: 29589439, pp. 4920–4928. DOI: 10.1021/acs.est.7b04732.

[15] Jeffery B. Greenblatt and Samveg Saxena. "Autonomous taxis could greatly reduce greenhouse-gas emissions of US light-duty vehicles". In: *Nature Climate Change* 5.9 (July 2015), pp. 860–863. DOI: 10.1038/nclimate2685.

[16] Colin J. R. Sheppard et al. "Private Versus Shared, Automated Electric Vehicles for U.S. Personal Mobility: Energy Use, Greenhouse Gas Emissions, Grid Integration and Cost Impacts". In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3575130.

[17] Susan Shaheen and Adam Cohen. "Shared ride services in North America: definitions, impacts, and the future of pooling". In: *Transport Reviews* 39.4 (2019), pp. 427–442. ISSN: 0144-1647. DOI: https://doi.org/10.1080/01441647.2018.1497728.

[18] Michael Ostrovsky and Michael Schwarz. "Carpooling and the Economics of Self-Driving Cars". In: *Proceedings of the 2019 ACM Conference on Economics and Computation*. EC '19: ACM Conference on Economics and Computation. Phoenix AZ USA: ACM, June 17, 2019, pp. 581–582. ISBN: 978-1-4503-6792-9. DOI: 10.1145/3328526.3329625.

[19] Jessica R. Lazarus et al. "To Pool or Not to Pool? Understanding opportunities, challenges, and equity considerations to expanding the market for pooling". In: *Transportation Research Part A: Policy and Practice* 148 (June 2021), pp. 199–222. ISSN: 09658564. DOI: 10.1016/j.tra.2020.10.007.

[20] Scheff J Liu L Miller HJ. "The impacts of COVID-19 pandemic on public transit demand in the United States". In: *PLoS ONE* 15.11 (2020). DOI: https://doi.org/10.1371/journal.pone.0242476.

[21] Yi Qi et al. "Impacts of COVID-19 on public transit ridership". In: *International Journal of Transportation Science and Technology* (2021). ISSN: 2046-0430. DOI: https://doi.org/10.1016/j.ijtst.2021.11.003.

[22] William S. Vickrey. "Pricing in Urban and Suburban Transport". In: *The American Economic Review* 53.2 (1963), pp. 452–465. ISSN: 00028282. URL: http://www.jstor.org/stable/1823886.

[23] Arthur Cecil Pigou. *The economics of welfare*. Palgrave Macmillan, 2013.

[24] Jonas Eliasson. "A cost–benefit analysis of the Stockholm congestion charging system". In: *Transportation Research Part A: Policy and Practice* 43.4 (2009), pp. 468–480. ISSN: 0965-8564. DOI: https://doi.org/10.1016/j.tra.2008.11.014.

[25] Anders Karlström and Joel P. Franklin. "Behavioral adjustments and equity effects of congestion pricing: Analysis of morning commutes during the Stockholm Trial". In: *Transportation Research Part A: Policy and Practice* 43.3 (2009). Stockholm Congestion Charging Trial, pp. 283–296. ISSN: 0965-8564. DOI: https://doi.org/10.1016/j.tra.2008.09.008.

[26] Georgina Santos, Kenneth Button, and Roger G. Noll. "London Congestion Charging". In: *Brookings-Wharton Papers on Urban Affairs* (2008), pp. 177–234. ISSN: 15287084. URL: http://www.jstor.org/stable/25609551.

[27] *National Inventory of Specialty Lanes and Highways: Technical Report*. Tech. rep. FHWA-HOP-20-043. Office of Operations, Federal Highway Administration, U.S. Department of Transportation, 2021. URL: https://ops.fhwa.dot.gov/publications/fhwahop20043/ch3.htm.

[28] Daniel G. Chatman and Michael Manville. "Equity in Congestion-priced Parking: A Study of SFpark, 2011 to 2013". In: *Journal of Transport Economics and Policy (JTEP)* 52.3 (2018), pp. 239–266. ISSN: 0022-5258.

[29] Joe Castiglione, Mark Bradley, and John Gliebe. *Activity-Based Travel Demand Models: A Primer*. 2016, pp. 7–8. ISBN: 9780309273992. DOI: 10.17226/22357.

[30] Sidney A. Feygin et al. "BISTRO: Berkeley Integrated System for Transportation Optimization". In: *ACM Trans. Intell. Syst. Technol.* 11.4 (June 2020). ISSN: 2157-6904. DOI: 10.1145/3384344.

[31] Jan Rouwendal and Erik T. Verhoef. "Basic economic principles of road pricing: From theory to applications". In: *Transport Policy* 13.2 (Mar. 2006), pp. 106–114. ISSN: 0967070X. DOI: 10.1016/j.tranpol.2005.11.007.

[32] Theodore Tsekeris and Stefan Voß. "Design and evaluation of road pricing: state-of-the-art and methodological advances". In: *NETNOMICS: Economic Research and Electronic Networking* 10.1 (Apr. 2009), pp. 5–52. ISSN: 1385-9587, 1573-7071. DOI: 10.1007/s11066-008-9024-z.

[33] Erik T. Verhoef. "The implementation of marginal external cost pricing in road transport". In: *Papers in Regional Science* 79.3 (July 1, 2000), pp. 307–332. ISSN: 1056-8190. DOI: 10.1007/PL00013611.

[34] Kenneth A. Small and Jia Yan. "The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation". In: *Journal of Urban Economics* 49.2 (Mar. 2001), pp. 310–336. ISSN: 00941190. DOI: 10.1006/juec.2000.2195.

[35] Yang Liu, Xiaolei Guo, and Hai Yang. "Pareto-improving and revenue-neutral congestion pricing schemes in two-mode traffic networks". In: *NETNOMICS: Economic Research and Electronic Networking* 10.1 (Apr. 2009), pp. 123–140. ISSN: 1385-9587, 1573-7071. DOI: 10.1007/s11066-008-9018-x.

[36] Jerome Mayet and Mark Hansen. "Congestion Pricing with Continuously Distributed Values of Time". In: *Journal of Transport Economics and Policy* 34 (2021), p. 12.

[37] Erik T. Verhoef, Kenneth A. Small, and Ken A. Small. "Product Differentiation on Roads: Constrained Congestion Pricing with Heterogeneous Users". In: *Journal of Transport Economics and Policy* 38.1 (2004), pp. 127–156. ISSN: 00225258.

[38] Hanna Armelius. "An Integrated Approach to Urban Road Pricing". In: *Journal of Transport Economics and Policy* 39 (2005), p. 19.

[39] Carlos F. Daganzo and Lewis J. Lehe. "Distance-dependent congestion pricing for downtown zones". In: *Transportation Research Part B: Methodological* 75 (2015), pp. 89–99. ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2015.02.010.

[40] Paolo Ferrari. "Road pricing and users' surplus". In: *Transport Policy* 12.6 (Nov. 2005), pp. 477–487.

[41] Amihai Glazer and Esko Niskanen. "Which Consumers Benefit from Congestion Tolls?" In: *Journal of Transport Economics and Policy* 34 (2021), p. 12.

[42] A.D May et al. "The impact of cordon design on the performance of road pricing schemes". In: *Transport Policy* 9.3 (July 2002), pp. 209–220. ISSN: 0967070X. DOI: 10.1016/S0967-070X(02)00031-8.

[43] Kenneth A. Small. "Road Pricing and Public Transport". In: *Research in Transportation Economics* 9 (2004), pp. 133–158. ISSN: 07398859. DOI: 10.1016/S0739-8859(04)09006-7.

[44] Younes Hamdouch et al. "Congestion pricing for multi-modal transportation systems". In: *Transportation Research Part B: Methodological* 41.3 (Mar. 2007), pp. 275–291. ISSN: 01912615. DOI: 10.1016/j.trb.2006.04.003.

[45] Takuya Maruyama and Agachai Sumalee. "Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-chain equilibrium model". In: *Transportation Research Part A: Policy and Practice* 41.7 (Aug. 2007), pp. 655–671. ISSN: 09658564. DOI: 10.1016/j.tra.2006.06.002.

[46] Kenneth A. Small et al. "Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design [with Comments]". In: *Brookings-Wharton Papers on Urban Affairs* (2006), pp. 53–96. ISSN: 15287084. URL: http://www.jstor.org/stable/25067428.

[47] A. Sumalee. "Optimal Toll Ring Design with Spatial Equity Constraint: An Evolutionary Approach". In: 2003.

[48] Hai Yang and Xiaoning Zhang. "Multiclass Network Toll Design Problem with Social and Spatial Equity Constraints". In: *Journal of Transportation Engineering* 128.5 (2002), pp. 420–428. ISSN: 0733-947X, 1943-5436. DOI: 10.1061/(ASCE)0733-947X(2002)128:5(420).

[49] Hai Yang and Hai-Jun Huang. "Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes". In: *Transportation Research Part A: Policy and Practice* 33.2 (Feb. 1999), pp. 139–155.

[50] Mark Burris et al. "The impact of HOT lanes on carpools". In: *Research in Transportation Economics* 44 (June 2014), pp. 43–51. ISSN: 07398859. DOI: 10.1016/j.retrec.2014.04.004.

[51] Jaimyoung Kwon and Pravin Varaiya. "Effectiveness of California's High Occupancy Vehicle (HOV) system". In: *Transportation Research Part C: Emerging Technologies* 16.1 (Feb. 2008), pp. 98–115. ISSN: 0968090X. DOI: 10.1016/j.trc.2007.06.008.

[52] Susan Shaheen and Ata Kahn. *Shared Mobility and Automated Vehicles: Responding to Socio-Technical Changes and Pandemics*. London, England: The Institution of Engineering and Technology (IET), 2021, pp. 1–12. ISBN: ISBN-13: 978-1-78561-862-8.

[53] Joakim Ekström, Leonid Engelson, and Clas Rydergren. "Heuristic algorithms for a second-best congestion pricing problem". In: *NETNOMICS: Economic Research and Electronic Networking* 10.1 (Apr. 2009), pp. 85–102. ISSN: 1385-9587, 1573-7071. DOI: 10.1007/s11066-008-9019-9.

[54] A.D May and D.S Milne. "Effects of alternative road pricing systems on network performance". In: *Transportation Research Part A: Policy and Practice* 34.6 (2000), pp. 407–436. ISSN: 0965-8564. DOI: https://doi.org/10.1016/S0965-8564(99)00015-4.

[55] André de Palma, Moez Kilani, and Robin Lindsey. "Congestion pricing on a road network: A study using the dynamic equilibrium simulator METROPOLIS". In: *Transportation Research Part A: Policy and Practice* 39.7 (Aug. 2005), pp. 588–611. ISSN: 09658564. DOI: 10.1016/j.tra.2005.02.018.

[56] Georgina Santos and Laurent Rojey. "Distributional impacts of road pricing: The truth behind the myth". In: *Transportation* 31.1 (Feb. 2004), pp. 21–42. ISSN: 0049-4488. DOI: 10.1023/B:PORT.0000007234.98158.6b.

[57] Yingyan Lou, Yafeng Yin, and Jorge A. Laval. "Optimal dynamic pricing strategies for high-occupancy/toll lanes". In: *Transportation Research Part C: Emerging Technologies* 19.1 (Feb. 2011), pp. 64–74. ISSN: 0968090X. DOI: 10.1016/j.trc.2010.03.008.

[58] Xijie Li et al. "Cordon- or Link-Based Pricing: Environment-Oriented Toll Design Models Development and Application". In: *Sustainability* 11 (Jan. 2019), p. 258. DOI: 10.3390/su11010258.

[59] Yafeng Yin and Siriphong Lawphongpanich. "Internalizing emission externality on road networks". In: *Transportation Research Part D: Transport and Environment* 11.4 (2006), pp. 292–301. ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2006.05.003.

[60] Zhiyuan Liu et al. "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics". In: *Transportation Research Part C: Emerging Technologies* 79 (June 2017), pp. 58–72. DOI: 10.1016/j.trc.2017.03.011.

[61] Simon Shepherd and Agachai Sumalee. "A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems". In: *Networks and Spatial Economics* 4 (June 2004), pp. 161–179. DOI: 10.1023/B:NETS.0000027771.13826.3a.

[62] Nan Zheng et al. "A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model". In: *Transportation Research Part A: Policy and Practice* 46.8 (2012), pp. 1291–1303. ISSN: 0965-8564. DOI: https://doi.org/10.1016/j.tra.2012.05.006.

[63] Chen Xiqun et al. "Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network." In: *Computer-Aided Civil and Infrastructure Engineering* 29.5 (2014), pp. 359–381.

[64] Ziyuan Gu, S. Travis Waller, and Meead Saberi. "Surrogate-based toll optimization in a large-scale heterogeneously congested network". In: *Computer-Aided Civil and Infrastructure Engineering* 34.8 (2019), pp. 638–653. DOI: 10.1111/mice.12444.

[65] L. Fleischer, K. Jain, and M. Mahdian. "Tolls for Heterogeneous Selfish Users in Multicommodity Networks and Generalized Congestion Games". In: *45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2004. DOI: 10.1109/focs.2004.69.

[66] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. "Pricing network edges for heterogeneous selfish users". In: *Proceedings of the thirty-fifth ACM symposium on Theory of computing - STOC '03*. ACM Press, 2003. DOI: 10.1145/780542.780618.

[67] Stella C. Dafermos. "Toll Patterns for Multiclass-User Transportation Networks". In: *Transportation Science* 7.3 (1973), pp. 211–223. DOI: 10.1287/trsc.7.3.211.

[68] Malachy Carey and Ashok Srinivasan. "Externalities, Average and Marginal Costs, and Tolls on Congested Networks with Time-Varying Flows". In: *Operations Research* 41.1 (1993), pp. 217–231. ISSN: 0030364X, 15265463. URL: http://www.jstor.org/stable/171955.

[69] André de Palma and Robin Lindsey. "Congestion Pricing with Heterogeneous Travelers: A General-Equilibrium Welfare Analysis". In: *Networks and Spatial Economics* 4.2 (June 2004), pp. 135–160. DOI: 10.1023/b:nets.0000027770.27906.82.

[70] Andrew W. Evans. "Road Congestion Pricing: When Is It a Good Policy?" In: *Journal of Transport Economics and Policy* 26.3 (1992), pp. 213–243. ISSN: 00225258. URL: http://www.jstor.org/stable/20052985.

[71] Liang Wen and Richard Eglese. "Minimizing CO2e emissions by setting a road toll". In: *Transportation Research Part D: Transport and Environment* 44 (2016), pp. 1–13. ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2015.12.019.

[72] Lili Du et al. "A Quantitative and Systematic Methodology to Investigate Energy Consumption Issues in Multimodal Intercity Transportation Systems". In: *International Journal of Transportation Science and Technology* 4.3 (2015). Special Issue on Travel Demand Management, pp. 229–256. ISSN: 2046-0430. DOI: https://doi.org/10.1260/2046-0430.4.3.229.

[73] Yafeng Yin and Hai Yang. "Optimal Tolls with a Multiclass, Bicriterion Traffic Network Equilibrium". In: *Transportation Research Record* 1882.1 (2004), pp. 45–52. DOI: 10.3141/1882-06.

[74] Devansh Jalota et al. "When Efficiency meets Equity in Congestion Pricing and Revenue Refunding Schemes". In: *Equity and Access in Algorithms, Mechanisms, and Optimization*. ACM, Oct. 2021. DOI: 10.1145/3465416.3483296.

[75] Kenneth A. Small. "Using the revenues from congestion pricing". In: *Transportation* 19.4 (1992), pp. 359–381. DOI: 10.1007/bf01098639.

[76] Leonardo J. Basso and Sergio R. Jara-Díaz. "Integrating congestion pricing, transit subsidies and mode choice". In: *Transportation Research Part A: Policy and Practice* 46.6 (2012), pp. 890–900. DOI: 10.1016/j.tra.2012.02.013.

[77] Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. "A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications". In: *IEEE Transactions on Evolutionary Computation* 22.2 (2018), pp. 276–295. DOI: 10.1109/TEVC.2017.2712906.

[78] Tim Roughgarden. *Selfish Routing and The Price of Anarchy*. Vol. 74. Jan. 2005. ISBN: 978-0-262-18243-0.

[79] Rafael H. M. Pereira, Tim Schwanen, and David Banister. "Distributive justice and equity in transportation". In: *Transport Reviews* 37.2 (Nov. 2016), pp. 170–191. DOI: 10.1080/01441647.2016.1257660.

[80] Todd Litman. *Evaluating Transportation Equity - Guidance for Incorporating Distributional Impacts in Transport Planning*. Tech. rep. Victoria Transport Policy Institute, 2022. URL: https://vtpi.org/equity.pdf.

[81] Susan Shaheen et al. *Travel Behavior: Shared Mobility and Transportation Equity*. Tech. rep. PL-18-007. Office of Policy and Governmental Affairs Federal Highway Administration, 2017. URL: https://www.fhwa.dot.gov/policy/otps/shared_use_mobility_equity_final.pdf.

[82] Mogeng Yin. "Activity-based Urban Mobility Modeling from Cellular Data". PhD thesis. UC Berkeley, 2018, p. 105. ISBN: 9780438324398.

[83] Hong Zheng. "A Primer for Agent-Based Simulation and Modeling in Transportation Applications". In: *Development* (2013), p. 75.

[84] Andreas Horni, Kai Nagel, and Kay W Axhausen. *The multi-agent transport simulation MATSim*. Ubiquity Press London, 2016.

[85] Sebastian Hörl. "Implementation of an autonomous taxi service in a multi-modal traffic simulation using MATSim Master thesis in Complex Adaptive Systems". In: June (2016). DOI: 10.13140/RG.2.1.2060.9523.

[86] Michael Grant et al. *Performance-based planning and programming guidebook*. Tech. rep. United States. Federal Highway Administration, 2013.

[87] United States Department of Transportation. *Trends in Statewide Long-Range Transportation Plans: Core and Emerging Topics*. 2012. URL: https://www.planning.dot.gov/documents/State%7B%5C_%7Dplans%7B%5C_%7Dreport%7B%5C_%7D508%7B%5C_%7DA.PDF.

[88] Federal Highway Administration/Federal Transit Administration. *Public Engagement: Case Studies and Notable Practices*. 2019. URL: https://planning.dot.gov/focus_caseStudies.aspx.

[89] Gian-Claudia Sciara. "Metropolitan Transportation Planning: Lessons From the Past, Institutions for the Future". In: *Journal of the American Planning Association* 83.3 (2017), pp. 262–276. DOI: 10.1080/01944363.2017.1322526.

[90] Federal Highway Administration/Federal Transit Administration. *The Transportation Planning Process Key Issues: A Briefing Book for Transportation Decisionmakers, Officials, and Staff*. Tech. rep. Washington, DC, USA, 2016.

[91] United States Government Accountability Office. *Report to the Ranking Member, Committee on Environment and Public Works: U.S. Senate– Metropolitan Planning Organizations: Options Exist to Enhance Transportation Planning Capacity and Federal Oversight*. Tech. rep. Washington, DC, USA, 2009.

[92] Joshua Auld et al. "POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations". In: *Transportation Research Part C: Emerging Technologies* 64 (2016), pp. 101–116.

[93] Ebru Vesile Ocalir-Akunal. "Decision Support Systems in Transport Planning". In: *Procedia Engineering* 161 (2016), pp. 1119–1126. ISSN: 18777058. DOI: 10.1016/j.proeng.2016.08.518.

[94] Zaheer Khan et al. "ICT enabled participatory urban planning and policy development: The UrbanAPI project". In: *Transforming Government: People, Process and Policy* 8.2 (2014), pp. 205–229. ISSN: 17506166. DOI: 10.1108/TG-09-2013-0030.

[95] Carlos Lima Azevedo et al. "Tripod: sustainable travel incentives with prediction, optimization, and personalization". In: *Transportation Research Board 97th Annual Meeting*. 2018.

[96] Maria Nadia Postorino and Giuseppe M.L. Sarné. "Agents meet traffic simulation, control and management: A review of selected recent contributions". In: *CEUR Workshop Proceedings* 1664 (2016), pp. 112–117. ISSN: 16130073.

[97] Association of Metropolitan Planning Organizations. *ActivitySim — An open platform for activity-based travel modeling*. 2019.

[98] Joe Castiglione, Mark Bradley, and John Gliebe. *Activity-Based Travel Demand Models: A Primer*. 2016, pp. 7–8. ISBN: 9780309273992. DOI: 10.17226/22357.

[99] L. Smith, R. Beckman, and K. Baggerly. "TRANSIMS: Transportation analysis and simulation system". In: *Fifth National Conference on Transportation Planning Methods Applications-Volume II: A Compendium of Papers Based on a Conference Held in Seattle, Washington in April 1995Transportation Research Boad and Washington State Department of Transportation*. 1995. DOI: 10.2172/88648. URL: http://www.osti.gov/servlets/purl/88648-fgWOUT/webviewable/.

[100] Michael Behrisch et al. "SUMO–simulation of urban mobility: an overview". In: *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind. 2011.

[101] Carlo Castagnari et al. "Tangramob: an agent-based simulation framework for validating urban smart mobility solutions". In: *arXiv preprint arXiv:1805.10906* (2018).

[102] Joseph Kamel et al. "Exploring the Impact of User Preferences on Shared Autonomous Vehicle Modal Split: A Multi-Agent Simulation Approach". In: *Transportation Research Procedia* 37.September 2018 (2019), pp. 115–122. ISSN: 23521465. DOI: 10.1016/j.trpro.2018.12.173.

[103] Paul Salvini and Eric J Miller. "ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems". In: *Networks and spatial economics* 5.2 (2005), pp. 217–234.

[104] Paul Waddell. "UrbanSim: Modeling urban development for land use, transportation, and environmental planning". In: *Journal of the American planning association* 68.3 (2002), pp. 297–314.

[105] T Nicolai and K Nagel. "Coupling transport and land use: Investigating accessibility indicators for feedback from a travel to a land-use model". In: *Latsis Symposium*. 2012, pp. 12–16.

[106] Rolf Moeckel et al. "Trends in integrated land use/transport modeling: An evaluation of the state of the art". In: *Journal of Transport and Land Use* 11.1 (2018), pp. 463–476. DOI: 10.5198/jtlu.2018.1205.

[107] Thomas W Nicolai and Kai Nagel. *Coupling MATSim and UrbanSim: Software design issues*. Tech. rep. SustainCity Working Paper, 2010.

[108] Dominik Ziemke, Kai Nagel, and Rolf Moeckel. "Towards an agent-based, integrated land-use transport modeling system". In: *Procedia computer science* 83 (2016), pp. 958–963.

[109] US Department of Energy. *Energy Efficient Mobility Systems: 2018 Annual Progress Report*. Tech. rep. US Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office, Dec. 2018.

[110] Daniel Delling, Thomas Pajor, and Renato F Werneck. "Round-based public transit routing". In: *Transportation Science* 49.3 (2014), pp. 591–604.

[111] Matthew Wigginton Conway, Andrew Byrd, and Marco van der Linden. "Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule and Headway-Based Networks". In: *Transportation Research Record: Journal of the Transportation Research Board* 2653.1 (2017), pp. 45–53. ISSN: 0361-1981. DOI: 10.3141/2653-06.

[112] Gul A Agha et al. "Actors: a model for reasoning about open distributed systems". In: *Formal methods for distributed processing: a survey of object-oriented approaches* (2001), pp. 155–176.

[113] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[114] Moshe E Ben-Akiva, Steven R Lerman, and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press, 1985.

[115] Carolina Osorio and Michel Bierlaire. "A simulation-based optimization framework for urban transportation problems". In: *Operations Research* 61.6 (2013), pp. 1333–1345.

[116] Andrew R Conn, Katya Scheinberg, and Luis N Vicente. *Introduction to derivative-free optimization*. Vol. 8. Siam, 2009.

[117] Russell R. Barton and Martin Meckesheimer. "Chapter 18 Metamodel-Based Simulation Optimization". In: *Handbooks in Operations Research and Management Science* 13.C (2006), pp. 535–574. ISSN: 09270507. DOI: 10.1016/S0927-0507(06)13018-2.

[118] Carl Edward Rasmussen. "Gaussian processes in machine learning". In: *Summer School on Machine Learning*. Springer. 2003, pp. 63–71.

[119] Bobak Shahriari et al. "Taking the human out of the loop: A review of Bayesian optimization". In: *Proceedings of the IEEE* 104.1 (2015), pp. 148–175.

[120] Ian Dewancker, Michael McCourt, and Scott Clark. "Bayesian Optimization Primer". 2015. DOI: 10.1016/j.jss.2013.05.010.

[121] Peter I Frazier. "A tutorial on Bayesian optimization". In: *arXiv:1807.02811* (2018).

[122] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. "Sequential model-based optimization for general algorithm configuration". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6683 LNCS (2011), pp. 507–523. ISSN: 03029743.

[123] James S Bergstra et al. "Algorithms for hyper-parameter optimization". In: *Advances in neural information processing systems*. 2011, pp. 2546–2554.

[124] James Bergstra, Daniel L K Yamins, and D Cox. "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures". In: *Proceedings of the 30th International Conference on Machine Learning*. 2013, pp. 115–123. URL: http://jmlr.org/proceedings/papers/v28/bergstra13.html.

[125] Donald R Jones, Matthias Schonlau, and William J Welch. "Efficient global optimization of expensive black-box functions". In: *Journal of Global optimization* 13.4 (1998), pp. 455–492.

[126] Peter Frazier, Warren Powell, and Savas Dayanik. "The knowledge-gradient policy for correlated normal beliefs". In: *INFORMS journal on Computing* 21.4 (2009), pp. 599–613.

[127] Jian Wu et al. "Bayesian optimization with gradients". In: *Advances in Neural Information Processing Systems*. 2017, pp. 5267–5278.

[128] Jasper Snoek, Hugo Larochelle, and Ryan Adams. "Practical Bayesian Optimization of Machine Learning Algorithms". In: *Advances in neural information processing systems*. 2012, pp. 2951–2959. DOI: 10.1016/s2468-2667(17)30214-1.

[129] Kevin Swersky, Jasper Snoek, and Ryan Prescott Adams. "Freeze-thaw Bayesian optimization". In: *arXiv:1406.3896* (2014).

[130] Carolina Osorio and Linsen Chong. "A Computationally Efficient Simulation-Based Optimization Algorithm for Large-Scale Urban Transportation Problems". In: *Transportation Science* 49.3 (2015), pp. 623–636. ISSN: 0041-1655. DOI: 10.1287/trsc.2014.0550.

[131] Carolina Osorio and Kanchana Nanduri. "Urban transportation emissions mitigation: Coupling high-resolution vehicular emissions and traffic models for traffic signal optimization". In: *Transportation Research Part B: Methodological* 81 (2015), pp. 520–538. ISSN: 01912615. DOI: 10.1016/j.trb.2014.12.007.

[132] Linsen Chong and Carolina Osorio. "A simulation-based optimization algorithm for dynamic large-scale urban transportation problems". In: *Transportation Science* 52.3 (2017), pp. 637–656.

[133] Ennio Cascetta. "A stochastic process approach to the analysis of temporal dynamics in transportation networks". In: *Transportation Research Part B* 23.1 (1989), pp. 1–17. ISSN: 01912615. DOI: 10.1016/0191-2615(89)90019-2.

[134] Gunnar Flötteröd. "A search acceleration method for optimization problems with transport simulation constraints". In: *Transportation Research Part B: Methodological* 98 (2017), pp. 1339–1351. ISSN: 01912615. DOI: 10.1016/j.trb.2016.12.009.

[135] Gregory Bucci, Chris Calley, Michael Green, et al. *FHWA Research and Technology Evaluation: Agent-Based Modeling and Simulation*. Tech. rep. United States. Federal Highway Administration. Office of Corporate Research, 2018.

[136] Peter Vovsha et al. "New Features of Population Synthesis". In: *94th Annual Meeting of the Transportation Research Board* 250 (2015), pp. 1–20.

[137] Artem Chakirov and Pieter J Fourie. "Enriched sioux falls scenario with dynamic and disaggregate demand". In: *Arbeitsberichte Verkehrs-und Raumplanung* 978 (2014).

[138] David E Goldberg and John H Holland. "Genetic algorithms and machine learning". In: *Machine learning* 3.2 (1988), pp. 95–99.

[139] Roger D Peng. "Reproducible research in computational science". In: *Science* 334.6060 (2011), pp. 1226–1227.

[140] Andrew Morin et al. "Shining light into black boxes". In: *Science* 336.6078 (2012), pp. 159–160.

[141] Geir Kjetil Sandve et al. "Ten Simple Rules for Reproducible Computational Research". In: *PLOS Computational Biology* 9.10 (Oct. 2013), pp. 1–4. DOI: 10.1371/journal.pcbi.1003285.

[142] Justine S. Hastings et al. "Unlocking Data to Improve Public Policy". Mar. 2019. DOI: 10.31219/osf.io/28krq.

[143] Andrea M. Robitaille, Jasmy Methipara, and Lei Zhang. "Effectiveness and Equity of Vehicle Mileage Fee at Federal and State Levels". In: *Transportation Research Record* 2221.1 (2011), pp. 27–38. DOI: 10.3141/2221-04.

[144] Cohen M. D'Agostino, P. Pellaton, and B. White. *Equitable Congestion Pricing*. Tech. rep. UC Office of the President: University of California Institute of Transportation Studies, 2020. DOI: 10.7922/G2RF5S92.

[145] Xiang He et al. "Optimal Time-Varying Pricing for Toll Roads Under Multiple Objectives: A Simulation-Based Optimization Approach." In: *Transportation Science* 51.2 (2017), pp. 412–426. ISSN: 00411655.

[146] Xiqun (Michael) Chen, Zheng Zhu, and Lei Zhang. "Simulation-based Optimization of Mixed Road Pricing Policies in a large Real-world Network". In: *Transportation Research Procedia* 8 (2015). Current practices in transport: appraisal methods, policies and models – 42nd European Transport Conference Selected Proceedings, pp. 215–226. ISSN: 2352-1465. DOI: https://doi.org/10.1016/j.trpro.2015.06.056.

[147] Nan Zheng, Guillaume Rérat, and Nikolas Geroliminis. "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment". In: *Transportation Research Part C: Emerging Technologies* 62 (2016), pp. 133–148. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2015.10.015.

[148] Chen Xiqun et al. "Simulation-based pricing optimization for improving network-wide travel time reliability." In: *Transportmetrica A: Transport Science* 14.1 (2018), pp. 155–176.

[149] Xiqun (Michael) Chen et al. "Time-of-day vehicle mileage fees for congestion mitigation and revenue generation: A simulation-based optimization method and its real-world application". In: *Transportation Research Part C: Emerging Technologies* 63 (2016), pp. 71–95. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2015.12.001.

[150] Linsen Chong and Carolina Osorio. "A Simulation-Based Optimization Algorithm for Dynamic Large-Scale Urban Transportation Problems." In: *Transportation Science* 52.3 (2018), pp. 637–656. ISSN: 00411655.

[151] USDOT. "Trends in Statewide Long-Range Transportation Plans: Core and Emerging Topics". In: (2012).

[152] Hong Zheng. "A Primer for Agent-Based Simulation and Modeling in Transportation Applications". In: *Development* (2013), p. 75.

[153] B. Shahriari et al. "Taking the Human Out of the Loop: A Review of Bayesian Optimization". In: *Proceedings of the IEEE* 104.1 (Jan. 2016), pp. 148–175. ISSN: 1558-2256. DOI: 10.1109/JPROC.2015.2494218.

[154] James Bergstra et al. "Hyperopt: a Python library for model selection and hyperparameter optimization". In: *Computational Science & Discovery* 8.1 (July 2015), p. 014008. DOI: 10.1088/1749-4699/8/1/014008.

[155] Jamey M. B. Volker, Amy E. Lee, and Susan Handy. "Induced Vehicle Travel in the Environmental Review Process". In: *Transportation Research Record* 2674.7 (2020), pp. 468–479. DOI: 10.1177/0361198120923365.

[156] Jr. Douglass B. Lee, Lisa A. Klein, and Gregorio Camus. "Induced Traffic and Induced Demand". In: *Transportation Research Record* 1659.1 (1999), pp. 68–75. DOI: 10.3141/1659-09.

[157] Transport for London (TFL). *The Greater London (Central Zone) Congestion Charging Order 2001*. Tech. rep. 2002. URL: https://web.archive.org/web/20080228180825/http://www.tfl.gov.uk/assets/downloads/Report_to_the_MayorcChapters_1-16.pdf.

[158] Jonas Eliasson. *The Stockholm congestion charges: an overview*. Tech. rep. 2014. URL: https://transportportal.se/swopec/cts2014-7.pdf.

[159] Los Angeles County Metropolitan Transportation Authority (LAMetro). *Metro Traffic Reduction Study - Reports to the Metro Board of Directors*. Tech. rep. 2021.

[160] San Francisco County Transportation Authority (SFCTA). *Downtown Congestion Pricing Study*. Tech. rep. 2021.

[161] Adam Stocker Elliot Martin Susan Shaheen. "Impacts of Transportation Network Companies on Vehicle Miles Traveled, Greenhouse Gas Emissions, and Travel Behavior Analysis from the Washington D.C., Los Angeles, and San Francisco Markets". In: (2021). DOI: 10.7922/G2BC3WV9. URL: https://escholarship.org/uc/item/90b6d7r3.

[162] Anne Brown. *Not All Fees are Created Equal: Equity Implications of Ride-hail Fee Structures*. preprint. Open Science Framework, May 28, 2021. DOI: 10.31219/osf.io/cpsqu.

[163] United States Department of Transportation (USDOT) Federal Highway Administration (FHWA). *Transit and Congestion Pricing - A Primer*. Tech. rep. 2009. URL: https://ops.fhwa.dot.gov/publications/fhwahop09015/fhwahop09015.pdf.

[164] Hai Yang and Hai-Jun Huang. "Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes". In: *Transportation Research Part A: Policy and Practice* 33.2 (Feb. 1999), pp. 139–155.

[165] Jonas Eliasson et al. "Accuracy of congestion pricing forecasts". In: *Transportation Research Part A: Policy and Practice* 52 (June 2013), pp. 34–46. DOI: 10.1016/j.tra.2013.04.004.

[166] Yafeng Yin and Siriphong Lawphongpanich. "Internalizing emission externality on road networks. Transp Res, Part D Transp Environ". In: *Transportation Research Part D: Transport and Environment* 11 (July 2006), pp. 292–301. DOI: 10.1016/j.trd.2006.05.003.

[167] Devansh Jalota et al. *When Efficiency meets Equity in Congestion Pricing and Revenue Refunding Schemes*. 2021. DOI: 10.48550/ARXIV.2106.10407.

[168] Zhiyuan Liu et al. "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics". In: *Transportation Research Part C: Emerging Technologies* 79 (2017), pp. 58–72. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2017.03.011.

[169] Qi Luo, Zhiyuan Huang, and Henry Lam. "Dynamic Congestion Pricing for Ridesourcing Traffic: a Simulation Optimization Approach". In: *2019 Winter Simulation Conference (WSC)*. 2019, pp. 2868–2869. DOI: 10.1109/WSC40007.2019.9004722.

[170] US Department of Energy. *Energy Efficient Mobility Systems: 2018 Annual Progress Report*. Tech. rep. US Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office, Dec. 2018.

[171] *Means of Transportation to Work*. 2019. URL: https://www.census.gov/programs-surveys/acs/.

[172] United States Department of Transportation (USDOT) Federal Highway Administration (FHWA). *National Household Travel Survey*. 2017. URL: https://nhts.ornl.gov.

[173] Jessica Lazarus; Gordon Bauer Jeffery Greenblatt Susan Shaheen. "Bridging the Income and Digital Divide with Shared Automated Electric Vehicles". In: (2021). Publisher: Institute of Transportation Studies, Berkeley. DOI: 10.7922/G2707ZQ4.

[174] San Francisco Municipal Transportation Agency (SFMTA). *Travel Decision Survey 2019*. data retrieved from SFMTA, https://www.sfmta.com/reports/travel-decision-survey-2019. 2019.

[175] *Uber Movement Travel Speeds*. Tech. rep. Uber, Inc., 2018. URL: https://movement.uber.com.

[176] Joe Castiglione et al. *TNCs Today: A Profile of San Francisco Transportation Network Company Activity*. Tech. rep. 2017. URL: https://www.sfcta.org/sites/default/files/2019-02/TNCs_Today_112917_0.pdf.

[177] Mark Bradley et al. "Results of the First Large-Scale Survey of Transportation Network Companies Use in the Bay Area". In: *Transportation Research Record* 2676.7 (2022), pp. 13–23. DOI: 10.1177/03611981221076441.

[178] Cheikh Touré et al. "Uncrowded Hypervolume Improvement: COMO-CMA-ES and the Sofomore Framework". In: *Proceedings of the Genetic and Evolutionary Computation Conference*. GECCO '19. Prague, Czech Republic: Association for Computing Machinery, 2019, pp. 638–646. ISBN: 9781450361118. DOI: 10.1145/3321707.3321852.

[179] Nikolaus Hansen. *The CMA Evolution Strategy: A Tutorial*. Tech. rep. 2016. URL: arXiv:1604.00772v1.

[180] Jessica Lazarus et al. "Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete – A case study of San Francisco". In: *Journal of Transport Geography* 84 (2020), p. 102620. ISSN: 0966-6923. DOI: https://doi.org/10.1016/j.jtrangeo.2019.102620.

[181] San Francisco Municipal Transportation Agency (SFMTA). *How People Traveled Through San Francisco in 2021*. 2019. URL: https://www.sfmta.com/blog/how-people-traveled-through-san-francisco-2021-0.

[182] John S. Evans et al. "Exposure efficiency: an idea whose time has come?" In: *Chemosphere* 49.9 (Dec. 2002), pp. 1075–1091. DOI: 10.1016/s0045-6535(02)00242-4.

[183] Transit American Public Transportation Association (APTA). *APTA Ridership Trends*. data retrieved from APTA, https://transitapp.com/APTA, 2022.

Appendix A

BISTRO - General System Specifications

A.1 Introduction

The *Berkeley Integrated System for Transportation Optimization* (BISTRO) framework harnesses advances in the machine learning and data science fields to enable data-driven transportation demand modeling and policy analysis at a resolution and scale that can empower and engage city officials, transportation system managers, the private sector, academics, and citizens to understand, analyze, and collaboratively plan for the rapidly evolving transportation realities shaping urban areas worldwide.

The *BISTRO platform* gives users an opportunity to envision a set of changes to existing transportation systems that bring about the greatest improvements across important indicators of transportation system performance in terms of system-wide *level of service* (LoS), congestion, and sustainability.

This document details the system specifications of BISTRO. First, Section A.2 provides a brief overview of the current state of transportation planning, agent-based simulations and BISTRO. Then, Sections A.4 and A.5 gives a detailed mathematical description and analysis of the model behind BISTRO.

A.2 Background

Related work

As widely available transportation modes are proliferating and changing on a monthly basis or faster (*e.g.*, micromobility, shared mobility, automated mobility, etc.), understanding the behavioral responses of users and the subsequent transportation outcomes merit investigation. Forward-looking studies have begun to look at the possible transportation and

environmental benefits offered by future and emerging shared modes, such as work conducted by the International Transport Forum[1] and the Lawrence Berkeley National Laboratory.[2]

In the U.S. context, many *metropolitan planning organizations* (MPOs) are equipped to analyze the long-range transportation effects of new modes in their regions, but just as many are not. The purpose of *BISTRO* is to provide a platform on which users can envision near-term implementable operational changes to a region's transportation system. These changes are intended to complement existing operational and infrastructure expenditure processes. Additionally, the outputs of the platform can provide a basis for discussions across many sets of stakeholders.

Agent-Based Simulation

Agent-based simulation (ABS) of travel demand, is a method by which to evaluate the network-wide effects of modifications to a transportation system. Agent-based travel demand microsimulation realizes the daily activity schedules and transportation choices of a socio-demographically heterogeneous population of citizens on a virtual representation of physical road networks.[3] This methodology enables an informative resolution of feedback loops and spatio-temporal constraints operating between travel purposes, road network congestion, household vehicle availability, and the levels of service provided by infrastructure and available transportation modes.

Person agents represent simulated individuals who make decisions about what transportation mode(s) to use to travel to and from their daily activities. During the simulation, person agents make one or more *tours* of travel to sequential activities, starting and ending each tour at home. Each *trip* in a tour represents travel from one activity to the next. Trips may consist of one or more *legs* of travel, each using a particular *mode* of transportation. A *mode choice model* characterizes the transportation mode preferences of agents by accounting for the sensitivity of the agent to the attributes of each alternative, such as wait time, in-vehicle travel time, and trip cost.[4] The simulator uses a realistic representation of the transportation network and a *routing algorithm* to determine the generalized cost of routing vehicles on the network as a function of the expected travel-time on links taking into account congestion (*i.e.*, movement slower than the maximum allowable speed due to the number of vehicles on a link exceeding capacity).

The inputs to one instance of the simulation include a representative population of synthetic agents together with their typical daily activity plans. A virtual road network, transit schedule, and parking infrastructure define the transportation supply. The simulation pro-

[1]For more information, see: https://www.itf-oecd.org/sites/default/files/docs/shared-mobility-liveable-cities.pdf

[2]For more information, see https://www.nature.com/articles/nclimate2685

[3]While the population and its plans are synthetic, econometric modeling techniques using census data and travel surveys together with calibration against observed mode splits and network volumes ensure that the simulation represents typical daily traffic conditions.

[4]For more information about mode choice models, see: https://eml.berkeley.edu/books/choice2.html

ceeds iteratively: evaluating the plans on the physical network and then permitting agents to replan components of their schedule in response to the generalized costs of travel. Once agents have settled on a set of plans that collectively maximize the average utility of their set of evaluated plans, the simulation engine reaches a fixed point or equilibrium condition.

Each simulation run produces outputs of the actual paths and travel times realized by each person agent and each vehicle, as well as a host of other data, further detailed in Section A.4. In practice, outputs of agent-based simulations may be used to communicate policy alternatives to stakeholders. For example, visualizations of congested roadways with millions of agents behaving independently can provide a concise method to communicate the effects of infrastructure interventions.

Agent-based simulation allows for the evaluation of counterfactual *scenarios*. A scenario is a simulation that implements a unique set of circumstances that differs in some way from a *base case*. The base case is calibrated using data representing the current state of the transportation system being simulated. Examples of scenarios include alteration of the population configuration representing population or employment growth, alteration of the transportation network such as unexpected road network restrictions due to sporting events, inclement weather or traffic accidents, as well as the introduction of new modes of transportation such as autonomous vehicles. Well-calibrated simulations of transportation systems, such as those just described, allow stakeholders to better understand the implications of policy proposals in hypothetical travel environments.

BISTRO

The *Berkeley Integrated System for TRansportation Optimization* (BISTRO) is an open-source *Collaborative Planning Support System* (CPSS) designed to assist stakeholders in addressing the increasingly complex problems arising in transportation systems worldwide. BISTRO includes an agent-based modeling and simulation (ABMS) framework—namely the *Behavior Energy Autonomy and Mobility* (BEAM) framework[5] developed at Lawrence Berkeley National Lab (LBNL)—and scenario development pipeline to build empirically-validated simulations of multimodal metropolitan transportation systems and algorithmically optimize system interventions that best align with policy and planning objectives. BISTRO was developed with the intent to leverage the distinct backgrounds of planners and computer scientists to facilitate a process to draw upon the strengths of two complementary areas of expertise to inform rather than direct public conversations about proposed transportation policy, investments, and regulations. In other words, it is the intent of the BISTRO developers to offer it as a tool to add complementary value to already existing transportation planning processes.

For an in-depth description of the BISTRO framework and all of its major components illustrated in Figure A.1, refer to Chapter 4.

[5]For more information about BEAM, see: http://beam.lbl.gov/

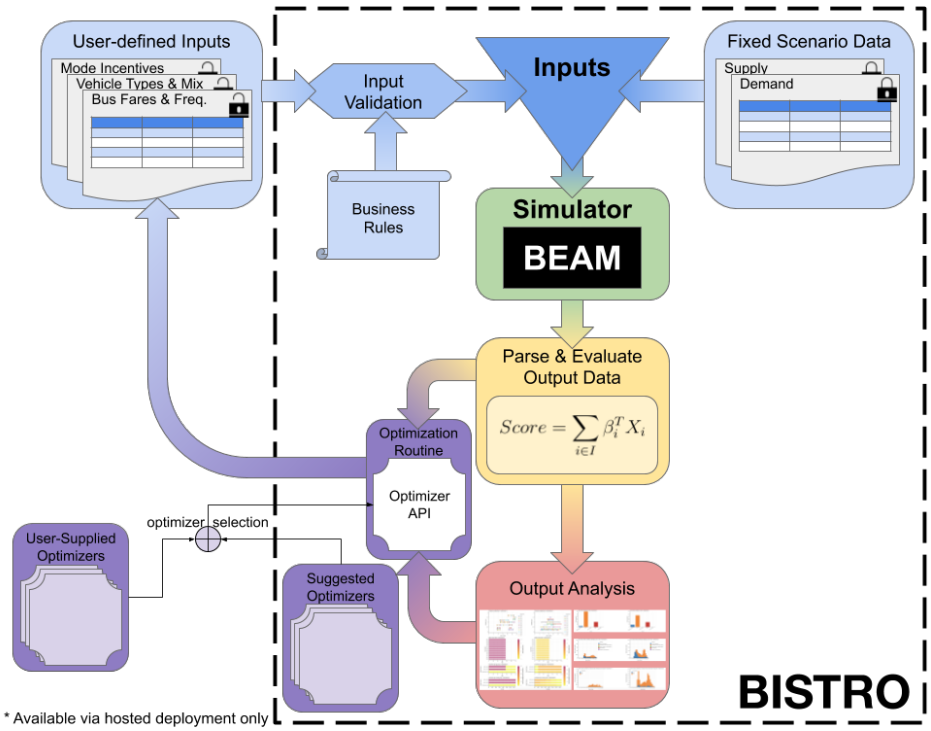


Figure A.1: BISTRO framework flow diagram, outlining processes, user defined inputs, outputs and optimization.

A.3 Study Configuration

BISTRO enables the study of optimal interventions to a given transportation system that is modeled within the BISTRO run environment. A BISTRO run environment is configured using a set of fixed input data that define the required transportation system supply elements (*e.g.*, road network, transit schedule, on-demand ride fleet) and demand elements (*e.g.*, synthetic population, activity plans, and mode choice function parameters). Precisely which aspects of the virtual transportation system should be represented in the simulation model depends on the strategic goals and system objectives defined as part of the planning and analysis process motivating a particular BISTRO use case.

Each simulation run takes as input a set of configuration and user-defined input variables. The *configuration variables* (also called *fixed scenario data* in Figure A.1) define the geographic and physical constraints of the transportation network, the characteristics of vehicles in the transit, private, and on-demand ride fleets, and the instantiation of each agent in the population, including their socio-demographic characteristics and activity plans. The configuration variables are not intended to be altered in any way during optimization, but may be altered through the scenario development pipeline to produce different scenarios across which the system interventions may be optimized.

The *user-defined inputs* (UDIs) represent the investment(*e.g.*, transit fleet mix modification, bus route modifications, parking supply, electric vehicle charge station locations, dynamic redistribution of e-bikes or on-demand vehicles), incentive (*e.g.*, incentives to specific socio-demographic groups for selected transportation modes, road pricing/toll roads, fuel tax), or policy/operational (*e.g.*, transit schedule adjustment, transit fare modification, parking pricing levers applicable to and available for the study at hand). The project owner may constrain the range of possible values upon which each UDI is valid by setting the corresponding input validation parameters and business rules.

In the following section, the specific scenario of Sioux Faux,

Scenario Configuration Inputs

System interventions may be optimized using one or more distinct representations of the supply and demand within the transportation system, each of which is called a *scenario*. The BISTRO scenario development pipeline enables project owners to generate multiple scenarios, each of which is defined by a unique set of configuration inputs that determine the physical transportation network and the dynamics of vehicles and person agents as well as the spatio-temporal and socio-demographic distribution and sensitivity of demand across the network.

Thus, the scenario configuration inputs include the following variables:

1. **Transportation Network Configuration**: inputs that specify the geography and physics of the transportation system as well as the operation of each transportation mode in the system.

    a) **Transportation network(s)**: one or more directed graph composed of nodes and road links that define the access to and physics of vehicles that move throughout the transportation system. For example there may be separate networks for each of the following modes: pedestrians, cyclists, road transport, and rail.

    b) **Facilities**: the locations of facilities in the network, each of which is located at a network node. Facilities include residences, places of work, transit stations, and additional locations at which agent activities may occur.

    c) **Transportation modes**: each transportation mode is configured by inputs that define the number and type of vehicles in the fleet (if applicable), the method of distributing those vehicles at the start of a simulation run, the operational costs associated with the operation of the mode, and the fare, if any, charged for use of the mode. Transit mode configuration inputs also define the facilities and paths used for each transit route.

2. **Population Configuration**: inputs that specify the distribution and characteristics of the population, their activity plans, and modal preferences.

a) **Population synthesis**: the spatial distribution of households and the person agents that are in them, including the spatial distribution of socio-demographic variables. Concretely, the population synthesis inputs define the number and size of households to be randomly assigned to homes as well as the distribution of household- and individual-level socio-demographics in within each geographic zone.

b) **Daily activity schedules**: the spatio-temporal distribution of activity plans. For example, a scenario thta investigates a special event may be implemented in the scenario configuration by altering the activity plans of a portion of the population. Other examples include altering the scenario configuration to mimic changes in the employment rate or the telecommute mode share.

c) **Mode choice**: the mode choices of person gents are determined by a multinomial logit model, the coefficients of which can be altered by scenario configuration inputs. Doing so would allow for a sensitivity analysis across scenarios of varying demand sensitivities.

User-Defined Inputs

A boundary separates external, exogenously defined inputs from the BISTRO simulation optimization pipeline. Outside of the boundary, the *user-defined inputs* (UDIs) represent the investment, incentive, and policy levers applicable to and available for the study at hand. Concretely, algorithm developers encode solutions as numeric values that represent vector-valued variables controlling aspects of the initialization and evolution of the simulation. For example, a UDI that alters frequency of buses on a route must specify a target transit agency, a route, a start time, an end time, and the desired headway.

BISTRO provides a library of possible inputs for scenario designers to adapt to specific use cases. The selection of UDIs is intended to be compatible with the system objective. UDIs may represent, for example, the investment (*e.g.*, transit fleet mix modification, bus route modifications, parking supply, electric vehicle charge station locations, dynamic re-distribution of e-bikes or on-demand vehicles), incentive (*e.g.*, incentives to specific socio-demographic groups for selected transportation modes, road pricing/toll roads, fuel tax), or policy/operational (*e.g.*, transit schedule adjustment, transit fare modification, parking pricing) levers applicable to the study at hand. The project owner may constrain the range of possible values upon which each UDI is valid by setting the corresponding input validation parameters and business rules. The example input file for bus scheduling shown in Table A.1 defines alteration of the headway of a particular bus route during a particular service period (defined by its start and end times). For discussion of the initially released BISTRO UDIs, refer to the BISTRO paper. An inventory and up-to-date descriptions of the UDIs may be found on the BISTRO website[6].

[6]https://sfwatergit.github.io/BISTRO-Website/

Table A.1: Example of bus frequency adjustment input file.

route_id	start_time	end_time	headway_secs	exact_times
1340	21600	79200	900	1
1341	21600	36000	300	1
1341	61200	72000	300	1

Input Validation and Business Rules

While BISTRO maintains a library of available interventions compatible with BEAM, scenario designers, policy makers, and other stakeholders will often want assurance that infeasible, regressive, or otherwise undesirable input combinations are prevented from being selected as "optimal." Together with syntactic and schematic validation of inputs, flexibly-defined *business rules* can effectively act as constraints on the search space—enhancing the interpretability and, thereby, the rhetorical and communicative value of BISTRO-derived solutions.

A.4 Evaluation and Scoring

The quality of a policy tested with BISTRO is judged based on a weighted combination of measurable outcomes from the simulation that emulate common operational and social goals considered by cities when evaluating the broader impacts of transportation policy and investment. The scoring components are derived from a discrete set of output variables produced for each simulation run.

The *outputs* of the simulation produced by users' solutions will determine the values of key performance metrics of the impact of the solutions on the accessibility, LoS, and congestion of the transportation network in the city of interest, as well as the resource constraints and environmental sustainability of the resulting network-wide travel equilibrium.

The following subsections detail the relevant person agent and vehicle movement outputs in the simulation as well as the scoring criteria used for evaluation.

Person output

For each trip taken by a person agent in the simulation, the following data is produced as output:

1. **Transportation mode(s):** for each trip, a person agent chooses one of the modes available to them. Person agents may use one or more modes to travel from their origin to their desired destination, as they may transfer between modes along the way.

a) **Mode(s) available:** the mode(s) available to person agents to use for each trip, including: walk, personal bicycle, personal car, on-demand ride, bus, rail, shared bicycle, shared scooter.

b) **Mode choice:** the mode chosen for each trip. Mode choices are made at the trip level, as the agent considers all available combinations of modes that may be used to travel from one activity to another. The use of a transit mode (bus or rail) involves the combination of that mode with one or more other modes used to access and egress the transit station. For example, a person agent may choose to drive to transit, which will result in a walking egress leg from the transit station to the agent's destination.

2. **Travel time:** the time spent by the person agent in the act of traveling during each leg of a trip. Travel time has several components, including:

   a) **In-vehicle travel time:** the time spent in a vehicle by an agent while traveling to an activity.

   b) **Wait time:** the time spent by an agent waiting for the arrival of a vehicle. Wait time may include time spent at a bus stop or time spent waiting for the arrival of an on-demand ride vehicle after the ride has been reserved.

   c) **Transfer time:** time spent walking from one transportation mode to another while completing a trip. Transfer time may include walking from the bus stop of one bus route to a bus stop of another bus route.

3. **Travel expenditure:** the cost incurred by a person agent during a trip. The net cost of travel incurred may include:

   a) Transit fares

   b) On-demand ride fares

   c) Gas consumption by a personal vehicle

   d) Tolls paid (if applicable)

   e) Applicable incentives

4. **Incentives:** the amount of monetary incentive available (based on the modes available) for a trip and the amount of incentive consumed by an agent during a trip. The amount of incentive consumed for a particular trip may not be more than the travel expenditure incurred for the trip.

5. **Trip purpose:** the nature of the primary activity to which a person agent is traveling during a trip. Trip purpose is segmented into two or more mutually exclusive categories. In the most simple case, there are two trip purpose categories: work trips and secondary trips. Additional trip purpose categories may be included in accordance with the availability of such categories in the activity plans used for the study at hand.

Vehicle output

Every vehicle movement during the simulation produces the following outputs:

1. **Origin-Destination-Time (ODT) record:** the origin location, destination location, time of departure, and time of arrival of a vehicle movement.
2. **Path:** an ordered list of the links traversed on the path from the origin to the destination of a vehicle movement.
3. **Fuel consumption:** the amount of fuel consumed by a vehicle during a movement.
4. **Vehicle occupancy:** the number of passengers in a vehicle during a movement.

Scoring criteria

Transportation system intervention alternatives are scored in BISTRO based on a function of score components evaluated using *key performance indicators* (KPIs) of the simulation. KPIs emulate common operational, environmental, and social goals considered by transportation planners and policymakers when evaluating the broader impacts of transportation policy and investment. BISTRO project planners may select KPIs to include in the scoring function from an existing library of options, or may choose to develop additional KPIs, as appropriate, for the goals and system objectives of the project. Additionally, the form of the scoring function may be designed by the analyst in consultation with the project planner.

These KPIs are all affected by the user-defined input variables, and care must be taken when optimizing to understand the interactions between metrics. There are two types of KPIs:

1. Those that measure the operational efficiency of the transportation system (*e.g.*, *vehicle miles traveled* [VMT], vehicle delay, operational costs, revenues)
2. Those that evaluate the experience of transportation system users (*e.g.*, generalized travel expenditure, bus crowding experienced, accessibility)

Both types of metrics are scored simultaneously by comparing the user-produced results to the *business-as-usual* (BAU) scenario.

This section describes the six categories of KPIs that have been developed and implemented in BISTRO at the time of publication of this document.

Accessibility

In an urban transportation planning setting, *accessibility* has often been defined as a measure of the ease and feasibility with which opportunities or points of interest can be reached via available modes of travel. Although there are many ways to measure accessibility, it is quantified herein as the average number of points of interest (of a specific type of activity)

reachable within a given duration of time. More specifically, accessibility is measured as the sum of the average number of points of interest reachable from network nodes by car or using public transit, within a specified amount of time during specific time periods. Examples of the how the accessibility metric may be disaggregated by trip purpose, mode, and time periods are shown include:

1. **Accessibility to work-based trips by car**: The sum of the average number of work locations accessible from each node by car within 15 minutes during the AM peak (e.g., 7–10AM).
2. **Accessibility to work-based trips by public transit**: The sum of the average number of work locations accessible from each node using public transit within 15 minutes during the AM peak (e.g., 7–10AM).
3. **Accessibility to secondary trips by car**: The sum of the average number of secondary locations accessible from each node by car within 15 minutes during the AM peak (e.g., 7–10AM) and midday period (e.g., 10AM–5PM) periods.
4. **Accessibility to secondary trips by public transit**: The sum of the average number of secondary locations accessible from each node using public transit within 15 minutes during the AM peak (e.g., 7–10AM) and midday period (e.g., 10AM–5PM) periods.

Figure A.2 displays the accessibility for work-based and other trips, by car and by public transit, in a benchmark BISTRO scenario.

Equity

The socio-demographic and spatial heterogeneity of travel behavior within BISTRO enables a variety of equity-focused impact analyses. One such metric, which is applicable in scenarios for which there is limited intuition about both the composition of the population demographics and the spatial distribution of resources (*e.g.*, transit access, car ownership, etc.), is presented as such:

- **Average generalized transportation cost burden**: the generalized transportation cost for a particular trip is computed as the sum of the travel expenditures of the trip (costs of fuel, fares minus incentives, as applicable) and the duration of the trip multiplied by the average value of time of the population. The cost burden is computed by dividing the generalized cost burden by the household income of the agent completing the trip. The average generalized transportation cost burden is thus computed for all work trips and all secondary trips, separately.

Although this is an aggregate measure and does not examine the changes in outcomes for specific population groups, the means to pay of each household is accounted for.

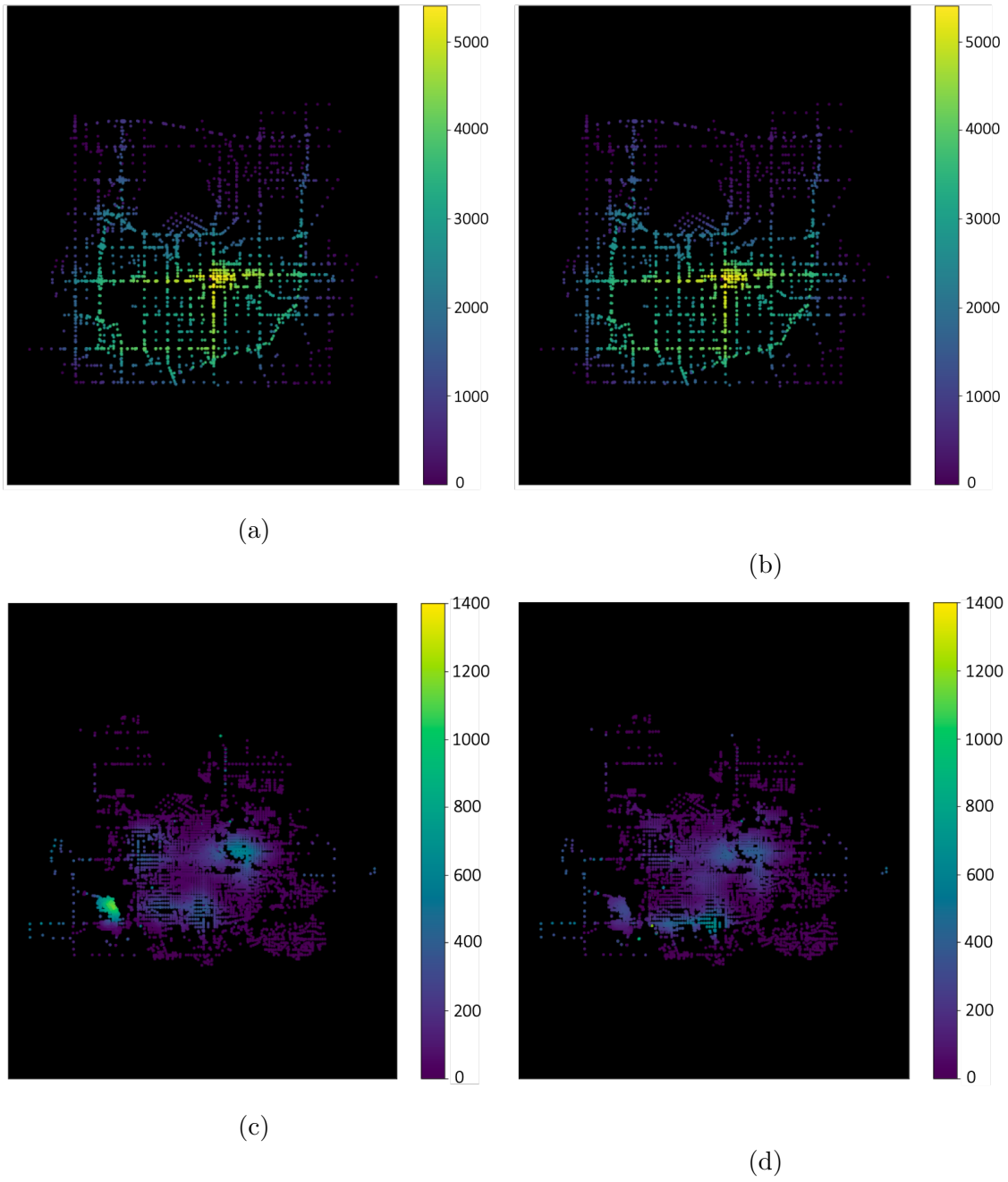


Figure A.2: Points of interest accessible within 15 minutes in the *Business as Usual* (BAU) of a BISTRO benchmark scenario: (a) accessibility to work locations by car, (b) accessibility to secondary locations by car, (c) accessibility to work locations by public transit, (d) accessibility to secondary locations by public transit.

Level of Service

The *level of service* (LoS) experienced by public transit passengers has a direct influence on short- and long-term demand for public transit service. In the short-term, passenger demand for a particular transit line is dependent on the time and cost of alternative travel options. Thus, the frequency and service period of transit service determines the availability, wait times, transfer times, and in-vehicle times of a prospective transit trip and thus the utility of that trip in comparison to the same trip completed with alternative transportation modes. In addition, the available capacity on a transit vehicle affects whether or not the passenger can board transit at the desired time. Furthermore, the comfort afforded by the available space on a transit vehicle has long-term effects on transit demand, as passengers internalize their experience during many transit trips over time and develop an additional aversion or affinity to transit based on their expectation of the LoS. Upon experiencing the discomfort of an overcrowded transit vehicle for the same 'trip' (*e.g.*, a traveler's 8AM home–work morning commute trip), a traveler will come to expect that LoS when considering whether to take transit for that trip in the future. Though the LoS may be measured in BISTRO by any one of the factors mentioned, here is an example of LoS in terms of passenger comfort:

1. **Average bus crowding experienced**: computed as the average over all transit legs of the total passenger-hours weighted by the value of time multipliers corresponding to the load factor (the ratio of total passengers to the seating capacity) of the bus during the leg.

Congestion

Congestion is measured in two primary ways: the total sum of miles traveled by all modes on the network, and by the delay incurred. Delay—calculated as the difference between actual and free flow time—is presented both as a sum across all vehicle movements in the simulation and as an average delay per agent trip. Using these three measures of congestion provides insight into the destination- or opportunity-independent level of mobility on a network, the overall network performance, and efficiency.

1. **Total Vehicle Miles Traveled (VMT)**: total miles traveled by all motorized vehicles of the system during the simulation.

2. **Average Vehicle Delay per Person Agent Trip**: the average across agent trips of vehicle hours of delay experienced by all vehicles while occupied by one or more passengers during the simulation.

Financial Sustainability

Financial Sustainability is considered on aggregate, as the sum of the operational costs of bus service and the total incentives used subtracted by the total bus revenue collected.

Table A.2: Default Vehicle Types currently used in simulations. Note: other vehicle types can be chosen by users when configuring a new scenario.

Vehicle type, $c \in C$	**Fuel type**	**Fuel consumption rate ($Joule/meter$)**	**Operational cost ($/hour$)**	**Seating capacity**	**Standing capacity**
CAR	gasoline	3655.98	n/a	4	0
BUS-DEFAULT	diesel	20048	89.88	37	20
BUS-SMALL-HD	diesel	18043.2	90.18	27	10
BUS-STD-HD	diesel	20048	90.18	35	20
BUS-STD-ART	diesel	26663.84	97.26	54	25

1. **Operational Costs**: total costs incurred by the public transit agency operations including the cost of fuel consumed, and hourly variable costs. Hourly variable costs include estimated labor, maintenance and operational costs. An example of the rates for each of these factors is specified in the vehicle fleet configuration variables (see Table A.2).

Table A.3: Fuel Types currently used in simulations. Note: other fuel types can be chosen by users when configuring a new scenario.

Fuel type	**Fuel cost ($/MJoule$)**
Gasoline	0.03
Diesel	0.02
Electricity	0.01
Biodiesel	0.01

2. **Incentives Used**: total incentives used by agents.

3. **Revenue**: sum of total bus fares collected.

Environmental Sustainability

Sustainability metrics provide a sense of the local externalities resulting from any transportation interventions.

1. **Total Particulate ($PM_{2.5}$) Emissions**: total $PM_{2.5}$ emissions produced by all motorized vehicles during the simulation. Using criteria pollutants, specifically particulate matter running exhaust emission factors, presents a mileage-based measure of local

air quality impacts based upon vehicle type. This metric provides a complementary addition to fuel/energy consumption-based metrics, which are captured elsewhere[7].

2. **Total Greenhouse Gas (GHG) Emissions**: total well-to-wheels[8] emissions produced by all vehicles. The well-to-wheels emissions are calculated using estimates of fuel consumed calculated as a function of vehicle speeds. For more information on the methodology followed to estimate fuel consumption, please refer to the BEAM documentation `https://beam.readthedocs.io/en/latest/index.html`.

A.5 Variable and Scoring Function Specification

The scoring criteria are defined explicitly as functions of the input and output variables of a simulation run. For ease of understanding, the variable notation reflects the meaning of each category of variable in the following manner: all network and vehicle configuration input variables are denoted with a Z, all population configuration input variables are denoted with a N, all user-defined input variables are denoted with a D, all person agent-level output variables are denoted with an X, and all vehicle movement output variables are denoted with a Y. The indices identifying the meaning of each variable will be defined in the following sections in addition to the corresponding units of measurement.

Input Variable Specification

Transportation network and fleet configuration

Network and fleet characteristic variables are defined during the configuration of a simulation, and remain static for each scenario. Tables A.4 and A.5 provide a summary of the network and fleet configuration input variables, respectively.

The road network is a directed graph, $\mathcal{G}(L, W)$, comprised of W nodes connected by L links. Road network configuration variables are indexed by the link identifier $l \in L$. A link instance is defined by the Boolean variable, $Z^l_{i,j}$, which indicates the origin and destination nodes connected by link l.

$$Z^l_{i,j} = \begin{cases} 1 \text{ if link } l \text{ connects node } i \text{ to node } j \\ 0 \text{ otherwise} \end{cases}$$

[7]For more information on the methodology followed to develop this metric, please refer to the California Air Resources Board documentation: `https://www.arb.ca.gov/cc/capandtrade/auctionproceeds/cci_emissionfactordatabase_documentation.pdf?_ga=2.94247453.1690201828.1547860553-1364631033.1545190476`

[8]Well-to-wheels analyses account for total life-cycle emissions, accounting for fuel or energy production, transport, and eventual consumption/burning in vehicles, among other aspects. More information can be found here: `https://ec.europa.eu/jrc/en/jec/activities/wtw`.

Table A.4: Transportation Network Configuration Input Notation

Transportation Network		
Notation	Description	Specification
L	All links in the network	$l \in L$
W	All nodes in the network	$w \in W$
F	All facilities in the network	$f \in F \subset W$
M	All transport modes in the network	$m \in M$
$Z^l_{i,j}$	Link instance between nodes $i, j \in W$	$Z^l_{i,j} \in \{0, 1\}$
Z^l_{length}	Length of link $l \in L$ (m)	$Z^l_{length} \in \mathbb{R}_+$
$Z^l_{capacity}$	Capacity of link $l \in L$ (vehicles/hour)	$Z^l_{capacity} \in \mathbb{R}_+$
$Z^l_{free-flow}$	Free-flow speed of link $l \in L$ (m/sec)	$Z^l_{free-flow} \in \mathbb{R}_+$
Z^K_w	Facility instance at node $w \in W$	$Z^K_w \in \{0, 1\}$
Public Transit Network		
Notation	Description	Specification
T	All transit agencies in the network	$t \in T$
R^t_{routes}	All routes operated by agency $t \in T$	$r \in R^t_{routes}$
$K_{transit}$	All transit stations in the network	$f \in K_{transit} \subset F$
$n^{t,r}_{stops}$	Number of stops on route $r \in R^t_{routes}$ of agency $t \in T$	$n^{t,r}_{stops} \in \mathbb{R},\ n^{t,r}_{stops} > 0$
$Z^{t,r}_{stops}$	Ordered list of stops on route $r \in R^t_{routes}$ of agency $t \in T$	$Z^{t,r}_{stops} \subset K_{transit}$
$n^{t,r}_{path}$	Number of links in the path for route $r \in R^t_{routes}$ of agency $t \in T$	$n^{t,r}_{path} \in \mathbb{R},\ n^{t,r}_{path} > 0$
$Z^{t,r}_{path}$	Ordered list of links on route $r \in R^t_{routes}$ of agency $t \in T$	$Z^{t,r}_{path} \subset K_{transit}$
Z^t_{fare}	The set of fares per ride with agency t charged to riders of each age	$Z^t_{fare} \in \mathbb{R}^A_+,\ Z^t_{fare,a} \in \mathbb{R}_+$
$Z^{t,r}_{headway}$	The headway on route $t \in R^t_{routes}$ (secs)	$Z^{t,r}_{headway} \in \mathbb{R}_+$
$n^{t,r}_{trips}$	Number of trips on route $r \in R^t_{routes}$ of agency $t \in T$	$n^{t,r}_{trips} \in \mathbb{R},\ n^{t,r}_{trips} \geq 0$
$Z^{t,r}_{trips}$	Set of all trips made on route $r \in R^t_{routes}$	$Z^{t,r}_{trips} \in \mathbb{R}^{n^{t,r}_{trips}}$
Z^v_{start}	Service start time of vehicle $v \in Z_{transit}$	$Z^v_{start} \in \mathbb{R},\ Z^v_{start} \geq 0$
Z^v_{end}	End time of vehicle $v \in Z_{transit}$	$Z^v_{end} \in \mathbb{R},\ Z^v_{end} \geq 0$
On-Demand Ride Services		
Notation	Description	Specification
S	All types of on-demand ride services	$s \in S := \{ride\ alone, pooled\}$
$Z^s_{fare,base}$	Base fare for service $s \in S$ (\$ / ride)	$Z^s_{fare,base} \in \mathbb{R}, Z^s_{fare,base} \geq 0$
$Z^s_{fare,dist}$	Distance-based fare for service $s \in S$ (\$ / m)	$Z^s_{fare,dist} \in \mathbb{R}, Z^s_{fare,dist} \geq 0$
$Z^s_{fare,dur}$	Duration-based fare for service $s \in S$ (\$ / min)	$Z^s_{fare,dur} \in \mathbb{R}, Z^s_{fare,dur} \geq 0$

Each link is given a length (in miles), Z^l_{length}, capacity (in vehicles per hour), $Z^l_{capacity}$, and free-flow speed (in miles per hour), $Z^l_{speed-limit}$. These variables are input to the routing operator during the simulation to determine the travel time of vehicle movements as a function of the number of vehicles on a link at a particular time.

Every point of interest in a BISTRO simulation (e.g., residences, work locations, transit stations, etc.) is located at one and only one node in the network such that the set of all facilities, F is a subset of the set of nodes, W. The Boolean variable Z^K_w denotes that facility $f \in F$ is located at node $w \in W$. The sets of all points of interest of a particular type, such as the set of all transit stations $K_{transit}$ or the set of all residences $K_{residence}$, are each a subset of the set of all facilities, F.

$$Z^f_w = \begin{cases} 1 \text{ if facility } f \text{ is located at node } w \\ 0 \text{ otherwise} \end{cases}$$

The transit network configuration follows a *General Transit Feed Specification* (GTFS) format. Each transit agency, $t \in T$ in a BISTRO simulation environment has one or more transit routes. Thus, the transit network is a subgraph of the road network comprised of t set of all transit facilities $K_{transit} \subset W$ and a set of routes, R^t_{routes}, operated by each agency, $t \in T$. Each transit route is defined by an ordered list of $n^{t,r}_{stops}$ transit stations at which the route stops, $Z^{t,r}_{stops} = \{K_{t,r,i}\}_{i=1,\ldots,n^{t,r}_{stops}}$, and an ordered list of $n^{t,r}_{path}$ links along which the route travels, $Z^{t,r}_{path} = \{l_i\}_{i=1,\ldots n^{t,r}_{path}}$.

In the default BISTRO transit network configuration, the fare structure for transit is age-based and identical across all routes in an agency. The fares for each transit agency t are defined by the set $Z^t_{fare} = \{Z^t_{fare,a}\}_{a=0,1,\ldots,max\ age}$ such that $Z^t_{fare,a}$ denotes the fare for passengers of age $a \in \{0, 1, \ldots, max\ age\}$.

There are two available types of on-demand ride services, S. The *ride alone* service type allows passengers to reserve a ride on-demand directly from their origin to destination for themselves (and companion(s)). The *pool* service type allows passengers to reserve a ride on-demand from their origin to destination for themselves and up to one companion in which the ride may be shared with other passengers traveling along a similar route. A pooled on-demand ride may deviate from the shortest path for any particular passenger in order to pick up or drop off another passenger. Each service type $s \in S$ is instantiated with fare parameters, including the base fare $Z^{on-demand,s}_{fare,base}$, a distance-based fare $Z^{on-demand,s}_{fare,distance}$, and a duration-based fare $Z^{on-demand,s}_{fare,duration}$. The total fare for on-demand rides is the sum of the base fare, the product of the total ride distance and the distance based fare, and the product of the total ride duration and the duration-based fare (see section A.5).

The set of all vehicles, V, is comprised of the union of the set of personally owned vehicles, $V_{personal}$, the set of buses, V_{bus}, and the set of on-demand vehicles, $V_{on-demand}$. The number and type of vehicles in the private and on-demand fleets is defined during the configuration of the simulation and remains independent of user-defined input. However, the number of

Table A.5: Vehicle Fleet Configuration Input Notation

Vehicle Fleet Configuration		
Notation	Description	Specification
V	All vehicles in the transportation system	$v \in V := V_{transit} \cup V_{on-demand} \cup V_{private}$
$V_{transit}$	All transit vehicles	$v \in V_{transit} \subset V$
$V_{fleet}^{t,r}$	All transit vehicles in the fleet servicing route $r \in R_{routes}^{t}$ of transit agency $t \in T$	$v \in V_{fleet}^{t,r} \subset V_{transit}$
$V_{private}$	All privately owned vehicles	$v \in V_{private} \subset V$
$Z_{hhd,h}^{v}$	Vehicle ownership instance of vehicle $v \in V_{private}$ owned by household $h \in H$	$Z_{hhd,h}^{v} \in \{0,1\}$
C	All possible vehicle types in the transportation system	$c \in C$
C_m	All possible vehicle types in the vehicle fleet for mode $m \in M$	$c \in C_m \subset C$
$Z_{seating}^{c}$	The passenger seating capacity of vehicle type $c \in C$	$Z_{seating}^{c} \in \mathbb{R},\ Z_{seating}^{c} > 0$
$Z_{standing}^{c}$	The passenger standing capacity of vehicle type $c \in C$	$Z_{standing}^{c} \in \mathbb{R},\ Z_{standing}^{c} \geq 0$
$Z_{fuel-type}^{c}$	Fuel type used by vehicle type $c \in C$	$Z_{fuel-type}^{c} \in \{gas, diesel, electricity\}$
$Z_{fuel-cons}^{c}$	The average fuel consumption rate of vehicle type $c \in C$ (mJ/m)	$Z_{fuel-cons}^{c} \in \mathbb{R},\ Z_{fuel-type}^{c} > 0$
$Z_{var-op-cost}^{c}$	The variable operational cost of vehicle type $c \in C$ (\$/hour)	$Z_{var-op-cost}^{c} \in \mathbb{R},\ Z_{var-op-cost}^{c} \geq 0$
$Z_{fuel-cost}^{f}$	The fuel cost of fuel type $f \in \{gas, diesel, electricity\}$ (\$ /mJ)	$Z_{fuel-cost}^{f} \in \mathbb{R},\ Z_{fuel-cost}^{c} \geq 0$
$Z_{veh-type,c}^{v}$	Vehicle type assignment of vehicle $v \in V$	$Z_{veh-type,c}^{v} \in \{0,1\}$

vehicles in the bus fleet is dependent on the service periods and headways (i.e., frequency) for each route, which can be altered by user-defined inputs (UDIs).

The schedules for each transit route are typically defined by GTFS input, which provides the headway, $Z^{t,r}_{headway}$, and the set of all trips, $Z^{t,r}_{trips} \in \mathbb{R}^{n^{t,r}_{trips}}$, to be completed for each route. Each trip on a particular transit route is assigned to a unique vehicle id. Thus, the set of all buses to be used on route $r \in R^t_{routes}$ is denoted as $V^{t,r}_{fleet} \in \mathbb{R}^{n^{t,r}_{trips}}$, and the set of all transit vehicles in the base case as $V_{transit} = \cup_{t \in T} \cup_{r \in R^t_{routes}} V^{t,r}_{fleet}$. Each transit vehicle, $v \in V_{transit}$, is given a start and end time (in seconds from the start of the simulation) for servicing the corresponding route, r, Z^v_{start} and Z^v_{end}, respectively. Thus, the configured service period for each route, t, is the tuple (Z^t_{start}, Z^t_{end}) where Z^t_{start} corresponds to the start time of the vehicle making the first trip on the route and Z^t_{end} corresponds to the end time of the vehicle making the last trip on the route.

The type of vehicles in certain fleets may also be altered by UDIs. Across all vehicles, the set of possible vehicle types is the set C. In a BISTRO study that includes a user-defined input (UDI) that alters the fleet mix of one or modes, it is necessary to define the subset of vehicle types that are eligible to be included in the fleet mix for each mode. Some vehicle fleets may include only a subset of all vehicle types in a particular simulation run. For example, the vehicle types available for the transit fleet, $C_{transit}$ will typically be mutually exclusive from the set of vehicle types available for the private fleet, $C_{private}$.

Each vehicle type, $c \in C$, is defined by a seating and standing capacity (in number of person agents), $Z^c_{seating}$ and $Z^c_{standing}$, a fuel type, $Z^c_{fuel-type}$, average fuel consumption rate (in units of fuel consumed per meter), $Z^c_{fuel-consumption}$, and a variable operational cost (in dollars per hour traveled), $Z^c_{var-op-cost}$. For each fuel type, $Z^K_{fuel-cost}$ defines the cost of fuel consumption (in dollars per mJ of fuel consumed).

The default set of fuel types available in BISTRO includes gasoline, diesel, and electricity. While the average fuel consumption rate for each vehicle type is provided, the actual fuel consumed by each vehicle movement is calculated in BEAM based on the speed profile on each link traversed. Therefore, the actual fuel consumed by a vehicle in a simulation run may not be equal to the product of the average fuel consumption rate and the total miles driven by that vehicle.

At the start of a simulation run, each vehicle is instantiated with an identifier, $v \in V$, and a Boolean identifier, $Z^v_{veh-type,c}$, that denotes the vehicle type of the vehicle such that

$$Z^v_{veh-type,c} = \begin{cases} 1 \text{ if vehicle } v \text{ is of vehicle type } c \\ 0 \text{ otherwise} \end{cases}$$

Additionally, all vehicles in the private vehicle fleet are identified by the household to which they belong using the Boolean identifier $Z^v_{hhd,h}$, defined in the following subsection.

Table A.6: Population Configuration Input Notation

Population Configuration		
Notation	Description	Specification
H	All households in a simulation scenario	$h \in H$
N	All person agents in a simulation scenario	$n \in N$
A	Maximum age for person agents in a simulation scenario	$A \in [0, 120]$
I	Maximum income for person agents in a simulation scenario	$I \in \mathbb{R},\ I < \infty$
P	Maximum number of activities completed by each agent in a simulation scenario	$P > 0$
$K_{residence}$	All residences in the network	$f \in K_{residence} \subset F$
$N^h_{residence}$	Residence location of household $h \in H$ at facility $f \in K_{residence}$	$N^h_{residence} \in K_{residence}$
$N^h_{hhd-income}$	The household income of household $h \in H$ (\$)	$N^h_{hhd-income}$
N^n_{hhd}	Household membership of person $n \in N$ in household $h \in H$	$N^n_{hhd,h} \in \mathbb{R},\ N^n_{hhd,h} \geq 0$
$N^h_{vehicles}$	The number of vehicles owned by household $h \in H$	$N^h_{vehicles} \geq 0$
N^n_{age}	The age of person agent $n \in N$	$N^n_{age} \in [0, A]$
N^n_{income}	The income of person agent $n \in N$ (\$)	$N^n_{income} \in [0, I]$
N^n_{gender}	The gender of person agent $n \in N$	$N^n_{gender} \in \{female,\ male\}$
N^n_{VOT}	The value of time (VOT) of person agent $n \in N$ (\$/hour)	$N^n_{VOT} \in \mathbb{R},\ N^n_{VOT} > 0$
$N_{avg-VOT}$	The average VOT of the population (\$/hour)	$N_{avg-VOT} \in \mathbb{R},\ N_{avg-VOT} > 0$
N^n_{plan}	An ordered list of facilities at which agent $n \in N$ complete planned activities $N^n_{activity,1}$ through $N^n_{activity,P}$	$N^n_{activity,i} \in F,\ i \in \{1, \ldots, P\}$
$N^{n,p}_{start}$	The desired start time of activity $p \in N^n_{plan}$	$N^{n,p}_{start} \in \mathbb{R},\ N^{n,p}_{start} \geq 0$
$N^{n,p}_{end}$	The end time of activity $p \in N^n_{plan}$	$N^{n,p}_{end} \in \mathbb{R},\ N^{n,p}_{end} \geq 0$

Population configuration

Prior to the simulation, a synthetic population is generated. The population consists of a set of households, H. Each household $h \in H$ owns a number of vehicles, $N^h_{vehicles} \geq 0$. The Boolean identifier, $Z^v_{hhd,h}$ denotes whether vehicle $v \in V_{private}$ is owned by household h such that

$$Z^v_{hhd,h} = \begin{cases} 1 \text{ if vehicle } v \text{ is owned by household } h \in H \\ 0 \text{ otherwise} \end{cases}$$

Each person agent in the population, $n \in N$, is a member of one household, denoted by the variable $N^n_{hhd} \in H$, with home location denoted by $N^h_{residence} \in K_{residence}$.

Additionally, each person agent is assigned fixed socio-demographic attributes. These include variables identifying age, N^n_{age}, gender, N^n_{gender}, and income, N^n_{income}. The maximum age and income, A and I, respectively, are configurable input variables. Additionally, the input $N^h_{hhd-income}$ denotes the household income of household h.

Finally, each person is instantiated with an activity plan, $N^n_{plan} = \{N^n_{plan,i}\}_{i=1,\ldots,P}$, which is an ordered list of facilities at which agents will complete their planned activities, starting and ending at home. The maximum number of activities in each plan is a configurable input variable, P. For simplified plans with just one activity outside of the home ($P = 3$), agents have plans of the form

$$N^n_{plan} = \{N^{N^n_{hhd}}_{residence}, N^n_{activity,2}, N^{N^n_{hhd}}_{residence}\}$$

where $N^n_{activity,2}$ may be a work facility or other, *secondary activity*, or point of interest. Each activity in a person's plan, $p \in N^n_{plan}$, has a desired start and end time, denoted (in seconds from the start of the simulation) by $N^{n,p}_{start}$ and $N^{n,p}_{end}$, respectively.

User-defined input

As mentioned in Section A.4, users may alter aspects of the transportation system by, for example, redefining the transit vehicle fleet composition, bus service schedules, and fares, and/or incentive amounts for particular demographic groups to use bus and/or on-demand ride services. Example specifications for these four UDIs mentioned are provided below. In a base case scenario, the transit fleet may be homogeneous, with all vehicles in the fleet configured with the same vehicle type:

$$Z^v_{veh-type,c} = Z^{v'}_{veh-type,c} \; \forall v, v' \in V_{transit}, c \in C$$

Users may choose to alter the vehicle type servicing each transit route, $t \in \{1, 2, \ldots, T\}$, by changing the value of the variable $D^t_{veh-type,c}$, which denotes the vehicle type for route $r \in R^t_{routes}$. Thus

$$D^{t,r}_{veh-type,c} = D^v_{veh-type,c} = D^{v'}_{veh-type,c} \; \forall v, v' \in V^{r,t}_{fleet}, c \in C_{transit}$$

Table A.7: Example User-Defined Input Notation

Example User-Defined Input: Transit Vehicle Fleet Configuration		
Notation	Description	Specification
$D^{t,r}_{veh-type}$	Vehicle type assignment for route $r \in R^t_{routes}$ of transit agency $t \in T$	$D^{t,r}_{veh-type,c} \in C_{transit}$
Example User-Defined Input: Transit Route Rescheduling		
Notation	Description	Specification
$n_{max-periods}$	The maximum number of allowable user-defined service periods	$n_{max-periods} \in \mathbb{Z},\ n_{max-periods} \geq 0$
$n^{t,r}_{periods}$	The number of user-defined service periods for route $r \in R^t_{routes}$ of transit agency $t \in T$	$n^{t,r}_{periods} \in \mathbb{Z},\ n^{t,r}_{periods} \in [0, n_{max-periods}]$
$D^{t,r}_{start,i}$	The start time of user-defined service period i for route $r \in R^t_{routes}$ of transit agency $t \in T$	$D^{t,r}_{start,i} \in \mathbb{Z},\ D^{t,r}_{start,i} \geq 0$
$D^{t,r}_{end,i}$	The end time of user-defined service period i for route $r \in R^t_{routes}$ of transit agency $t \in T$	$D^{t,r}_{end,i} \in \mathbb{Z},\ D^{t,r}_{start,i} < D^{t,r}_{end,i} \leq D^{t,r}_{start,i+1}$
$D^{t,r}_{schedule}$	A vector of $n^{t,r}_{periods}$ service period tuples for transit route $r \in R^t_{routes}$ of transit agency $t \in T$	$D^{t,r}_{schedule} \in \mathbb{Z}^{n^{t,r}_{periods},2}$
$n_{min-headway}$	Minimum allowable user-defined headway (seconds)	$n_{min-headway} \in \mathbb{Z},\ n_{min-headway} \geq 0$
$n_{max-headway}$	Maximum allowable user-defined headway (seconds)	$n_{max-headway} \in \mathbb{Z},\ n_{max-headway} \geq 0$
$D^{t,r}_{headway}$	User-defined headway for route $r \in R^t_{routes}$ of transit agency $t \in T$	$D^{t,r}_{headway} \in \mathbb{Z}^{n^{t,r}_{periods}},\ D^{t,r}_{headway} \in [n_{min-headway}, n_{max-headway}]$
Example User-Defined Input: Transit Pricing		
Notation	Description	Specification
D^t_{fare}	A vector of user-defined fares charged to transit riders of each age for using transit agency $t \in T$	$D^t_{fare} \in \mathbb{R}^A,\ D^t_{fare,a} \in \mathbb{R}, D^t_{fare,a} \geq 0$
Example User-Defined Input: Multi-modal Incentives		
Notation	Description	Specification
$D^m_{incentive}$	A matrix of user-defined incentive amount for agents of age a and income group i to use mode $m \in M$	$D^m_{incentive} \in \mathbb{R}^{A,I},\ D^m_{incentive,a,i} \in \mathbb{R},\ D^m_{incentive,a,i} \geq 0$

Users may redefine the service period of a transit route by appending one or more tuples to the mutable, ordered array of tuples, $D^{t,r}_{schedule}$, corresponding to the bus route t. Thus, a bus route with $n^{t,r}_{periods}$ user-defined service periods has the form:

$$D^{t,r}_{schedule} = \{(D^{t,r}_{start,1}, D^{t,r}_{end,1}), (D^{t,r}_{start,2}, D^{t,r}_{end,2}), \dots (D^{t,r}_{start,n^{t,r}_{periods}}, D^{t,r}_{end,n^{t,r}_{periods}})\}$$

where

$$D^{r,t}_{start,i} < D^{r,t}_{end,i} \; \forall i \in \{1, 2, \dots n^{t,r}_{periods}\}$$

and

$$D^{r,t}_{end,i} \leq D^{r,t}_{start,i+1} \; \forall i \in \{1, 2, \dots n^{t,r}_{periods} - 1\}$$

A business rule that constrains the maximum number of service periods that can be defined during optimization, $n_{max-periods}$ is recommended.

Users may redefine the headway of a route using the ordered array of $n^{t,r}_{periods}$ variables, $D^{t,r}_{headway} = \{D^{t,r}_{headway,i}\}_{i=1,\dots,n^{t,r}_{periods}}$, corresponding to each service period. An additional set of business rules that constrain the range of possible user-defined headway values, $n_{min-headway}$ and $n_{max-headway}$, are recommended.

The fare for riding with transit agency $t \in T$ can be redefined by the set of user-defined variables, $D^{t}_{fare} = \{D^{t}_{fare,a}\}_{a=0,1,\dots,A}$. All fares must be nonnegative.

Finally, the last example UDI demonstrate how users may choose to allocate multi-modal incentives. An incentive for mode $m \in M$ may be defined based on any number of parameters. Here, an age- and income-based incentive is shown:

$$D^{m}_{incentive} = \{D^{m}_{incentive,a,i}\}_{a=0,1,\dots,A; i \in I}$$

where the indices a, i correspond to the age and income group of riders, respectively. All incentive values must be nonnegative.

Output Variable Specification

Each BEAM simulation run outputs records of all person and vehicle agent events that occur during the run. In this subsection, specification is provided only for output variables that are directly used in the default KPIs provided in the BISTRO KPI library.

Person output

The person agent output reports the choices, movements, and expenditures of each person agent in the population during a simulation run. Each person agent, $n \in N$ takes $R^{n}_{trips} = X^{n}_{activities} - 1$ trips during the simulation, where $X^{n}_{activities}$ is the number of activities in the agent plans for a given scenario.

The mode(s) available to person agent n for trip $r \in \{1, \dots, R^{n}_{trips}\}$ are output as a set of modes, $X^{n,r}_{available}$.

Table A.8: Person Output Notation

Person Output		
Notation	Description	Specification
N	All person agents in a simulation scenario	$n \in N$
P	Maximum number of activities completed by each agent in a scenario	$P > 0$
$X^n_{activities}$	Number of activities in the plan of agent $n \in N$	$m^n_{activities} \in \mathbb{Z}_+$
R^n_{trips}	Number of trips made by agent $n \in N$	$R^n_{trips} = X^n_{activities} - 1$
$G_{n,r}$	Number of legs in trip $r \in \{1, \ldots, R^n_{trips}\}$ of agent $n \in N$	$G_{n,r} \geq 1$
$X^{n,r}_{available}$	Set of mode(s) available to person agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r}_{available,i} \in M$
$X^{n,r}_{choice}$	The mode chosen by person agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r}_{choice} \in X^{n,r}_{available}$
$X^{n,r,g}_{choice}$	The mode chosen by person agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{choice} \in M$
$X^{n,r,g}_{vehicle}$	The vehicle in which person agent $n \in N$ completed leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{vehicle} \in V$
$X^{n,r,g}_{route}$	The agency and route chosen by agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{route,1} \in T \cup \emptyset,\ X^{n,r,g}_{route,2} \in R^{X^{n,r,g}_{route,1}}_{routes} \cup \emptyset$
$p_{n,r,g}$	The number of links traversed by person agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$p_{n,r,g} \in \mathbb{Z},\ p_{n,r,g} > 0$
$X^{n,r,g}_{path}$	List of links traversed by agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{path} := \{X^{n,r,g}_{path,i} \in L\}_{i=1,\ldots,p_{n,r,g}}$
$X^{n,r,g}_{distance}$	Total distance traveled by agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$ (m)	$X^{n,r,g}_{distance} \in \mathbb{R}_+$
$X^{n,r,g}_{duration}$	Total duration traveled by agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$ (sec)	$X^{n,r,g}_{duration} \in \mathbb{Z}_+$
$X^{n,r,g}_{fare,m}$	Fare paid by agent $n \in N$ for mode $m = X^{n,r,g}_{choice}$ during leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{fare,m} \in \mathbb{R}_+$
$X^{n,r,g}_{fuel-cons}$	Total fuel consumed by person agent $n \in N$ by driving a personal vehicle during leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$ (mJ)	$X^{n,r,g}_{fuel-cons} \in \mathbb{R}_+$
$X^{n,r,g}_{fuel-cost}$	Total fuel cost to agent $n \in N$ of driving during leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$ (\$)	$X^{n,r,g}_{fuel-cost} \in \mathbb{R}_+$
$X^{n,r}_{incentive}$	Total incentive amount available to agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$ (\$)	$X^{n,r}_{incentive} \in \mathbb{R}_+$
$X^{n,r}_{exp}$	Total (net) expenditure incurred by agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$ (\$)	$X^{n,r}_{exp} \in \mathbb{R}_+$

The mode ultimately chosen by a person agent for trip $r \in \{1, \ldots, R^n_{trips}\}$ is output as the variable, $X^{n,r}_{choice} \in X^{n,r}_{available}$. where

$$X^{n,r}_{choice,m} = \begin{cases} 1 \text{ if mode } m \text{ is chosen by person } n \text{ for trip } r \in \{1, \ldots, R^n_{trips}\} \\ 0 \text{ otherwise} \end{cases}$$

Trips may include multiple legs. Thus, the number of legs, $G_{n,r}$, included in trip $r \in \{1, \ldots, R^n_{trips}\}$ of agent $n \in N$, is

$$G_{n,r} \begin{cases} = 1 \text{ if } X^{n,r}_{choice,walk} = 1 \text{ or } X^{n,r}_{choice,on-demand} = 1 \text{ or } X^{n,r}_{choice,bicycle} = 1, \text{ etc.} \\ >= 2 \text{ otherwise} \end{cases}$$

Transit trips include an access and egress leg as well as one or more transit legs, depending on the use of one or more transit routes during the trip. The mode used during each leg is denoted by the variable $X^{n,r,g}_{mode}$.

$$X^{n,r,g}_{mode} = \begin{cases} 1 \text{ if mode } m \text{ is used by person } n \text{ for leg } g \text{ of trip } r \\ 0 \text{ otherwise} \end{cases}$$

Each trip leg is made using a vehicle, recorded by the output variable, $X^{n,r,g}_{vehicle} \in V$. The output variable $X^{n,r,g}_{route} = \{X^{n,r,g}_{route,1}, X^{n,r,g}_{route,2}\}$ records the transit agency, $X^{n,r,g}_{route,1}$, and route, $X^{n,r,g}_{route,2}$ used by agent n during leg g of trip r such that $X^{n,r,g}_{route} = \{\emptyset, \emptyset\}$ if the agent did not use transit during leg g.

Each leg of a trip traverses a path, recorded as an ordered list of $p_{n,r,g}$ links traversed,

$$X^{n,r,g}_{path} = \{X^{n,r,g}_{path,i}\}_{l \in \{1,\ldots,p_{n,r,g}\}}$$

where $X^{n,r,g}_{path,i} \in L$ is a link in the transportation network.

The distance traveled (in miles) by agent n during leg g of trip r is

$$X^{n,r,g}_{distance} = \sum_{l \in X^{n,r,g,}_{path}} Z^{X^{n,r,g}_{path,l}}_{length}$$

The duration (in seconds) of each leg is a result of the traffic dynamics of the simulation and is recorded by the ouput variable $X^{n,r,g}_{duration} \geq 0$. The expenditure (in dollars \$) incurred by a person agent during a trip is the sum of all fare(s) paid for the legs of the trip, if any, plus additional expenditures from driving a personal vehicle, if applicable.

The applicable fare for person agent $n \in N$ to use mode $m \in M$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$ is given by the mode-specific fare output, $X^{n,r,g}_{fare,m}$. The default mode-specific fares for transit and on-demand rides are defined below. If additional fare-charging modes are included in the simulation, such as bikesharing, carsharing, or other shared mobility, additional mode-specific fares may be defined.

By default, the transit fare incurred by agent n during a trip leg is defined either by the corresponding configuration input variable, or by the user-defined transit fare input variable, in the case that the transit fare UDI is used for a particular simulation run:

$$X^{n,r,g}_{fare,transit} = \begin{cases} Z^{X^{n,r,g}_{route,1}}_{fare,N^n_{age,a}} if \;\; D^t_{fare} = \emptyset \\ D^{X^{n,r,g}_{route,1}}_{fare,N^n_{age,a}} otherwise \end{cases}$$

The on-demand ride fare for service type $s \in S$ incurred by agent n during any trip leg is

$$X^{n,r,g}_{fare,on-demand,s} = Z^{on-demand,s}_{fare,base} + X^{n,r,g}_{distance} Z^{on-demand,s}_{fare,distance} + \frac{1}{60} X^{n,r,g}_{duration} Z^{on-demand,s}_{fare,duration}$$

The cost of fuel consumed by agent n during any trip leg is the product of the total amount of fuel consumed by driving during that leg, $X^{n,r,g}_{fuel-consumed}$, and the corresponding fuel cost of the vehicle used during the leg, as follows:

$$X^{n,r,g}_{fuel-cost} = X^{n,r,g}_{choice,drive} X^{n,r,g}_{fuel-consumed} \sum_{c \in C_{personal}} Z^{X^{n,r,g}_{vehicle}}_{veh-type,c} Z^{Z^c_{fuel-type}}_{fuel-cost}$$

Where $Z^v_{veh-type,c}$ is a Boolean indicator denoting which vehicle type $c \in C_{private}$ corresponds to the vehicle $X^{n,r,g}_{vehicle} \in V_{personal}$ that was used during the trip leg, $Z^c_{fuel-type}$ denotes the fuel type used by that vehicle type, and $Z^K_{fuel-cost}$ is the cost per mJ consumed of that fuel type.

The incentive available to agent n during trip r is given by the output variable:

$$X^{n,r}_{incentive} = D^{X^{n,r}_{choice}}_{incentive,N^n_{age},N^n_{income}}$$

Where $D^m_{incentive,a,i}$ denotes the user-defined incentive amount available for a person agent of age a and income i, N^n_{age} and N^n_{income} are the age and income, respectively, of agent n.

The total expenditure incurred per person agent trip is captured by the output variable,

$$X^{n,r}_{exp} = \left(\sum_{g=1}^{G_{r,n}} \left(X^{n,r,g}_{fare,X^{n,r,g}_{choice}} + X^{n,r,g}_{fuel-cost} \right) - X^{n,r}_{incentive} \right)_+$$

where

$$(x)_+ = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

In the event that the incentive amount available to an agent for a particular trip exceeds the total fare and/or fuel cost of the trip, the agent receives an incentive amount equal to the total fare and fuel cost incurred during the trip.

Table A.9: Vehicle Output Notation

Vehicle Output		
Notation	Description	Specification
V	All vehicles in the transportation system	$v \in V := V_{transit} \cup V_{on-demand} \cup V_{private}$
Q_v	The total number of movements made by vehicle $v \in V$	$Q_v \in \mathbb{Z},\ Q_v \geq 0$
$p_{v,q}$	The number of links traversed by vehicle $v \in V$ during movement $q \in [1, Q_v]$	$p_{v,q} \in \mathbb{Z},\ p_{v,q} > 0$
$Y^{v,q}_{path}$	The ordered set of links traversed by vehicle $v \in V$ during movement $q \in [1, Q_v]$	$Y^{v,q}_{path} := \{Y^{v,q}_{path,i}\}_{i=1,...,p_{v,q}},\ Y^{v,q}_{path,l} \in L$
$Y^{v,q}_{distance}$	The total distance traveled by vehicle $v \in V$ during movement $q \in [1, Q_v]$ (meters)	$Y^{v,q}_{distance} \in \mathbb{R},\ Y^{v,q}_{distance} > 0$
$Y^{v,q}_{duration}$	The total duration of movement $q \in [1, Q_v]$ of vehicle $v \in V$ (seconds)	$Y^{v,q}_{duration} \in \mathbb{Z},\ Y^{v,q}_{duration} > 0$
$Y^{v,q}_{fuel-consumed}$	The total fuel consumed by vehicle $v \in V$ during movement $q \in [1, Q_v]$ (mJ)	$Y^{v,q}_{fuel-consumed} \in \mathbb{R},\ Y^{v,q}_{fuel-consumed} > 0$
$Y^{v,q}_{fuel-cost}$	The total cost of fuel consumed by vehicle $v \in V$ during movement $q \in [1, Q_v]$ ($)	$Y^{v,q}_{fuel-cost} \in \mathbb{R},\ Y^{v,q}_{fuel-cost} \geq 0$
$Y^{v,q}_{op-cost}$	The total operational cost of movement $q \in [1, Q_v]$ by vehicle $v \in V$ ($)	$Y^{v,q}_{op-cost} \in \mathbb{R},\ Y^{v,q}_{op-cost} \geq 0$
$Y^{v,q}_{pax}$	The number of passengers on board vehicle $v \in V$ during movement $q \in [1, Q_v]$ ($)	$Y^{v,q}_{pax} \in \mathbb{Z},\ Y^{v,q}_{pax} \geq 0$

Vehicle output

The vehicle output reports the movements of all vehicles during a simulation run. Each vehicle, $v \in V$, makes $Q_v \geq 0$ movements. For buses, a movement consists of travel between two bus stops. For personal vehicles, a movement consists of travel between two parking facilities. Finally, for on-demand ride vehicles, a movement consists of the travel from origin to destination during any one of the three phases of service: empty, fetch, and fare.

The path, fuel consumption, and occupancy of each vehicle v is recorded upon every vehicle movement. Similar to paths for person agent legs, the path traversed by vehicle v during a movement, $q = \{1, 2, ..., Q_v\}$, is recorded as an ordered list of $p_{v,q}$ links traversed,

$$Y^{v,q}_{path} = \{Y^{v,q}_{path,l}\}_{l \in \{1,...,p_{v,q}\}}$$

where $Y^{v,q}_{path,l} \in L$ is a link in the network.

Thus, the distance traveled by vehicle $v \in V$ during movement $q \in Q_v$ is given by:

$$Y^{v,q}_{distance} = \sum_{l \in Y^{v,q}_{path}} Z^{l}_{length}$$

Where Z^{l}_{length} is the length of link $l \in L$ in the transportation network. The duration of the movement (in seconds) is recorded by the output variable $Y^{v,q}_{duration} > 0$, and the fuel consumed by the variable, $Y^{v,q}_{fuel-consumed}$.

Thus the cost of fuel consumed by vehicle v during movement q is the product of the fuel consumed during the movement and the fuel cost of the corresponding vehicle type:

$$Y^{v,q}_{fuel-cost} = Y^{v,q}_{fuel-consumed} \sum_{c \in C} Z^{v}_{veh-type,c} Z^{Z^{c}_{fuel-type}}_{fuel-cost}$$

The total operational cost, if applicable, for vehicle v during movement q is the product of the duration of the movement and the variable operational cost for the corresponding vehicle type, as follows:

$$Y^{v,q}_{op-cost} = \frac{1}{3600} Y^{v,q}_{duration} \sum_{c \in C} Z^{v}_{veh-type,c} Z^{c}_{var-op-cost}$$

The number of passengers in vehicle v during movement q is recorded by the output variable $Y^{v,q}_{pax} \geq 0$.

KPI Specification

All score components are assessed as the ratio of the value of the corresponding KPI in a simulation run to the corresponding KPI value in the business as usual simulation run; that is, the scenario run without any inputs. The following sections detail each of the KPI functions included in the composite score, which is explained in Section A.5.

Accessibility

Accessibility measurements utilize the network characteristics resulting for any given submission (*i.e.*, links $l \in L$ are weighted with average travel-times, and average bus headways are assigned to routes during $n_{periods}$ periods of interest). Accessibility is then calculated as the sum of the average number of points of interest (work or secondary) reachable from all nodes $w \in W$ by a specified road network mode (car or transit) within a specified amount of time. Separate calculations are made for car (incorporating drive alone and a lower-bound estimate for on-demand rides) and transit trips to compare the changes in accessibility across network users. For the purposes of calculating the accessibility KPI, the output variable $Y^{m,e,u,w}_{shortest}$ reports the set of links that correspond to the *shortest path directed network distance* (in units of time) for mode $m \in M$ to travel from node $u \in W$ to node $w \in W$ during time

Table A.10: Accessibility KPI Notation

KPI: Accessibility		
Notation	Description	Specification
$n_{periods}$	The number of time periods of interest for the accessibility KPI	$n_{periods} \in \mathbb{Z},\ n_{periods} > 0$
K_{work}	All work facilities in the network	$f \in K_{work} \subset F$
$K_{secondary}$	All secondary facilities in the network	$f \in K_{secondary} \subset F$
Z_w^K	Facility instance at node $w \in W$	$Z_w^K \in \{0, 1\}$
$Y_{shortest}^{m,e,u,w}$	The set of links in the shortest path from node $u \in W$ to node $w \in W$ using mode $m \in M$ during time period $e = 1 \dots n_{periods}$	$Y_{shortest}^{m,e,u,w} := \{Y_{shortest,i}^{m,e,u,w}\}_{i=1,\dots,n_{periods}},\ Y_{shortest,i}^{m,e,u,w} \in L$
$Y_{avg-tt}^{m,l,e}$	The average travel time for mode $m \in M$ on link $l \in L$ during time period e	$Y_{avg-tt}^{m,l,e} \in \mathbb{R},\ Y_{avg-tt}^{m,l,e} \geq 0$
τ	The travel time within which points of interest must be accessible in order to be counted for the the accessibility KPI (minutes)	$\tau \in \mathbb{R},\ \tau > 0$

period $e \in \{1, \dots, n_{periods}$. Thus, $Y_{shortest}^{m,e,u,w} = \{Y_{s-p,i}^{m,e,u,w}\}_{i=1\dots n_{sp}^{m,e,u,w}}$, where each element, of $Y_{s-p,i}^{m,e,u,w} \in Y_{shortest}^{m,e,u,w}$ is a link in the network ($Y_{shortest}^{m,e,u,w} \in L$). The average travel time for mode $m \in M$ on link $l \in L$ during time period e is recorded by the output variable $Y_{avg-tt}^{m,l,e}$. Thus, the accessibility KPI for trip purpose *purpose* (identified by the facility type of the point of interest) and mode m is calculated as follows:

$$K_{accessibility,purpose,m} = \frac{1}{n_{periods} \sum_{w \in W} 1} \sum_{e=1}^{n_{periods}} \sum_{u \in W} \sum_{w \in W} \sum_{f \in K_{purpose}} Z_w^f I_{\tau \geq \sum_{l \in Y_{shortest}^{m,e,u,w}} Y_{avg-tt}^{m,l,e}}$$

where Z_w^K is a Boolean indicator of whether facility $f \in K_{purpose}$ of type *purpose* is located at node $w \in W$ and the term $I_{\tau \geq \sum_{l \in Y_{shortest}^{m,e,u,w}} Y_{avg-tt}^{m,l,e}}$ indicates whether the total average travel time on the shortest path for mode m from node u to node w is less than or equal to the threshold travel time of τ as follows:

$$I_{\tau \geq \sum_{l \in Y_{shortest}^{m,e,u,w}} Y_{avg-tt}^{m,l,e}} = \begin{cases} 1 \ if\ \ \tau \geq \sum_{l \in Y_{shortest}^{m,e,u,w}} Y_{avg-tt}^{m,l,e} \\ 0 \ otherwise \end{cases}$$

1. **Work-based trips:**

$$K_{accessibility,work,m} = \frac{1}{n_{periods} \sum_{w \in W} 1} \sum_{e=1}^{n_{periods}} \sum_{u \in W} \sum_{w \in W} \sum_{f \in K_{work}} Z_w^f I_{\tau \geq \sum_{l \in Y_{shortest}^{m,e,u,w}} Y_{avg-tt}^{m,l,e}}$$

2. **Secondary trips:**

$$K_{accessibility,secondary,m} = \frac{1}{n_{periods} \sum_{w \in W} 1} \sum_{e=1}^{n_{periods}} \sum_{u \in W} \sum_{w \in W} \sum_{f \in K_{secondary}} Z_w^f I_{\tau \geq \sum_{l \in Y_{shortest}^{m,e,u,w}} Y_{avg-tt}^{m,l,e}}$$

Measures of Level of Service

The LoS of the transportation system are evaluated on average, across person trips. Each person, $n \in N$ completes R_{trips}^n trips from one activity to another throughout a simulation run. Each trip, $r \in \{1, \ldots, R_{trips}^n\}$ includes $G_{n,r}$ legs, each using a single mode of transportation.

1. **Average Generalized Transportation Cost Burden**: As detailed in Section A.4:3, the total travel expenditure, $X_{exp}^{n,r}$, for a person agent, $n \in N$, during any trip, $r \in \{1, \ldots, R_{trips}^n\}$, may include bus and/or on-demand ride fares, and the cost of fuel consumed less any applicable incentives.

   The average transportation cost burden is considered by trip purpose: for example, by work trips and secondary activity trips. Thus the average transportation cost burden for work trips is

$$K_{cost-burden,work} =$$
$$\frac{1}{\sum_{n \in N} \sum_{r=1}^{R_{trips}^n} I_{N_{plan,r+1}^n \in K_{work}}} \sum_{n \in N} \sum_{r=1}^{R_{trips}^n} I_{N_{plan,r+1}^n \in K_{work}} \frac{X_{exp}^{n,r} + X_{duration}^{n,r} \cdot N_{avg-VOT}}{\hat{N}_{hhd-income}^{N_{hhd}^n}}$$

$$\hat{N}_{hhd-income}^h = max(1, N_{hhd-income}^h)$$

$$I_{N_{plan,r+1}^n \in K_{work}} = \begin{cases} 1 \ if \ N_{plan,r+1}^n \in K_{work} \\ 0 \ otherwise \end{cases}$$

   where $N_{hhd-income}^{N_{hhd}^n}$ is the income of household $N_{hhd}^n \in H$ of which individual $n \in N$ is a member, and $N_{avg-VOT}$ is the population-wide average value of time (VOT).

Table A.11: Level of Service (LoS) KPI Notation

KPI: LoS- Average Transportation Cost Burden		
Notation	Description	Specification
N	All person agents in a simulation scenario	$n \in N$
R^n_{trips}	The number of trips made by agent $n \in N$ during a simulation run	$0 < R^n_{trips} \leq P - 1$
$G_{n,r}$	The number of legs in trip $r \in \{1, \ldots, R^n_{trips}\}$ of person agent $n \in N$	$G_{n,r} \geq 1$
$X^{n,r}_{exp}$	The total (net) expenditure incurred by person agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$ ($)	$X^{n,r}_{exp} \in \mathbb{R},\ X^{n,r}_{exp} \geq 0$
$X^{n,r}_{duration}$	The total duration traveled by person agent $n \in N$ during trip $r \in \{1, \ldots, R^n_{trips}\}$ (seconds)	$X^{n,r}_{duration} \in \mathbb{R},\ X^{n,r}_{duration} > 0$
$N^h_{hhd-income}$	The household income of household $h \in H$ ($)	$N^h_{hhd-income}$
$N_{avg-VOT}$	The average VOT of the population ($/hour)	$N_{avg-VOT} \in \mathbb{R},\ N_{avg-VOT} > 0$
KPI: LoS- Average Bus Crowding Experienced		
Notation	Description	Specification
$X^{n,r,g}_{choice}$	The mode chosen by person agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{choice} \in M$
G_{bus}	The total number of bus trip legs made by all person agents	$G_{bus} \in \mathbb{Z},\ G_{bus} \geq 0$
V_{bus}	The set of all vehicles in the bus fleet	$v \in V_{bus} \subset V$
$Y^{v,q}_{duration}$	The total duration of movement $q \in [1, Q_v]$ of vehicle $v \in V$ (seconds)	$Y^{v,q}_{duration} \in \mathbb{Z},\ Y^{v,q}_{duration} > 0$
$Z^v_{veh-type,c}$	Vehicle type assignment of vehicle $v \in V$	$Z^v_{veh-type,c} \in \{0, 1\}$
$Z^c_{seating}$	The passenger seating capacity of vehicle type $c \in C$	$Z^c_{seating} \in \mathbb{Z},\ Z^c_{seating} > 0$
$Z^c_{standing}$	The passenger standing capacity of vehicle type $c \in C$	$Z^c_{standing} \in \mathbb{Z},\ Z^c_{standing} \geq 0$
$Y^{v,q}_{pax}$	The number of passengers on board vehicle $v \in V$ during movement $q \in [1, Q_v]$	$Y^{v,q}_{pax} \in \mathbb{R},\ Y^{v,q}_{pax} \geq 0$
$Tm^{v,q}_{seated}$	The value of time multiplier for seated passengers on board vehicle $v \in V$ during movement $q \in [1, Q_v]$	$Tm^{v,q}_{seated} \in \mathbb{R},\ Tm^{v,q}_{seated} \geq 0$
$Tm^{v,q}_{standing}$	The value of time multiplier for standing passengers on board vehicle $v \in V$ during movement $q \in [1, Q_v]$	$Tm^{v,q}_{standing} \in \mathbb{R},\ Tm^{v,q}_{standing} \geq 0$

Similarly, the average transportation cost burden for secondary trips is

$$K_{cost-burden,secondary} =$$

$$\frac{1}{\sum_{n\in N}\sum_{r=1}^{R^n_{trips}} I_{N^n_{plan,r+1}\in K_{secondary}}} \sum_{n\in N}\sum_{r=1}^{R^n_{trips}} I_{N^n_{plan,r+1}\in K_{secondary}} \frac{X^{n,r}_{exp} + X^{n,r}_{duration}\cdot N_{avg-VOT}}{N'^{N^n_{hhd}}_{hhd-income}}$$

2. **Average Transit Crowding Experienced**: The average transit crowding KPI is by default calibrated with time multipliers corresponding to the perceived value of time (VOT) in buses. However, a similar KPI may be implemented for crowding on other types of transit services, such as rail, using mode-specific functions for the VOT multipliers. The average transit crowding KPI is computed from simulation output as follows (using the fleet configuration UDI):

$$K_{transit-crowding} =$$

$$\frac{1}{G_{transit}} \sum_{v\in V_{transit}} \sum_{q=1}^{Q_v} \left(Y^{v,q}_{duration} \sum_{c=1}^{C} D^v_{veh-type,c} (Tm^{v,q}_{seated} min(Z^c_{seating}, Y^{v,q}_{pax}) + Tm^{v,q}_{standing} (Y^{v,q}_{pax} - Z^c_{seating})) \right)$$

where G_{bus} is the total number of trip legs in which public transit was used such that

$$G_{transit} = \sum_{n\in N}\sum_{r=1}^{R^n_{trips}}\sum_{g=1}^{G_{n,r}} I_{X^{n,r,g}_{choice}=transit}$$

$Tm^{v,q}_{seated}$ and $Tm^{v,q}_{standing}$ are the time multipliers for seated and standing passengers to account for the relative disutility of time spent in public transit vehicles that are filled over seating capacity, respectively such that

$$Tm^{v,q}_{seated} = \begin{cases} 1.1 + log(\frac{Y^{v,q}_{pax}}{Z^c_{seating}}) \text{ if } Y^{v,q}_{pax} \geq Z^c_{seating} \\ 0 \text{ otherwise} \end{cases}$$

$$Tm^{v,q}_{standing} = \begin{cases} 1.1 + 2.5log(\frac{Y^{v,q}_{pax}}{Z^c_{seating}}) \text{ if } Y^{v,q}_{pax} > Z^c_{seating} \\ 0 \text{ otherwise} \end{cases}$$

Measures of congestion

The first two measures of congestion, total vehicle miles traveled (VMT) and total vehicle hours of delay, are assessed on aggregate across all vehicle movements in a simulation run.

Table A.12: Measures of Congestion KPI Notation

KPI: Congestion- Total Vehicle Miles Traveled (VMT)		
Notation	Description	Specification
V	All vehicles in the transportation system	$v \in V := V_{transit} \cup V_{on-demand} \cup V_{private}$
Q_v	The total number of movements made by vehicle $v \in V$	$Q_v \in \mathbb{Z},\ Q_v \geq 0$
$Y^{v,q}_{distance}$	The total distance traveled by vehicle $v \in V$ during movement $q \in [1, Q_v]$ (meters)	$Y^{v,q}_{distance} \in \mathbb{R},\ Y^{v,q}_{distance} > 0$
KPI: Congestion- Average Vehicle Delay per Passenger Trip		
Notation	Description	Specification
$R^n_{trips,motorized}$	The number of trips taken by agent $n \in N$ using a motorized mode $m \in \{drive, on-demand, walk-transit, drive-transit\}$	$R^n_{trips,motorized} \in \mathbb{Z},\ R^n_{trips,motorized} \geq 0$
$X^{n,r}_{choice}$	The mode chosen by person agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r}_{choice} \in M$
$X^{n,r,g}_{duration}$	The total duration traveled by person agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$ (seconds)	$X^{n,r,g}_{duration} \in \mathbb{Z},\ X^{n,r,g}_{duration} > 0$
$X^{n,r,g}_{path}$	The ordered set of links traversed by person agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{path} \in \mathbb{R}^{p_{n,r,g}},\ X^{n,r,g}_{path,l} \in L$
Z^l_{length}	Length of link $l \in L$ (meters)	$Z^l_{length} \in \mathbb{R},\ Z^l_{length} > 0$
$Z^l_{speed-limit}$	Free-flow speed of link $l \in L$ (meters/second)	$Z^l_{speed-limit} \in \mathbb{R},\ Z^l_{speed-limit} > 0$

1. **Total VMT**:

$$K_{vmt} = \sum_{v \in V} \sum_{q=1}^{Q_v} Y^{v,q}_{distance}$$

where $Y^{v,q}_{distance}$ is the total distance (in miles) traveled by vehicle v during movement q.

2. **Average Vehicle Delay per Passenger Trip**:

$$K_{pax-trip-delay} = \frac{1}{\sum_{n \in N} R^n_{trips,motorized}} \sum_{n \in N} \sum_{r \in \{1,\ldots,R^n_{trips}\}} \left(\sum_{g=1}^{G_{n,r}} X^{n,r,g}_{duration} - \sum_{l \in X^{n,r,g,}_{path}} \frac{Z^l_{length}}{Z^l_{speed-limit}} \right)$$

where $R^n_{trips,motorized}$ is the number of trips taken by agent n using a motorized mode during the simulation run:

$$R^n_{trips,motorized} = \sum_{r=1}^{R^n_{trips}} I_{X^{n,r}_{choice} \in \{drive, on-demand, walk-transit, drive-transit\}}$$

Financial Sustainability

Any intervention on the transportation system is likely to result in costs and benefits for the operation of mass transit in the city of interest. The financial sustainability component is computed as the difference between the total revenue collected through transit fares $K_{revenue}$ and the net cost of transit operations $K_{op-cost}$ and multi-modal incentives used to manage demand $K_{incentives-used}$:

$$K_{financial-sust} = K_{revenue} - (K_{op-cost} + K_{incentives-used})$$

1. **Operational costs**: Operational costs include fixed costs, variable hourly costs, and fuel costs for transit operations.

$$K_{op-cost} = \sum_{v \in V_{transit}} \sum_{q \in Q_v} Y^{v,q}_{fuel-cost} + Y^{v,q}_{op-cost}$$

where $Y^{v,q}_{fuel-cost}$ and $Y^{v,q}_{op-cost}$ are the total fuel and operational costs produced during movement q of vehicle v in the transit fleet during the simulation run.

2. **Incentives used**

$$K_{incentives-used} = \sum_{n \in N} \sum_{r=1}^{R^n_{trips}} X^{n,r}_{incentive} - X^{n,r}_{exp}$$

where $X^{n,r}_{exp}$ and $X^{n,r}_{incentive}$ are the expenditure and incentives used, respectively, by agent n during trip r. Both the trip expenditure and incentives must be nonnegative, as defined in section A.5. Incentives may only be used by qualifying agents, as defined by user-defined inputs.

3. **Revenue**:

$$K_{revenue} = \sum_{n \in N} \sum_{r=1}^{R^n_{trips}} \Big(\sum_{g \in G_{n,r}} I_{X^{n,r,g}_{choice} = transit} X^{n,r,g}_{fare,transit} \Big) - X^{n,r}_{incentive}$$

where $X^{n,r,g}_{choice}$ indicates the mode that agent n chose to use for leg g of trip r. The variable $X^{n,r,g}_{fare,transit}$ is the transit fare paid, and $X^{n,r}_{incentive}$ is the incentive amount received by agent n for trip r.

Table A.13: Financial and Environmental Sustainability KPI Notation

KPI: Financial Sustainability		
Notation	Description	Specification
$V_{transit}$	All transit vehicles	$v \in V_{transit} \subset V$
Q_v	Number of movements made by vehicle $v \in V$	$Q_v \in \mathbb{Z}_+$
$Y^{v,q}_{op-cost}$	Total operational cost of movement $q \in [1, Q_v]$ by vehicle $v \in V$ (\$)	$Y^{v,q}_{op-cost} \in \mathbb{R}_+$
$Y^{v,q}_{fuel-cost}$	Total cost of fuel consumed by vehicle $v \in V$ during movement $q \in [1, Q_v]$ (\$)	$Y^{v,q}_{fuel-cost} \in \mathbb{R}_+$
N	All person agents in a scenario	$n \in N$
$X^n_{activities}$	The number of activities in the activity plan of person agent $n \in N$	$m^n_{activities} \in \mathbb{Z}_+$
R^n_{trips}	The number of trips made by agent $n \in N$ during a simulation run	$R^n_{trips} = X^n_{activities} - 1$
$G_{n,r}$	The number of legs in trip $r \in \{1, \ldots, R^n_{trips}\}$ of person agent $n \in N$	$G_{n,r} \geq 1$
$X^{n,r}_{incentive}$	Total incentive amount available to agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$ (\$)	$X^{n,r}_{incentive} \in \mathbb{R}_+$
$X^{n,r}_{exp}$	Total (net) expenditure incurred by agent $n \in N$ for trip $r \in \{1, \ldots, R^n_{trips}\}$ (\$)	$X^{n,r}_{exp} \in \mathbb{R}_+$
$X^{n,r,g}_{choice}$	The mode chosen by person agent $n \in N$ for leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{choice} \in M$
$X^{n,r,g}_{fare,m}$	The fare paid by person agent $n \in N$ for the use of mode $m = X^{n,r,g}_{choice}$ during leg $g \in G_{n,r}$ of trip $r \in \{1, \ldots, R^n_{trips}\}$	$X^{n,r,g}_{fare,m} \in \mathbb{R}_+$
KPI: Environmental Sustainability		
Notation	Description	Specification
$Y^{v,q}_{distance}$	The total distance traveled by vehicle $v \in V$ during movement $q \in [1, Q_v]$ (meters)	$Y^{v,q}_{distance} \in \mathbb{R}_+$
$Z^c_{fuel-type}$	The fuel type used by vehicle type $c \in C$	$Z^c_{fuel-type} \in \{gas, diesel, electricity\}$
$Z^v_{veh-type,c}$	Vehicle type assignment of vehicle $v \in V$	$Z^v_{veh-type,c} \in \{0, 1\}$
$PM^{c,f}_{2.5}$	$PM_{2.5}$ emission factor for vehicle type $c \in C$ using fuel type $f \in \{gas, diesel, electricity\}$	$PM^{c,f}_{2.5} \in \mathbb{R}_+$

Measures of environmental sustainability

Environmental sustainability will be assessed as the total particulate matter emitted from $PM_{2.5}$ running exhaust (RUNEX) and the GHG emissions from all vehicle movements in a simulation run. $PM_{2.5}$ emissions vary by mode and by fuel type; for Sioux Faux, there are only two possible vehicle-fuel type combinations: gasoline auto and diesel bus, simplifying the summation to:

1. **Total $PM_{2.5}$ Emissions**:

$$K_{PM_{2.5}} = \sum_{v \in V} \sum_{q=1}^{Q_v} Y_{distance}^{v,q} \sum_{c \in C} Z_{veh-type,c}^{v} \sum_{f \in F} PM_{2.5}^{c,Z_{fuel-type}^{c}}$$

   where $Y_{distance}^{v,q}$ is the total distance (in miles) traveled by vehicle v during movement q, $Z_{veh-type,c}^{v}$ is a Boolean indicator that vehicle v is of type c, $Z_{fuel-type}^{c}$ is the fuel type used by vehicle type c, and $PM_{2.5}^{c,f}$ is defined as follows:

$$PM_{2.5}^{c,diesel} = \begin{cases} 0.259366648 \ grams/mile \text{ if the vehicle is a bus} \\ 0.018403666 \ grams/mile \text{ if the vehicle is a car} \\ 0 \text{ otherwise} \end{cases}$$

$$PM_{2.5}^{c,gasoline} = \begin{cases} 0.002517723 \ grams/mile \text{ if the vehicle is a bus} \\ 0.001716086 \ grams/mile \text{ if the vehicle is a car} \\ 0 \text{ otherwise} \end{cases}$$

2. **Total *GHG* Emissions**: BEAM uses a high resolution data-driven vehicle energy consumption model based on a similar model[9]. Consequently, similarly detailed vehicle emissions statistics can be backed out of BEAM energy usage outputs.

Composite Score

The BISTRO scoring function serves as the objective function by which the UDIs are optimized. The selection and/or definition of the objective function is left to the decision of the project owner, according to project directives; herein, a general structure is defined to facilitate users in the creation of custom objective functions. Multiple project objectives (referred here specifically as *score components*) may be included in the scoring function, either as individual elements within a vector of scalar-valued score components to be minimized or as parameters to a function that aggregates the objectives into a one-dimensional scalar score. The score components are computed as the normalized ratio of the value of the corresponding KPI in the given simulation run to the value of the same KPI in the

[9]To find the description of this model, read: https://www.nrel.gov/docs/fy17osti/69121.pdf

Table A.14: Composite Score Function Notation

Composite Score Parameters		
Notation	Description	Specification
$\vec{C_s}$	A vector of all simulation inputs for alternative solution, s	See specifications in SectionA.5
$\vec{K}$	A vector of all KPIs evaluated for solution s, $K_i(C_s)$	$K_i(C_s) \in \mathbb{R}, \forall i$
$\vec{\sigma}$	A vector of standard deviations of each KPI produced by random search	$\sigma_i \in \mathbb{R}, \forall i$
$\vec{\mu}$	A vector of means of each KPI produced by random search	$\mu_i \in \mathbb{R}, \forall i$
$\vec{\alpha}$	A vector of user-defined parameters for each score component	n/a
$\vec{z}$	A vector of normalized KPIs	$\vec{z_i} \in \mathbb{R}, \forall i$

business-as-usual (BAU) run.[10] The improvement ratios are normalized using KPI values produced by a randomized sample of the UDI space, the size of which can be defined by the BISTRO project owner. This normalization accounts for differences in variance across KPIs, thus allowing the score components to provide meaningful feedback on the improvement achieved for each KPI relative to the distribution of the ratios of KPI to BAU produced by the random search. The composite score is thus a function of the normalized relative improvements of the candidate input to the BAU in each metric, as follows:

$$F\left(\vec{C}_a, \vec{K}, \vec{\sigma}, \vec{\mu}, \vec{\alpha}\right) = f\left(\vec{z}, \vec{\alpha}\right), \tag{A.5.1}$$

where $\vec{K}$ is the vector of all KPIs evaluated for a given set of inputs (see Table A.15 for example specified in Section A.5), $\vec{C_s}$; $\vec{\mu}$ and $\vec{\sigma}$ are the vectors of normalization parameters; and $\vec{z}$ is a vector of each KPI's z-scores, *i.e.*,

$$z_i = \frac{\frac{K_i(C_s)}{K_i(C_{BAU})} - \mu_i}{\sigma_i}, \tag{A.5.2}$$

for the i-th KPI. The value of the i-th score component in the BAU case is simply $K_i(C_{BAU})$.

The default objective is to minimize the composite score function, since an increase in many of the score components actually represents a scenario that is worse than the *status quo* (*e.g.*, decreasing VMT over BAU results in a lower unscaled score than increasing VMT). To maintain consistency in this regard, the scoring function may include an additional parameter

[10]In the BAU of a given scenario, the simulation is run without alteration from the initial configuration of that scenario.

Table A.15: Example KPI Notation

Example KPIs		
Notation	Description	Specification
K_i	The scoring function for KPI i	$K_i \in \mathbb{R}$
$K_{accessibility,purpose,m}$	Accessibility by mode $m \in M$ to POIs for $purpose \in \{work, secondary\}$ (number of POIs)	$K_{accessibility,purpose,m} \geq 0$
$K_{cost-burden,purpose}$	The average generalized transportation cost burden for trips of $purpose \in \{work, secondary\}$ ($)	$K_{generalized\ cost-burden,purpose} \geq 0$
$K_{bus-crowding}$	The average bus crowding experienced (seconds)	$K_{bus-crowding} \geq 0$
K_{vmt}	The total vehicle miles traveled (VMT) across all vehicle in a simulation run (miles)	$K_{vmt} > 0$
$K_{pax-trip-delay}$	The average vehicle delay per passenger trip (seconds)	$K_{pax-trip-delay} \geq 0$
$K_{financial-sust}$	The net revenue of total transit operational costs, transit revenue and multi-modal incentives distributed ($)	
$K_{PM_{2.5}}$	The total $PM_2.5$ emissions from all vehicle movements in the simulation run (g)	$K_{PM_{2.5}} \geq 0$

$\vec{\alpha}$ to allow for transformation of score components that are positively related to desirable outcomes (*e.g.*, improvements in accessibility). For example, if the scoring function takes the form of a sum over all score components, the parameter $\vec{\alpha}$ may be used as a coefficient of each score component that determines whether the component will be summed or subtracted, as follows:

$$\alpha_i = \begin{cases} -1 \text{ if it is desirable for score component } i \text{ to increase} \\ 1 \text{ otherwise} \end{cases} \tag{A.5.3}$$

Appendix B

Congestion Pricing Optimization Case Study II: San Francisco Bay Area: Supplementary Material

B.1 Methods

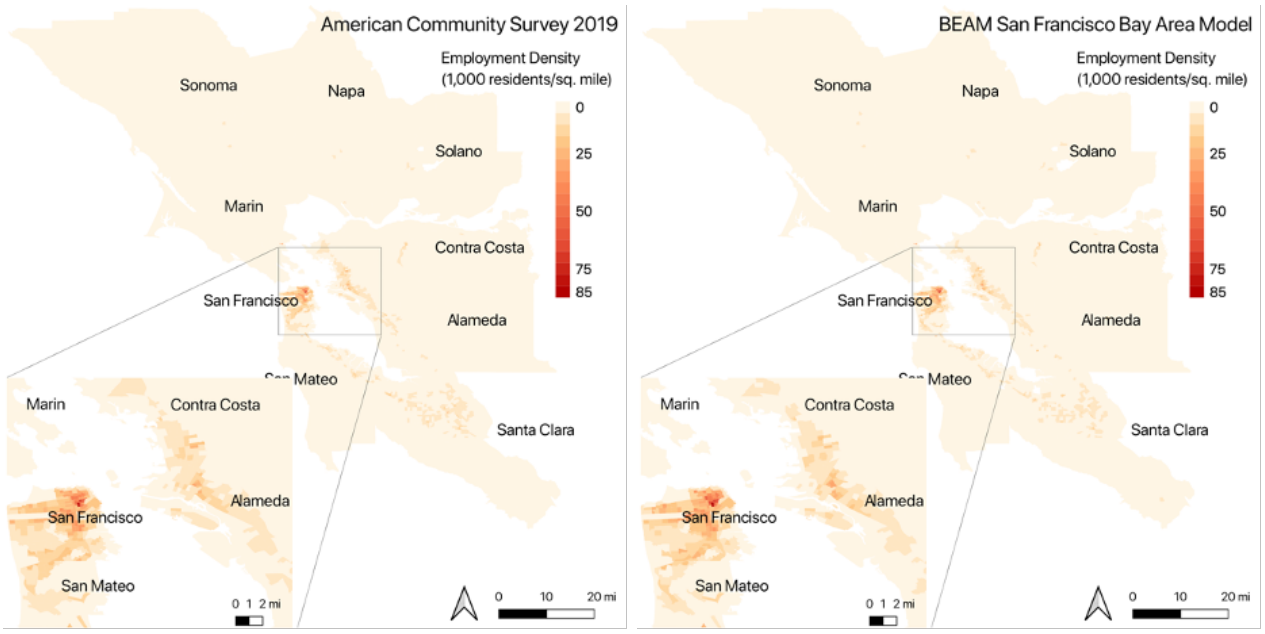
SFSim Model Specification

Figure B.1 displays the employment density per Traffic Analysis Zone (TAZ) of the San Francisco Bay Area as estimated by the 2015 American Community Survey (ACS) in B.1a, and as modeled in the BEAM travel model of the San Francisco Bay Area in B.1b. As seen in the figure, the San Francisco Bay Area BEAM model achieves a similar spatial distribution of commuters per square mile across the TAZ in the region.

The GHG and $PM_{2.5}$ emissions KPIs are calculated using the VMT and fuel consumption and emissions rates corresponding to each vehicle and fuel type specified in the SFSim model. Table B.1 documents the fuel consumption rate and capacity for each private vehicle type in the SFSim model and table B.2 documents the emissions rates of GHG and $PM_{2.5}$ for each fuel type.

	Fuel consumption (Joule/meter)		**Fuel capacity** (Joules)	
Vehicle type	Gasoline	Electricity	Gasoline	Electricity
ICEV	3656 (21 mpg)		3655980000	
Hybrid	2153 (35 mpg)			
PHEV	3656 (21 mpg)	671	3655980000	269999983
BEV		671		53999997

Table B.1: Vehicle type specification for SFSim model



(a) American Community Survey (2015) (b) BEAM San Francisco Bay Area Model

Figure B.1: Employment Density per Traffic Analysis Zone in the San Francisco Bay Area

	Emission rates	
Fuel type	GHG (g CO_2e /gal)	$PM_{2.5}$ (g $PM_{2.5}$/ mile)
Gasoline	11405.84	0.001716086
Diesel	13718.04	0.002593665
Electricity	378.54	0

Table B.2: Vehicle emissions rate specification for SFSim model

Estimation of mode split targets

The mode split targets used for calibration of the SFSim model were estimated using the San Francisco Metropolitan Transportation Agency's (SFMTA) 2019 Travel Decision Survey (TDS). The survey (n=841) was conducted via random telephone sampling of residents of the San Francisco Bay Area from May to August 2019. Respondents reported the details of their most recent two days of travel to/from/within the City of San Francisco, including the purpose and mode of transportation used for each trip. The publicly available survey data includes individual weights that were calculated based on age and location.

The sample of commute trips consisted of 137 responses, resulting in a margin of error of about 9%. In order to estimate target mode splits for commute trips by income group for the SFSim model, we computed the weighted estimates of the number of commute trips

made by each of four mode groups within San Francisco and between San Francisco and other counties. The mode groups included:

- drive: drive alone, drive with others, drive a carpool;
- public transit: any public transit service, such as Muni, BART, Caltrain, ferry or public bus;
- TNC: any TNC or regular taxi service; and
- other: walk, bike, motorcycle, scooter, or self-reported "other" mode.

Finally, household income data for each individual in the sample of commute trips was used to estimate the mode splits for each income group.

Cordon search space specification

The search space for the latitude, longitude, and radius of the congestion pricing cordon (d_1, d_2, and d_3, respectively) were determined using a heuristic algorithm, summarized in algorithm 1 below. The algorithm was designed to identify multiple search spaces for the optimization of more than one cordon at a time. The goal of the algorithm is to optimize the coverage of the cordons to target areas with the highest volume of cars across all simulated time periods. Car volume is determined by the congestion level (g_l) on each link $l \in \mathcal{L}$ where g_l is calculated as the ratio of the total car volume vol_l on the link to the capacity cap_l of the link. In step 2 we clustered the congested links into several groups, where group number was the number of cordons. In step 3 we constructed a polygon for each cluster with the convex hull of all the links. In step 4 we formed the cordon search spaces using the circumscribed circles of the convex hulls and adjust the radius of cordons if they were overlapping.

Algorithm 1: Cordon parameters initialization

Result: Initial cordon parameters $d_1^{i_0}$,$d_2^{i_0}$,$d_3^{i_0}$

1. Calculate the congestion level $g_l = vol_l/cap_l$ of every road link l using the output from BEAM;

*Pick the α most congested links;

2. Cluster the picked links into N clusters using K-Means algorithm;
3. Find the convex hull H_i of each cluster ;
4. Find the circumscribed circle C_i of each H_i, where $d_1^{i_0}$,$d_2^{i_0}$ is the center of C_i and $r_{i_0} = C_{radius}$. If there is overlap between C_i and C_j then we adjust the $d_3^{i_0} = d_3^{j_0} = max(d_3^{i_0}, d_3^{j_0})\beta$ where beta is a factor less than 1.

The parameters α and β are hyper-parameters that determine the number and degree of overlap of the cordons, respectively. Specifically, the hyper-parameter α is the percentile of links that will be regarded as congested as follows: let $\mathcal{L}$ be the set of all links, then the set of congested links is defined as $\mathcal{L}' = \{l \in \mathcal{L} : g_l \geq g_i, \forall i \in \mathcal{L} \setminus \mathcal{L}'\}$ where $|\mathcal{L}'| = \alpha|\mathcal{L}|$. The lower α is, the sparser the distribution of cordons will be. The sparsity will make the

Cordon		Longitude																													
param-	Latitude	0.19	Latitude																												
eters	Radius	0.26	0.15	Radius																											
	Toll rate	0.20	0.17	0.25	Toll rate																										
Quadrant	NW	-0.36	0.42	-0.03	0.04	NW																									
	NE	0.49	0.59	0.14	0.12	-0.26	NE																								
	SE	0.47	-0.39	0.12	0.07	-0.28	-0.34	SE																							
	SW	-0.61	-0.51	-0.22	-0.21	-0.31	-0.39	-0.41	SW																						
Incentive	1^{st} threshold	0.15	0.05	0.25	0.11	-0.07	0.06	0.09	-0.08	1^{st} threshold																					
param-	2^{nd} threshold	0.21	0.07	0.24	0.12	-0.11	0.12	0.10	-0.12	0.52	2^{nd} threshold																				
eters	PT - 1^{st} level	0.21	0.12	0.27	0.14	-0.11	0.15	0.11	-0.16	0.38	0.42	PT - 1^{st} level																			
	PT - 2^{nd} level	0.21	0.15	0.31	0.18	-0.07	0.16	0.09	-0.18	0.34	0.36	0.74	PT - 2^{nd} level																		
	PTNC - 1^{st} level	0.08	0.05	0.14	0.11	0.01	0.03	0.06	-0.09	0.28	0.27	0.20	0.17	PTNC - 1^{st} level																	
	PTNC - 2^{nd} level	0.14	0.05	0.26	0.14	-0.02	0.03	0.11	-0.12	0.30	0.28	0.25	0.25	0.70	PTNC - 2^{nd} level																
KPIs	VD	-0.48	-0.31	-0.60	-0.47	0.14	-0.37	-0.12	0.34	-0.33	-0.49	-0.46	-0.51	-0.26	-0.36	VD															
	VMT	-0.40	-0.31	-0.73	-0.51	0.09	-0.34	-0.09	0.33	-0.34	-0.46	-0.46	-0.53	-0.19	-0.31	0.93	VMT														
	GHG	-0.40	-0.31	-0.73	-0.51	0.09	-0.34	-0.09	0.33	-0.35	-0.46	-0.46	-0.53	-0.20	-0.31	0.93	1.00	GHG													
	PM	-0.40	-0.32	-0.71	-0.49	0.10	-0.34	-0.09	0.33	-0.36	-0.49	-0.48	-0.55	-0.21	-0.32	0.94	1.00	1.00	PM												
	CB	-0.30	-0.11	-0.07	-0.07	0.21	-0.23	-0.07	0.11	-0.34	-0.58	-0.50	-0.44	-0.40	-0.35	0.65	0.47	0.47	0.51	CB											
	TC	-0.39	0.22	-0.18	-0.15	0.38	-0.08	-0.31	0.06	-0.29	-0.47	-0.42	-0.47	0.03	0.00	0.43	0.38	0.38	0.40	0.45	TC										
	NPR	-0.27	-0.29	-0.48	-0.34	-0.05	-0.25	-0.02	0.29	0.10	0.19	0.05	0.08	0.09	0.07	0.35	0.43	0.43	0.40	-0.16	-0.05	NPR									
Revenue	Toll revenue	0.40	0.34	0.86	0.44	-0.04	0.33	0.09	-0.36	0.24	0.25	0.29	0.34	0.16	0.28	-0.80	-0.88	-0.88	-0.87	-0.24	-0.21	-0.62	Total toll revenue								
break-	PT revenue	0.20	0.19	0.09	0.20	0.05	0.17	0.03	-0.22	-0.06	0.05	0.09	0.02	0.00	-0.04	-0.19	-0.20	-0.20	-0.20	-0.02	-0.07	-0.72	0.17	Total PT revenue							
down	Incentives	0.31	0.18	0.47	0.24	-0.10	0.20	0.14	-0.24	0.46	0.77	0.61	0.66	0.40	0.49	-0.74	-0.72	-0.73	-0.76	-0.66	-0.49	0.10	0.51	0.07	Total incentives						
Mode	Drive alone	-0.40	-0.30	-0.77	-0.57	0.06	-0.30	-0.12	0.34	-0.35	-0.45	-0.43	-0.49	-0.25	-0.37	0.91	0.97	0.97	0.96	0.40	0.36	0.46	-0.92	-0.19	-0.72	Drive alone					
splits	Public transit	0.40	0.29	0.77	0.54	-0.08	0.31	0.12	-0.33	0.32	0.42	0.41	0.49	0.10	0.21	-0.88	-0.97	-0.97	-0.96	-0.36	-0.43	-0.51	0.92	0.21	0.65	-0.97	Public transit				
	Ride alone TNC	-0.37	-0.25	-0.83	-0.44	0.04	-0.24	-0.15	0.33	-0.42	-0.54	-0.45	-0.47	-0.34	-0.47	0.74	0.82	0.82	0.81	0.28	0.31	0.33	-0.82	-0.15	-0.75	0.89	-0.82	Ride alone TNC			
	Pooled TNC	0.04	-0.02	-0.03	-0.12	-0.03	-0.02	0.06	-0.02	0.24	0.45	0.17	0.11	0.60	0.67	-0.15	-0.02	-0.03	-0.06	-0.41	0.04	0.41	-0.06	-0.12	0.53	-0.08	-0.13	-0.25	Pooled TNC		
	Active	0.26	0.28	0.66	0.65	0.05	0.21	0.07	-0.30	0.20	0.08	0.23	0.26	0.33	0.42	-0.65	-0.71	-0.71	-0.68	-0.06	0.02	-0.51	0.77	0.16	0.37	-0.80	0.69	-0.75	0.09	Active (walk, bike)	
TNC	Deadheading	0.13	0.11	0.35	0.20	-0.02	0.12	0.01	-0.10	-0.06	-0.14	0.06	0.14	-0.56	-0.51	-0.18	-0.33	-0.32	-0.29	0.22	-0.27	-0.50	0.38	0.17	-0.16	-0.27	0.47	-0.12	-0.88	0.09	Deadheading
ops	PTNC 4+ pax	0.07	0.01	0.12	-0.04	-0.02	-0.01	0.08	-0.05	0.27	0.48	0.15	0.11	0.60	0.68	-0.23	-0.12	-0.13	-0.16	-0.39	0.03	0.30	0.10	-0.09	0.57	-0.20	0.01	-0.38	0.95	0.19	-0.79

Table B.3: Correlation matrix of the Pearson correlation coefficients for all pairs of input parameters, KPIs, mode splits, and other relevant metrics across all samples (n = 879)

cordons different from each other and thus have more opportunity to make more informed decisions on the parameters in the optimization process. At the same time, the lower α is, the fewer links are considered as congested. If the α is too low, we will have falsely neglected some congestion links.

The hyper-parameter β reduces the initial overlap of the cordons. When the clusters are located close to each other, β should be smaller to avoid overlap. For a pair of overlapping cordons, the maximum radius between the two is chosen to avoid having very small cordons which are not practical. Given a fixed number of cordons, overlapping cordons reduce the dimensions of the search space for optimization. Thus, we sought to force each cordon to be different as a way to preserve complexity and consequently increase exploration of the search space for a fixed value of N.

After the initial parameter vector $(d_1^{i_0}, d_2^{i_0}, d_3^{i_0})$ is calculated, the search space for d_1^i, d_2^i, d_3^i are specified as follows: $d_1^i \in [d_1^{i_0} - 0.5 * d_2^{i_0}, d_1^{i_0} + 0.5 * d_3^{i_0}]$, $d_2 \in [d_2^{i_0} - 0.5 * d_3^{i_0}, d_2^{i_0} + 0.5 * d_3^{i_0}]$, $d_3 \in [0, d_3^{i_0}]$.

B.2 Results

Table B.3 presents the Pearson's correlation coefficients for the input parameters, KPIs, and other metrics across all 879 samples in the sample history of the case study.