
# Transportation Science



# Transportation Science and Logistics Society Best Dissertation Award Competition: Abstracts of 2025 Winners

informs
https://pubsonline.informs.org/journal/trsc

TRANSPORTATION SCIENCE
Vol. 60, No. 1, January–February 2026, pp. v–xiii
ISSN 0041-1655 (print), ISSN 1526-5447 (online)



# Transportation Science and Logistics Society Best Dissertation Award Competition: Abstracts of 2025 Winners

**Published Online in Articles in Advance:**
January 5, 2026

**https://doi.org/10.1287/trsc.2025_dissertation_award**



**Abstract.** The journal is pleased to bring back the tradition of publishing the abstracts of the winners of the TSL Best Dissertation Award. The 2025 dissertation prize committee was chaired by Hai Wang. The other committee members were Jelmer Pier van der Gaast, Sophie Parragh, Tal Raviv, and Joan Walker.

**Funding:** Fabian Akkerman conducted his research in the project DynaPlex: Deep Reinforcement Learning for Data-Driven Logistics, made possible by TKI Dinalog and funded by the Ministry of Economic Affairs and Climate Policy (EZK) of the Netherlands. Bo Lin was supported by City of Toronto Transportation Services and the Natural Sciences and Engineering Research Council of Canada [NSERC Alliance Grant ALLRP 561212-20].

The 2025 prize winners are as follows:

**Winner**
*Machine Learning for Sequential Decisions in Logistics*
Fabian Akkerman, University of Twente
Advisors: Martijn Mes, Maria Iacob, Willem van Jaarsveld

**Runner-up**
*Analytics for Better Urban Cycling*
Bo Lin, University of Toronto
Advisors: Timothy Chan and Shoshanna Saxe

# Machine Learning for Sequential Decisions in Logistics

**Fabian Akkerman**

University of Twente, 7522 NB Enschede, Netherlands
**Contact:** f.r.akkerman@utwente.nl, https://orcid.org/0000-0001-8055-9864

The movement of goods is central to modern economies: Efficient logistics ensure timely deliveries, stocked shelves, and the functioning of global trade. As consumer expectations and complexity grow, companies must adapt quickly, making dynamic automated decision making essential. This thesis explores how machine learning (ML) can support sequential decision making in logistics. Operations research (OR) has provided robust and structured methods for tackling challenges in logistics. As these challenges increasingly involve complexity, uncertainty, and real-time dynamics, ML offers complementary tools that can enhance and extend traditional OR approaches. Unlike rule-based systems, ML can adapt to new information, making it well suited to sequential decision making under uncertainty. The study investigates how ML can be effectively integrated into such decision processes, drawing on real-world collaborations that span the logistics spectrum: from warehouse operations at a logistics company and spare part sourcing with a leading energy provider to municipal waste collection planning and last-mile delivery optimization for an e-grocery retailer.

## Scope of the Thesis

Logistics activities can be classified by their position in the production and distribution process: supply logistics (before production), internal logistics (during production), and distribution logistics (after production). The thesis focuses on *external logistics* (before and after production) highlighted in gray in Figure 1 (Ghiani, Laporte, and Musmanno 2013). It is organized into three parts: supply logistics, distribution logistics, and revenue management within distribution logistics. Each part showcases different applications of ML in addressing sequential decision-making challenges under uncertainty.

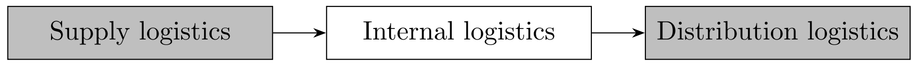
## Supply Logistics

We focus on inventory control, addressing the challenge of replenishing items at the right time and

Transportation Science, 2026, vol. 60, no. 1, pp. v–xiii, © 2026 INFORMS



**Figure 1.** High-Level Overview of Logistics Activities




quantity to meet demand. In the United States, inventory is valued at $2,500 billion—around 10% of the gross domestic product (GDP) (U.S. Census Bureau 2024). Overstocks and stockouts cost companies $562 billion and $1.2 trillion, respectively (IHL Group 2023). Policies balance availability and cost, but added complexities, for example, batch reordering (Topan, Bayindir, and Tan 2017), dual sourcing (Svoboda, Minner, and Yao 2021), and multiechelon systems (de Kok et al. 2018), make exact solutions intractable. ML enables more adaptive and efficient approximate policies through data-driven prediction and optimization.

ML is widely applied in inventory control, particularly for demand forecasting, which supports decision making indirectly by improving demand estimates. More recently, attention has shifted toward using ML to support decisions more directly, learning policies or actions that guide operational choices under uncertainty. A seminal study by Gijsbrechts et al. (2022) applied reinforcement learning (RL) to various inventory problems.

In Akkerman, Prak and Mes (2025), published in the *European Journal of Operational Research*, we propose a hybrid OR-ML model that combines neural networks and mixed-integer programming to improve replenishment and inspection decisions. The problem is based on the real-world operations of Bolk Transport, an industry partner in logistics and warehousing. The issue of inventory record inaccuracy (IRI)—caused by spoilage, theft, or obsolescence—leads to discrepancies between physical inventory and system records, posing a major challenge in inventory management. Although prior work often considers reordering and inspection in isolation or under highly stylized assumptions, our approach addresses the full dynamics of joint decision making in large warehouses with many heterogeneous stock-keeping units subject to shrinkage and limited inspection capacity. We introduce an algorithmic pipeline in which a neural network recommends dynamic replenishments based on historical demand and stock patterns, whereas a mixed-integer program selects an optimal inspection subset under operational constraints, accounting for both inspection and travel times in a warehouse. This hybrid architecture allows the system to flexibly adapt reordering based on past observations while optimizing which items to inspect under resource limitations. Numerical experiments show that ignoring shrinkage can increase total costs by up to 95.3%, whereas inspection becomes cost-effective at relatively low levels of IRI. Compared with benchmark policies, our method achieves cost savings of 8.3% over a static reordering and inspection policy, and 21.8% over a no-inspection policy.

In Akkerman et al. (2025), published in the *International Journal of Production Research*, we study a dual sourcing problem for spare parts, where the failure rates are supply mode dependent. This is especially relevant in downtime-critical environments, where dual sourcing with conventional and additive manufacturing enhances supply resilience. A key challenge is that parts produced by different methods exhibit distinct failure characteristics, which in turn influence future demand. The problem is modeled as a Markov decision process (MDP) and solved using RL methods, enhanced with an endogenous parameterized learning (EPL) framework. EPL enables a single policy to generalize across varying input parameters and item types, allowing for scalable, instance-independent policy deployment. We propose both a novel iterative heuristic and several RL-based approaches, showing that our best-performing policy achieves an average optimality gap of just 0.4% in stylized settings. In a real-world case study with TotalEnergies, a global energy company, using data from offshore platforms in remote locations, our methods outperform a baseline in 91.1% of instances, yielding average cost savings of up to 22.6%. This aligns with recent work applying RL to inventory control (Gijsbrechts et al. 2022, Oroojlooyjadid et al. 2022).

In Akkerman et al. (2024), presented at the *International Conference on Learning Representations*, we address the curse of dimensionality in batch reordering problems, where decision makers must select from large, combinatorial, sets of discrete decisions. In such high-dimensional settings, traditional methods like dynamic programming quickly become intractable. The core of our approach lies in exploiting the structured nature of these decision spaces, common in logistics, production, and transportation systems. To tackle the large decision space problem, we propose a scalable RL-based method called dynamic neighborhood construction (DNC), designed specifically for structured large discrete decision spaces. DNC introduces a novel neighborhood exploration paradigm that efficiently searches around a continuous proxy decision to identify promising discrete alternatives. This allows our method to handle decision spaces with up to $10^{73}$ decisions, far beyond the limits of current benchmarks. We benchmark DNC against three state-of-the-art approaches and show that it matches or outperforms them in terms of solution quality while being significantly more computationally efficient. As a result, DNC provides a practical path forward for applying RL in logistics settings that involve extremely large and structured decision spaces.

## Distribution Logistics

In this part, containing two papers, we focus on challenges in vehicle routing and last-mile delivery. In 2021, the global courier, express, and parcel market was valued at $407.7 billion, with last-mile delivery—especially in urban areas—accounting for 40%–50% of total logistics costs (Allied Market Research 2022). Classical vehicle routing problems (VRPs) literature has addressed deterministic problems and their extensions, including multiple depots, time windows, and pickups (Braekers, Ramaekers, and Van Nieuwenhuyse 2016). More recent work incorporates real-world complexities, such as stochastic demand (Zhang and Van Woensel 2023), travel time (Schilde, Doerner, and Hartl 2014), and resource availability (Arslan et al. 2019). These uncertainties make such problems inherently sequential and intractable for traditional methods (Hildebrandt, Thomas, and Ulmer 2023).

In distribution logistics, we see applications of ML both for forecasting and decision making. Supervised ML predicts demand locations (Hess, Spinler, and Winkenbach 2021), estimates arrival times (Hildebrandt and Ulmer 2022), and enhances classical VRP solvers (Baty et al. 2024). Applications of ML to support decision making primarily consider RL-based methods to learn routing policies for settings such as same-day delivery (Ulmer et al. 2019).

In Akkerman and Mes (2022), published at *Annals of Operations Research*, we use ML-based prediction models to estimate transportation costs in multiperiod customer selection, aiding in the customer selection process for routing problems. The setting involves selecting a subset of customers to serve in a given period, where computing the exact routing cost for each possible subset is computationally expensive. To address this, we train supervised regression models to predict the total distance of traveling salesman problems (TSPs) and VRPs based on spatial characteristics of customer sets. In addition to established features from the literature, we introduce new classes of spatial features that improve predictive accuracy. The predictions are then used within a selection heuristic that reduces problem dimensionality and enables rapid decision making, making the approach suitable for both offline and online settings. Although the predictions are static, the decision problem itself is inherently sequential because each inclusion affects future routing plans. The models are validated on both a fictional case—featuring various spatial layouts and demand configurations—and a real-world case involving dynamic waste collection in the city of Amsterdam. In the fictional case, our model achieves up to 25.3% savings in distance compared with a heuristic approximation. In the real case, we introduce a combined cost function that balances travel distance with service levels and demonstrate that our method reduces routing distance by up to 17% while maintaining service quality. This work connects to recent approaches that embed ML into vehicle routing workflows (Hess, Spinler, and Winkenbach 2021, Hildebrandt and Ulmer 2022, Baty et al. 2024).

In Akkerman, Mes, and van Jaarsveld (2025), published in *Computers and Industrial Engineering*, we examine dynamic VRPs (DVRPs), where customer requests arrive in real time and decisions must be made sequentially and under uncertainty. These problems are naturally modeled as MDPs and require methods that account for the future consequences of current decisions. A commonly used framework in this context is approximate dynamic programming or RL, often paired with parametric value or policy function approximations. We explore four such RL approaches: linear value function approximation (LVFA), linear policy function approximation (LPFA), neural network value function approximation (NNVFA), and neural network policy function approximation (NNPFA). Although LVFA is a straightforward and interpretable baseline, more complex neural network-based approximations may capture nonlinear patterns and richer contextual dependencies. To clarify the tradeoffs between these options, we provide a structured comparison of the four policy classes, highlighting their design differences and practical implications. We benchmark these methods on multiple variants of the DVRP with stochastic customer requests. Our evaluation includes two realistic use cases: same-day parcel pickup and delivery in the city of Amsterdam and robot routing in an automated storage and retrieval system (AS/RS) warehouse. The results show that slight variations of the problem might favor another RL approach. Particularly when using neural network approximators, we can learn anticipative vehicle positioning strategies that significantly outperform classical myopic heuristics. Our findings illustrate both the promise and complexity of deploying neural network-based RL for dynamic distribution logistics.

## Revenue Management in Distribution Logistics

In this part, containing three papers, we explore problems at the intersection of last-mile delivery and operational revenue management, a growing area fueled by the rise of e-commerce. Global retail e-commerce sales have increased by more than 370% in the past decade, reaching $6.3 trillion, with 22.6% of sales now online and one-third of the global population shopping digitally (Statista 2024). Rising expectations drive up costs—last-mile delivery makes up 53% of e-commerce logistics costs (Statista 2024). Revenue management helps mitigate this by managing demand (Snoeck, Merchán, and Winkenbach 2020). Although most literature focuses on offering differentiated time slots to individual customers (Agatz et al. 2013, Klein et al. 2018), our work aims to steer customer choices through incentives and dynamically alter delivery options to improve cost efficiency. These decisions must be made sequentially and in real time as customers arrive on the platform.

In Akkerman, Dieter, and Mes (2025), published in *Transportation Science*, we introduce a model that dynamically offers and prices out-of-home (OOH) delivery options, such as parcel lockers and parcel shops, using ML to adapt offerings in real time. OOH delivery provides a promising alternative to home delivery, which is often hampered by failed delivery attempts, traffic congestion, and handling inefficiencies—factors that negatively impact last-mile profitability. Although existing academic models for OOH delivery typically rely on static assumptions, we model the problem as a sequential decision process, where delivery locations and incentives must be dynamically selected for each arriving customer (triggered by a checkout on an e-commerce platform) based on the current system state and anticipated future demand. To address this, we propose dynamic selection and pricing of OOH (DSPO), an algorithmic pipeline that uses a spatial-temporal state encoding as input to a convolutional neural network. The network suggests location-incentive combinations that balance customer acceptance, locker capacity, and routing cost. Retraining during simulation enables adaptation to changing customer behavior and operational conditions. In an extensive numerical study informed by real-world data from Amazon delivery operations in Seattle, DSPO outperforms two state-of-the-art benchmarks: It reduces total costs by 19.9 percentage points compared with a scenario without OOH delivery, by 7 percentage points compared with a static selection and pricing policy, and by 3.8 percentage points compared with a demand management benchmark. Our findings highlight the complex interplay between customer behavior and delivery incentives and suggest that dynamic selection and pricing policies can yield significant efficiency gains in urban logistics.

Next, we focus on attended home delivery (AHD), where customers book delivery time slots in advance and companies aim to influence these choices to balance operational efficiency with customer satisfaction. The key challenge lies in dynamically pricing and offering time slots in a way that steers customer decisions while respecting routing and capacity constraints. In Akkerman, Mes, and Lalla-Ruiz (2022), presented at the *International Conference on Computational Logistics*, we develop a supervised ML-based pricing model trained on routing costs estimated via a cheapest insertion heuristic, comparing a rule-based approximation and a random forest model with limited features. This model enables dynamic pricing by estimating the marginal cost of each slot and adjusting incentives accordingly. In stylized experiments assuming full customer responsiveness, our method reduces per-customer travel costs by 11% and allows 6% more customers to be served. In a real case in collaboration with ORTEC and their client, a European e-grocery retailer, it achieves a 6% cost reduction and 1% increase in planned customers. In Akkerman, Visser, and Mes (2025) (working paper), we build on this by introducing decision-focused learning (DFL), which integrates a combinatorial optimization (CO) layer directly into the training process. This aligns the learning objective with routing performance rather than predictive accuracy. We extend DFL to a multistage setting that captures the sequential nature of customer arrivals over the booking horizon and demonstrate its scalability in complex AHD scenarios. DFL reduces the gap to an oracle solution by up to 44.3 percentage points in restrictive settings and consistently outperforms state-of-the-art heuristics in challenging configurations with many time slots or sparse road networks. Together, these separate papers highlight ML's role in dynamic time slot management and the need to align predictions with decision quality.

## Conclusion

Across these applications, the findings highlight two distinct roles for ML in logistics: *data analytics* and *decision analytics* (Powell 2022). Data analytics focuses on learning predictive models from static datasets to support better decisions through insights, typically in a one-shot fashion without ongoing feedback. This is seen, for example, in Akkerman and Mes (2022), Akkerman, Mes, and Lalla-Ruiz (2022), and Akkerman, Visser, and Mes (2025), where ML is used to estimate routing costs or customer preferences based on historical data.

In contrast, decision analytics aims to directly improve the quality of the decisions themselves. Here, the learning process interacts with the decision environment: each decision affects future states and costs, and the ML model must be updated accordingly. This is especially important in dynamic and stochastic settings, such as in Akkerman et al. (2025), Akkerman et al. (2024), Akkerman, Mes, and van Jaarsveld (2025), and Akkerman, Dieter, and Mes (2025).

These two perspectives can be formalized mathematically. In data analytics, the goal is to learn a predictive model $f(x|\theta)$, where $x \in \mathcal{X}$ represents the input features, $y \in \mathcal{Y}$ the corresponding labels, and $\theta$ the model parameters (e.g., weights in a neural network).

Given a data set of $N$ samples $\{(x^1, y^1), \ldots, (x^N, y^N)\}$, the model is trained to minimize the average prediction error:

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^{N} (y^n - f(x^n | \theta))^2.$$

In contrast, decision analytics focuses on learning a policy $\pi_\theta(x)$ that maps input features $x$ directly to a decision $a \in \mathcal{A}$, again parameterized by $\theta$. The objective is to minimize a downstream cost function $C(x, a)$, which evaluates the quality of the chosen decision in context. The corresponding learning objective becomes

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^{N} C(x^n, \pi_\theta(x^n)).$$

This distinction clarifies the difference between predicting outcomes (data analytics) and optimizing decisions (decision analytics), each requiring different learning objectives and feedback structures.

Based on the different papers in this thesis, which study a range of logistics problems, we propose a guiding framework (Akkerman 2025) that begins with data analytics to extract foundational insights and progresses to decision analytics when the operational context requires dynamic adaptation. This structured approach allows for increasing model complexity while retaining control over interpretability and implementation feasibility.

To conclude, this thesis advances the understanding of how ML can enhance sequential decision making in logistics by demonstrating that ML is not simply a substitute for OR but a complementary tool that excels in handling complexity, uncertainty, and adaptability in dynamic environments. As ML and OR continue to evolve, the integration of these fields offers a fertile ground for innovation. By strategically combining the data-driven capabilities of ML with the optimization and decision-making capabilities from OR, we can create more powerful, efficient, and transparent solutions to the complex challenges faced in logistics and beyond.

## References

Agatz N, Campbell A, Fleischmann M, Nunen J, Savelsbergh M (2013) Revenue management opportunities for internet retailers. *J. Revenue Pricing Management* 12:128–138.

Akkerman F (2025) Machine learning for sequential decisions in logistics. PhD thesis, University of Twente, Twente, Netherlands.

Akkerman F, Mes M (2025) Distance approximation to support customer selection in vehicle routing problems. *Ann. Oper. Res.* 350:269–297.

Akkerman F, Dieter P, Mes M (2025) Learning dynamic selection and pricing of out-of-home deliveries. *Transportation Sci.* 59(2): 250–278.

Akkerman F, Mes M, Lalla-Ruiz E (2022) Dynamic time slot pricing using delivery costs approximations. de Armas J, Ramalhinho H, Voβ S, eds. *Computational Logistics* (Springer International Publishing, Cham, Switzerland), 214–230.

Akkerman F, Mes M, van Jaarsveld W (2025) A comparison of reinforcement learning policies for dynamic vehicle routing problems with stochastic customer requests. *Computers Industrial Engrg.* 200:110747.

Akkerman F, Prak D, Mes M (2025) Dynamic reordering and inspection for the multi-item inventory record inaccuracy problem. *Eur. J. Oper. Res.* 321(2):428–444.

Akkerman F, Visser T, Mes M (2025) Decision-focused learning for time slot offering in attended home delivery. Working paper.

Akkerman F, Knofius N, van der Heijden M, Mes M (2025) Solving dual sourcing problems with supply mode dependent failure rates. *Internat. J. Production Res.*

Akkerman F, Luy J, van Heeswijk W, Schiffer M (2024) Dynamic neighborhood construction for structured large discrete action spaces. *Proc. 12th Internat. Conf. Learn. Representations*, 1–10.

Allied Market Research (2022) CEP market opportunities and forecasts. Accessed December 12, 2025, https://www.alliedmarketresearch.com/courier-express-and-parcel-market-A11516.

Arslan A, Agatz N, Kroon L, Zuidwijk R (2019) Crowdsourced delivery—A dynamic pickup and delivery problem with ad hoc drivers. *Transportation Sci.* 53(1):222–235.

Baty L, Jungel K, Klein P, Parmentier A, Schiffer M (2024) Combinatorial optimization-enriched machine learning to solve the dynamic vehicle routing problem with time windows. *Transportation Sci.* 58(4):708–725.

Braekers K, Ramaekers K, Van Nieuwenhuyse I (2016) The vehicle routing problem: State of the art classification and review. *Computers Industrial Engrg.* 99:300 –313.

de Kok T, Grob C, Laumanns M, Minner S, Rambau J, Schade K (2018) A typology and literature review on stochastic multi-echelon inventory models. *Eur. J. Oper. Res.* 269(3):955–983.

Ghiani G, Laporte G, Musmanno R (2013) *Introduction to Logistics Systems Management* (John Wiley & Sons, Hoboken, NJ).

Gijsbrechts J, Boute R, Van Mieghem J, Zhang D (2022) Can deep reinforcement learning improve inventory management? Performance on lost sales, dual-sourcing, and multi-echelon problems. *Manufacturing Service Oper. Management* 24(3):1349–1368.

Hess A, Spinler S, Winkenbach M (2021) Real-time demand forecasting for an urban delivery platform. *Transportation Res. Part E: Logist. Transportation Rev.* 145:102147.

Hildebrandt F, Ulmer M (2022) Supervised learning for arrival time estimations in restaurant meal delivery. *Transportation Sci.* 56(4): 1058–1084.

Hildebrandt F, Thomas B, Ulmer M (2023) Opportunities for reinforcement learning in stochastic dynamic vehicle routing. *Comput. Oper. Res.* 150:106071.

IHL Group (2023) OnDemand–The good, the bad, the ugly–inventory distortion. Accessed December 12, 2025, https://www.ihlservices.com/news/analyst-corner/2023/07/inventory-distortion-out-of-stocks/.

Klein R, Mackert J, Neugebauer M, Steinhardt C (2018) A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum* 40(4):969–996.

Oroojlooyjadid A, Nazari M, Snyder LV, Takáč M (2022) A deep q-network for the beer game: Deep reinforcement learning for inventory optimization. *Manufacturing Service Oper. Management* 24(1):285–304.

Powell W (2022) *Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions* (Wiley-Blackwell, Hoboken, NJ).

Schilde M, Doerner K, Hartl R (2014) Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-a-ride problem. *Eur. J. Oper. Res.* 238(1):18–30.

Snoeck A, Merchán D, Winkenbach M (2020) Revenue management in last-mile delivery: State-of-the-art and future research directions. *Transportation Res. Proc.* 46:109–116.

Statista (2024) Digital & trends: e-commerce worldwide. Accessed December 12, 2025, https://www.statista.com/study/10653/e-commerce-worldwide-statista-dossier/.

Svoboda J, Minner S, Yao M (2021) Typology and literature review on multiple supplier inventory control models. *Eur. J. Oper. Res.* 293(1):1–23.

Topan E, Bayındır ZP, Tan T (2017) Heuristics for multi-item two-echelon spare parts inventory control subject to aggregate and individual service measures. *Eur. J. Oper. Res.* 256(1):126–138.

Ulmer M, Goodson J, Mattfeld D, Hennig M (2019) Offline–Online approximate dynamic programming for dynamic vehicle routing with stochastic requests. *Transportation Sci.* 53(1): 185–202.

U.S. Census Bureau (2024) Manufacturing and trade inventories and sales. Accessed December 12, 2025, https://www.census.gov/.

Zhang J, Van Woensel T (2023) Dynamic vehicle routing with random requests: A literature review. *Internat. J. Production Econom.* 256:108751.

# Analytics for Better Urban Cycling

**Bo Lin**

University of Toronto, Toronto, Ontario M5S 3G8, Canada
**Contact:** imbo.lin@mail.utoronto.ca, https://orcid.org/0000-0002-4225-9171



Cycling has become increasingly popular due to its positive impact on public health (De Hartog et al. 2010, Mueller et al. 2018), urban mobility (Bruno and Nikolaeva 2020, Liu, Siddiq, and Zhang 2025), economic growth (Blondiau, Van Zeebroeck, and Haubold 2016), social equity (Giuffrida, Pilla, and Carroll 2023), and environmental sustainability (Larsen et al. 2013, Kou et al. 2020). The COVID-19 pandemic further accelerated this growth as cities worldwide built thousands of kilometers of bike lanes, recognizing cycling as a safe, low-cost alternative to public transit and driving that enabled outdoor activities and access to essential services (Buehler and Pucher 2021). This infrastructure expansion helped drive an estimated 11%–48% increase in global cycling demand, generating health benefits valued at 1–7 billion U.S. dollars (Kraus and Koch 2021). Notably, recent empirical studies indicate that much of this surge in cycling has persisted beyond the pandemic, suggesting a lasting shift in travel behavior (Büchel, Marra, and Corman 2022, Younes et al. 2023).

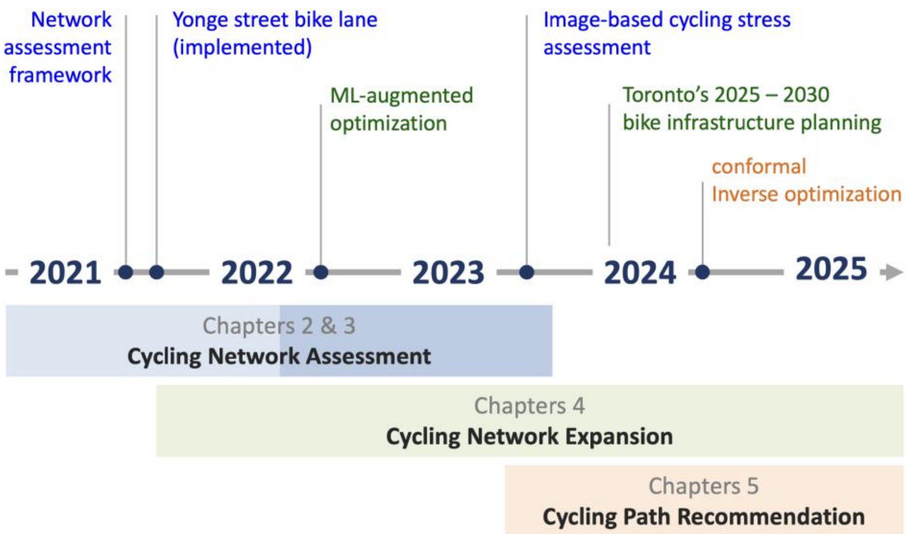
Despite recent growth, cycling mode share continues to be low, with safety and comfort concerns repeatedly identified as major barriers to cycling uptake (Dill and McNeil 2016, Manaugh, Boisjoly, and El-Geneidy 2017). In response, municipalities have recently made significant investments in bike infrastructure—one of the most effective interventions for reducing cycling stress and encouraging adoption (Furth, Mekuria, and Nixon 2016). New York City, for instance, is committed to building more than 230 miles of protected bike lanes from 2022 to 2026 (NYC DoT 2021), whereas Singapore aims to create a 1,300-km island-wide cycling network by 2030 (Singapore MoT 2024). Paris also plans to construct 180 km of bike lanes from 2021 to 2026 as part of its vision to become a 100% cycle-friendly city (Ville de Paris 2024). However, despite these strong commitments, available infrastructure budgets are usually limited relative to cities' extensive existing road networks. To maximize the impact of these investments, systematic planning, assessment, and utilization strategies are essential.

Toward the vision of making biking safer and more inviting, this thesis develops and implements a suite of quantitative methods to support the *assessment*, *planning*, and *utilization* of urban cycling infrastructure, with a particular focus on encouraging cycling adoption in Toronto, Canada. Our work demonstrates both practical and theoretical contributions: On the practical side, we demonstrate how operations research tackles key transportation planning challenges with tangible real-world impact; on the theoretical side, our methodologies—spanning descriptive, predictive, and prescriptive analytics—are principled, contributing to advancements in ML and optimization.

As visualized in Figure 1, the development of this thesis, spanning from 2020 to 2025, progressed through three successive phases, each building on the previous one. Although the development followed a sequential order, these tools remained in use beyond their initial phases, continuing to support Toronto's cycling infrastructure planning in later stages. A brief summary of these methods is provided below.

In Chapter 2, we review a descriptive framework to evaluate urban cycling networks (Lin et al. 2021). This framework comprises two steps. First, we assess the cycling stress—the discomfort cyclists experience—for each road segment and intersection by analyzing detailed road network data, including traffic conditions, road geometry, and the presence of bike infrastructure. Second, we calculate the network's low-stress cycling accessibility, defined as the total number of opportunities, such as jobs and grocery stores, that can be reached within 30 minutes from 3,702 population centers using low-stress routes. This accessibility metric, which measures the potential benefits provided by the cycling

**Figure 1.** (Color online) Thesis Structure




network, is strongly correlated with cycling mode choice in Toronto (Imani, Miller, and Saxe 2019). By analyzing accessibility metrics before and after new cycling infrastructure is built, this framework provides a systematic, data-driven approach to impact assessment, supporting infrastructure planning, policy discussions, and public engagement.

To illustrate the utility of this framework, we present a case study that evaluates the impact of bike infrastructure implemented by the City of Toronto during the COVID-19 pandemic. Since its development, this framework has been used to evaluate more than 10 major cycling infrastructure projects in Toronto, including the ActiveTO Midtown Yonge Complete Street Project (City of Toronto 2021b), which was implemented in 2021 in part based on a similar analysis. The derived insights have been widely referenced in news articles (CBC News 2021, Toronto Star 2021, U of T News 2021, TechXplore 2024) and policy documents (City of Toronto 2021a, b, 2024b) to facilitate public engagement and political discussions. This chapter lays the foundation for the rest of this thesis.

In Chapter 3, we address the challenge that existing cycling stress assessment methods rely on detailed road network data, which are often difficult to collect and come from inconsistent sources.[1] To overcome this limitation, we present the first computer vision approach for assessing cycling stress using street-view images (Lin, Saxe, and Chan 2024). Specifically, we frame cycling stress assessment as a two-step learning task: First, we predict cycling stress-related road features from street-view images and learning high-quality image representations; next, we combine the image embeddings with the predicted road features to predict the cycling stress of a road segment. This flexible framework allows available road features to be directly incorporated by substituting predicted features with observed ones in the second step. To improve performance, we propose two algorithms: one that leverages the ordinal nature of cycling stress labels and another that exploits the spatial structure of the road network. By utilizing the widespread availability of up-to-date street-view images, this method enables timely, accurate, and large-scale assessments of urban cycling networks. It is particularly beneficial for small- and midsized cities, where transportation planners are often overloaded and may lack the GIS expertise needed for such assessments, even when data are available.

In Chapter 4, building on the cycling network assessment method in Chapter 2, we propose an optimization method to determine the locations of new bike lanes under a given budget, with the goal of maximizing Toronto's low-stress cycling accessibility (Chan, Lin, and Saxe 2025). Given the size and complexity of Toronto's road network, the resulting optimization model is a large-scale mixed-integer bilevel problem with more than one million followers, each representing a cyclist traveling between an origin–destination pair. Existing solution methods struggle to find even a feasible solution before running out of memory. To address this challenge, we develop a ML-augmented optimization approach that is computationally tractable and capable of generating provably high-quality solutions.

Specifically, we explicitly model a sampled subset of followers, which forms a data set used to train a ML model that predicts the objective value of an individual follower based on relevant features. This model is then used to approximate the objective values of unsampled followers, bypassing the need to model their decisions explicitly. As a result, the optimization model is significantly smaller and more tractable than the original bilevel formulation. Despite this simplification, we theoretically demonstrate that the resulting model can still produce high-quality solutions to the original problem, as the ML model effectively captures the behavior of the unsampled followers. Unlike the conventional "predict-then-optimize" paradigm, where a ML model is trained to estimate unknown parameters in an optimization problem, our approach embeds ML model training within the optimization process and uses the trained model to approximate part of the objective function that depends on the decisions being optimized. This "predict-and-optimize" framework represents a new paradigm for integrating ML and optimization. Beyond cycling infrastructure planning, this approach is also applicable to two-stage stochastic programming and a broad class of problems commonly found in transportation planning, energy pricing, and financial markets.

From 2020 to 2024, our method achieved comparable performance to Toronto's implemented plan while reducing the required length of bike lanes by 25%, resulting in a cost saving of more than 8 million Canadian dollars. Because of its strong performance, the optimization method developed in this chapter has been used to support Toronto's 2025–2030 bike infrastructure planning (City of Toronto 2024b). As of March 2025, 29.1 km of bike lanes recommended by our optimization model had been approved by Toronto City Council for construction between 2025 and 2027, with an additional 27.6 km under study or in the design phase for future implementation (City of Toronto 2024a).

In Chapter 5, we develop a principled approach to prescribing decisions that are both high-quality and perceived as such by users and apply this method to provide cycling path recommendations (Chan, Delage, and Lin 2024, Lin, Delage, and Chan 2024). This methodology was motivated by observations in food-delivery applications, where platforms like Uber Eats and DoorDash provide path recommendations to cycling couriers. When these recommendations do not align with couriers' intuitions, couriers may choose to deviate, leading to increased delivery times and inefficiencies in order batching and assignment—both of which rely heavily on accurate delivery route modeling. To address this challenge, we develop an inverse optimization method to infer human perception (e.g., routing preferences) from historical decision data (e.g., past delivery paths). Unlike existing inverse optimization methods that provide a single point estimate representing the aggregated perception of all decision makers, our approach calibrates an uncertainty set that captures a decision maker's perception with high probability. This uncertainty set is subsequently used in a robust optimization model to prescribe decisions (e.g., recommend new delivery paths). We theoretically demonstrate that the prescribed decisions achieve both high absolute and perceived quality.

In a case study, we apply our decision pipeline to recommend last-mile delivery paths, significantly improving delivery path adherence without compromising delivery times compared with current industry practices—a rare win-win that benefits both platforms and workers.

In summary, this thesis demonstrates the transformative potential of integrating advanced analytics into urban cycling infrastructure planning. By developing a comprehensive suite of tools spanning descriptive (cycling stress and accessibility modeling), predictive (computer vision, contrastive learning, conformal prediction), and prescriptive (ML-augmented optimization, large-scale bilevel programming, inverse optimization, robust optimization) analytics, we provide an end-to-end, data-driven solution to support the assessment, planning, and utilization of cycling infrastructure. These tools have generated actionable insights that directly informed Toronto's infrastructure policy. As urban areas increasingly strive to promote sustainable and equitable transportation, our work offers a scalable, generalizable, and user-friendly solution to support these goals—paving the way for future research and collaboration in this critical domain.

## Endnote

[1] For example, in Toronto, road geometry and traffic speed data were accessed via Open Data Canada (Government of Canada 2020), whereas road type and cycling infrastructure data were obtained from Open Data Toronto (City of Toronto 2020). Integrating these data sets manually required approximately two months of effort at the start of my PhD.

## References

Blondiau T, Van Zeebroeck B, Haubold H (2016) Economic benefits of increased cycling. *Transportation Res. Proc.* 14:2306–2313.

Bruno M, Nikolaeva A (2020) Towards a maintenance-based approach to mode shift: Comparing two cases of Dutch cycling policy using social practice theory. *J. Transportation Geography* 86:102772.

Büchel B, Marra AD, Corman F (2022) COVID-19 as a window of opportunity for cycling: Evidence from the first wave. *Transportation Policy (Oxford)* 116:144–156.

Buehler R, Pucher J (2021) COVID-19 impacts on cycling, 2019–2020. *Transportation Rev.* 41(4):393–400.

CBC News (2021) How pandemic bike lanes made some Canadian cities more accessible. Accessed April 1, 2025, https://www.cbc.ca/news/science/pandemic-bike-lanes-canada-1.5951863.

Chan T, Delage E, Lin B (2024) Conformal inverse optimization for adherence-aware prescriptive analytics. Preprint, submitted November 7, https://doi.org/10.2139/ssrn.4970102.

Chan TCY, Lin B, Saxe S (2025) Machine learning-augmented optimization of large bilevel and two-stage stochastic programs: Application to cycling network design. *Manufacturing Service Oper. Management* 27(6):1851–1868.

City of Toronto (2020) City of Toronto open data. Accessed September 15, 2020, https://www.toronto.ca/city-government/data-research-maps/open-data/.

City of Toronto (2021a) 2021 cycling network plan update. Accessed February 6, 2025, https://www.toronto.ca/legdocs/mmis/2021/ie/bgrd/backgroundfile-173663.pdf.

City of Toronto (2021b) ActiveTO: Lessons learned from 2020 and next steps for 2021. Accessed February 6, 2025, https://www.toronto.ca/legdocs/mmis/2021/ie/bgrd/backgroundfile-164864.pdf.

City of Toronto (2024a) Cycling impact analysis. Accessed February 6, 2025, https://www.toronto.ca/legdocs/mmis/2024/ie/bgrd/backgroundfile-245676.pdf.

City of Toronto (2024b) Cycling network plan update (2025-2027). Accessed February 6, 2025, https://www.toronto.ca/legdocs/mmis/2024/ie/bgrd/backgroundfile-245671.pdf.

De Hartog JJ, Boogaard H, Nijland H, Hoek G (2010) Do the health benefits of cycling outweigh the risks? *Environ. Health Perspect.* 118(8):1109–1116.

Dill J, McNeil N (2016) Revisiting the four types of cyclists: Findings from a national survey. *Transportation Res. Rec.* 2587(1):90–99.

Furth PG, Mekuria MC, Nixon H (2016) Network connectivity for low-stress bicycling. *Transportation Res. Rec.* 2587(1):41–49.

Giuffrida N, Pilla F, Carroll P (2023) The social sustainability of cycling: Assessing equity in the accessibility of bike-sharing services. *J. Transportation Geography* 106:103490.

Government of Canada (2020) Government of Canada open data. Accessed September 15, 2020, https://open.canada.ca/en/open-data.

Imani AF, Miller EJ, Saxe S (2019) Cycle accessibility and level of traffic stress: A case study of Toronto. *J. Transportation Geography* 80:102496.

Kou Z, Wang X, Chiu SFA, Cai H (2020) Quantifying greenhouse gas emissions reduction from bike share systems: A model considering real-world trips and transportation mode choice patterns. *Resources Conservation Recycling* 153:104534.

Kraus S, Koch N (2021) Provisional COVID-19 infrastructure induces large, rapid increases in cycling. *Proc. Natl. Acad. Sci. USA* 118(15): e2024399118.

Larsen J, Patterson Z, El-Geneidy A (2013) Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. *Internat. J. Sustainable Transportation* 7(4):299–317.

Lin B, Chan T, Saxe S (2021) The impact of COVID-19 cycling infrastructure on low-stress cycling accessibility: A case study in the City of Toronto. *Findings.*

Lin B, Delage E, Chan TC (2024) Conformal inverse optimization. Globerson A, Mackey L, Belgrave D, Fan A, Paquet U, Tomczak J, Zhang C, eds. *Adv. Neural Inform. Processing Systems* 37 (Curran Associates, Inc., Red Hook, NY), 63534–63564.

Lin B, Saxe S, Chan TC (2024) AutoLTS: Automating cycling stress assessment via contrastive learning and spatial post-processing. Wooldridge M, Dy J, Natarajan S, eds. *Proc. AAAI Conf. Artificial Intelligence*, vol. 38 (AAAI Press, Washington, DC), 22222–22230.

Liu S, Siddiq A, Zhang J (2025) Planning bike lanes with data: Ridership, congestion, and path selection. *Management Sci.* 71(9): 7631–7654.

Manaugh K, Boisjoly G, El-Geneidy A (2017) Overcoming barriers to cycling: Understanding frequency of cycling in a university setting and the factors preventing commuters from cycling on a regular basis. *Transportation (Amsterdam)* 44:871–884.

Mueller N, Rojas-Rueda D, Salmon M, Martinez D, Ambros A, Brand C, De Nazelle A, et al. (2018) Health impact assessment of cycling network expansions in European cities. *Prevention Medicine* 109:62–70.

NYC DoT (2021) NYC streets plan. Technical report, New York City Department of Transportation, New York.

Singapore MoT (2024) Active mobility infrastructure. Accessed April 1, 2025, https://www.mot.gov.sg/what-we-do/active-mobility/active-mobility-infrastructure.

TechXplore (2024) Machine learning analysis sheds light on who benefits from protected bike lanes. Accessed April 1, 2025, https://techxplore.com/news/2024-10-machine-analysis-benefits-bike-lanes.html.

Toronto Star (2021) New study points out the benefits—Relaxed cycling access to jobs and more —Of Toronto's new bike—Lanes. Accessed April 1, 2025, https://shorturl.at/GMO1z.

U of T News (2021) Toronto's COVID-19 bike lane expansion boosted access to jobs, retail: U of T study. Accessed April 1, 2025, https://shorturl.at/YxDXf.

Ville de Paris (2024) Un nouveau plan vélo pour une ville 100% cyclable. Accessed April 1, 2025, https://www.paris.fr/pages/un-nouveau-plan-velo-pour-une-ville-100-cyclable-19554.

Younes H, Noland RB, Von Hagen LA, Sinclair J (2023) Cycling during and after COVID: Has there been a boom in activity? *Transportation Res. Part F: Traffic Psych. Behavior* 99:71–82.