# Constrained Distance-Based Congestion Pricing in Multi-Modal Urban Networks

Nasser Parishad*[1], Mehmet Yildirimoglu[1], and Mark Hickman[1]

[1]School of Civil Engineering, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Australia

## Short summary

While previous studies have examined the fairness and equity implications of alternative pricing schemes, designing an equitable congestion pricing methodology for complex multi-modal traffic systems that explicitly accounts for modal interactions and traveller heterogeneity remains a significant challenge. This study proposes a constrained, data-driven congestion pricing framework based on reinforcement learning (RL) to maximise network outflow while maintaining equity. A trip-based three-dimensional Macroscopic Fundamental Diagram (3D-MFD) simulation model is developed to serve as the RL environment, capturing multi-modal interactions among cars, buses, and walking. Traveller heterogeneity is incorporated through variations in origin–destination pairs, trip lengths, departure times, and individual values of time (VoT). Results indicate that the RL agent significantly mitigate congestion and reduce total travel time by up to 50%, while distributing cost burdens more equitably among travellers.
**Keywords**: Distance-based pricing, Equity, Reinforcement learning, Trip-based 3D-MFD

## 1 Introduction

Urban traffic congestion poses significant economic, environmental, and social challenges for cities worldwide. Congestion pricing has become a pivotal tool for managing urban traffic demand and promoting sustainable mobility (Balzer & Leclercq, 2022). By levying tolls on congested roads or areas, pricing policies influence travel behaviour—shifting departure times (Liu et al., 2023), altering route choices (Li & Ramezani, 2022), and encouraging modal shifts (Basso & Jara-Díaz, 2012). However, despite strong evidence of its effectiveness, designing efficient congestion pricing schemes for urban networks remains a complex challenge.

A practical pricing framework must explicitly account for multimodal interactions and for the differing impacts of transport modes on overall network performance. It must also accommodate substantial heterogeneity among travellers in terms of trip characteristics, socioeconomic conditions, behavioural preferences, and decision-making mechanisms. Beyond efficiency, congestion pricing must also address equity considerations (Ecola & Light, 2009), ensuring that pricing burdens are allocated fairly across user groups and that charges are perceived as reasonable and commensurate with network use. As a result, developing tolling schemes that simultaneously deliver efficiency, equity, and perceived fairness remains a central challenge in congestion pricing research.

Effective pricing models require a network-level representation of urban traffic, which Macroscopic Fundamental Diagrams (MFDs) offer by providing a parsimonious, region-level view of urban traffic dynamics. The MFD captures the relationship between vehicle accumulation in a homogeneous region and the network's overall throughput, generally displaying a uni-modal, low-scatter, demand-insensitive pattern (Geroliminis & Daganzo, 2008). However, real urban transportation systems comprise multiple transport modes operating under mixed-traffic conditions, where service stops significantly impact large-scale urban performance. To account for these complexities, Geroliminis et al. (2014) introduced a bi-modal MFD (3D-MFD) that integrates mode differentiation into traffic flow modelling.

This macroscopic perspective substantially reduces implementation complexity and operational costs, making congestion management more feasible in real-world settings. Numerous studies have employed MFD-based models to design optimal tolling policies for both single-region and multi-region networks (Zheng et al., 2012; Li & Ramezani, 2022). Common strategies include cordon-based, time-based, distance-based, and hybrid schemes such as joint time-distance and

joint distance-delay pricing. While alternative pricing structures, such as distance-based and hybrid schemes, may lead to more equitable outcomes, to the best of our knowledge, no prior study has directly incorporated quantitative equity measures into a simulation-based control framework for dynamic, real-time, network-scale pricing.

To overcome these limitations, this study combines Reinforcement Learning (RL) with a trip-based three-dimensional Macroscopic Fundamental Diagram (3D-MFD) model to design a distance-based pricing scheme for urban traffic networks (Sutton & Barto, 2018). Given that equity remains one of the most challenging aspects of congestion pricing, a novel Constrained Markov Decision Process (CMDP) framework is proposed to explicitly balance network outflow maximisation against equity preservation (Altman, 2021). By incorporating equity directly into the learning objective, the proposed framework seeks to limit disproportionate toll burdens across heterogeneous traveller groups.

## 2 Traffic dynamics and modelling

The simulation is built upon a multi-modal, grid-based network that includes three main transport modes: private cars, public transport (bus), and walking. In addition to supporting three distinct modes, the network enables multi-modal decision-making. All links are bi-directional, and all modes can move in all four cardinal directions. Cars and pedestrians are permitted to traverse the entire network, whereas buses are restricted to structured routes. Bus stations are placed at every third node along these corridors in both horizontal and vertical directions. Figure 1 presents a schematic representation of a portion of the network to aid visual understanding.

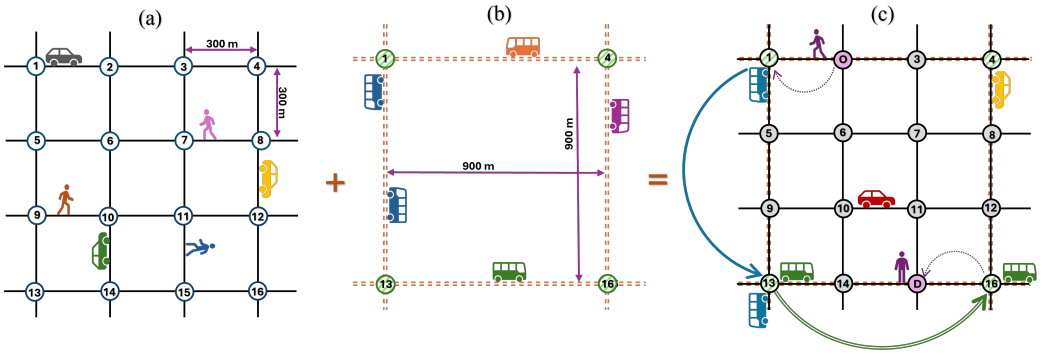



Figure 1: A portion of network layout, a) car/pedestrian network, b) public transport network, c) integrated network.

A single-peaked demand profile is developed to simulate the number of travellers entering the network during the peak hour. To capture traveller heterogeneity, each individual is randomly assigned an origin–destination pair, a departure time, and a Value of Time (VoT). For each traveller, the shortest path is computed for all available modes using Dijkstra's algorithm (Dijkstra, 1959), serving as the basis for mode-specific travel length estimation. To capture operational heterogeneity in the bus system, random headway values are assigned across all services, and each individual bus is given a randomised departure time upon reaching its scheduled headway. Additionally, a fixed dwelling time is incorporated at each bus station to reflect real-world boarding and alighting processes. Buses operate with fixed passenger capacities, and travellers are unable to board if a bus is full.

To estimate the average speed of motorised traffic across the network, this study adopts the 3D-MFD framework proposed by Geroliminis et al. (2014), which accounts for the interactions between cars and buses in a congested urban environment. Let $n_c$ and $n_b$ denote the accumulation of cars and buses, respectively. The total circulating flow $Q(n_c, n_b)$ in the network (veh / h) is given by the following exponential function:

$$Q(n_c, n_b) = a(n_c + n_b) \cdot e^{bn_c^2 + cn_b^2 + dn_c n_b + en_c + fn_b} \tag{1}$$

Where $a, b, c, d, e, f$ are model parameters calibrated to reflect the system's congestion sensitivity and mode interactions (Geroliminis et al., 2014). Given the total network length $L$, the average network speed is derived as $V(n_c, n_b) \approx Q(n_c, n_b) \cdot L/(n_b + n_c)$. Figure 2 illustrates the resulting

3D speed profile, showing how increasing accumulations of cars and buses degrade overall network performance due to congestion effects (Ampountolas et al., 2017).

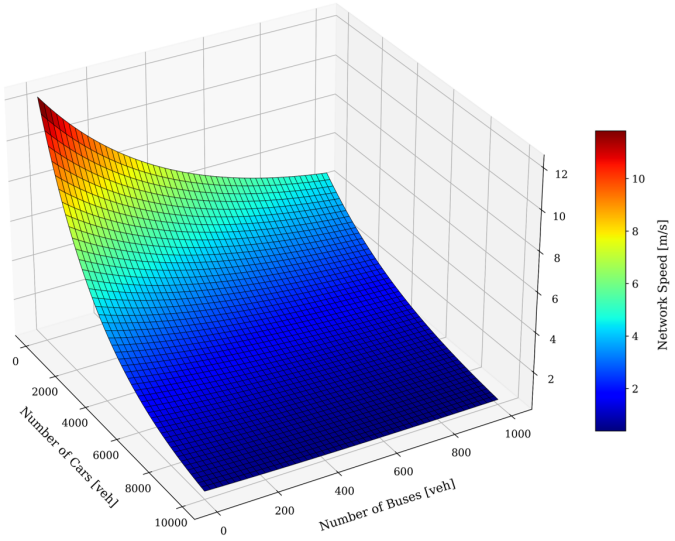



Figure 2: Average speed network profile.

For the walking mode, a constant speed is assumed based on empirical studies (Aghabayk et al., 2021), as pedestrians travel along dedicated footpaths and are assumed not to interact with vehicular traffic.

To simulate network-level traffic evolution, this study adopts a trip-based simulation framework (Jin, 2020). In this framework, every entry into and exit from the network is treated as an individual event. Travellers are sorted in ascending order based on their departure times and sequentially injected into the simulation. Remaining trip lengths are updated dynamically using trip-based macroscopic modelling principles; for further details, refer to Yildirimoglu et al. (2021).

Based on the trip-based simulation model (Mariotte et al., 2017), the travel time for car user $i$, denoted $T_i^{car}$, is computed as a function of their assigned trip length $l_i^{\text{car}}$, departure time $t_i$, and the network speed $V$, which is dynamically determined by the prevailing accumulation of cars and buses in the network. The network speed is assumed to be uniform across all users at any given moment, provided that vehicle accumulations remain unchanged during that period. Accordingly, the arrival time for user $i$ is calculated by discretising the travelled distance over time, accounting for changes in speed resulting from fluctuations in car and bus accumulations between the departure and arrival times:

$$l_i^{\text{car}} = \int_{t_i}^{t_i + T_i^{\text{car}}} V(n^{\text{car}}(t), n^{\text{bus}}(t))\, dt \tag{2}$$

Similarly, applying the same dynamic traffic evolution for buses as a separate module, the travel time for bus $j$, denoted $T_j^{bus}$ is determined as:

$$l_j^{\text{bus}} = \int_{t_j}^{t_j + T_j^{\text{bus}}} V(n^{\text{car}}(t), n^{\text{bus}}(t))\, dt \tag{3}$$

For the walking mode—whether used as a standalone option or in combination with public transport—the travel time $T_i^{walk}$ is computed assuming a constant walking speed:

$$V^{\text{walk}} = \frac{l_i^{\text{walk}}}{T_i^{\text{walk}}} \tag{4}$$

The simulation continues until all travellers have reached their respective destinations.

***Mode choice***

Since mode choice decisions are typically made at the start of each trip, this study evaluates the instantaneous travel time of each mode alongside the toll price at the time of departure. The monetary costs associated with each mode at time $t$ are defined as follows:

$$\begin{cases} C_i^{\text{car}} & = -VoT_i \cdot T_i^{\text{car}}(t_i) - \omega(t_i) \cdot l_i \\ C_i^{\text{walk}} & = -VoT_i \cdot T_i^{\text{walk}} \\ C_i^{\text{bus}} & = -VoT_i \cdot (T_i^{\text{bus}} + T_i^{\text{wait}} + T_i^{\text{transit}} + T_i^{\text{access}}) \end{cases} \tag{5}$$

where $VoT_i$ represents the Value of Time (VoT), $\omega$ denotes the toll price, $l_i$ indicates the traveller's trip length. For bus users, a uniformly random waiting time based on the bus headway is included. Additionally, if a trip requires a transfer to another bus line, a randomly assigned transition time is added. Moreover, users' access time to bus station is considered. This study also incorporates a multi-modal decision-making process. In multi-modal decision-making scenarios, three different alternatives arise, where the traveller's origin, destination node, or both are not aligned with bus station nodes. In these cases, the traveller walks to the nearest bus station, takes the bus, and then walks to the destination node. The total cost for these trips combines the monetary costs associated with walking and bus travel.

This study assumes that mode alternatives are independent and the unobserved cost term follows a Gumbel distribution. Therefore, the decision-making process is performed by a multinomial logit model. The probability of traveller $i$ choosing the car mode $Pr_i^{\text{car}}(t)$ is represented by the following equation:

$$Pr_i^{\text{car}}(t) = \frac{e^{\mu C_i^{\text{car}}}}{e^{\mu C_i^{\text{car}}} + e^{\mu C_i^{\text{bus}}} + e^{\mu C_i^{\text{walk}}}} \tag{6}$$

After calculating the probability of choosing each mode, a random utility theory is employed to assign modes to each traveller.

## 3 Methodology

This section presents a constrained multi-objective reinforcement learning framework that jointly considers network efficiency and equity in dynamic distance-based congestion pricing. The proposed agent seeks to maximise network outflow while limiting traveller cost, and explicitly constraining inequality through a Constrained Markov Decision Process (CMDP).

To quantify inequality, the study uses the Gini coefficient. The framework uses each traveller's experienced cost at arrival as a non-negative proxy for burden, normalised by trip length:

$$\bar{C}_i = \frac{\text{VoT}_i\, T_i + \omega(t_i)\, l_i}{l_i}, \tag{7}$$

where $\text{VoT}_i$ is the traveller's value of time, $T_i$ is travel time, $\omega(t_i)$ is the distance-based toll rate at departure time, and $l_i$ is trip length. This normalisation reduces the mechanical dependence of total cost on distance and instead measures cost intensity.

The optimisation problem is then defined as:

$$\max_{\pi} \quad \alpha\, \mathbb{E}[\text{Outflow}] - \beta\, \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \bar{C}_i\right] \tag{8}$$

subject to

$$\mathbb{E}[Gini(\bar{C})] \le \epsilon, \tag{9}$$

where $\pi$ is the pricing policy, $\alpha$ and $\beta$ are objective weights, and $\epsilon$ is the allowable inequality threshold. The constrained problem is solved using a Lagrangian relaxation:

$$L(\lambda, \theta) = J_R(\theta) - \lambda\big(J_C(\theta) - \epsilon\big), \tag{10}$$

where $\theta$ denotes the policy parameters, $J_R$ is the expected cumulative reward, $J_C$ is the expected cumulative inequality cost, and $\lambda \ge 0$ is the Lagrange multiplier. A DDPG agent is adopted because its continuous action space allows flexible toll adjustments and its actor-critic structure is well suited to constrained learning.

At each control step, the reward combines efficiency and traveller cost:

$$r_t = \alpha\, r_t^o - \beta\, r_t^{\bar{c}}, \tag{11}$$

where $r_t^o$ is the scaled network outflow term and $r_t^{\bar{c}}$ is the scaled mean normalised cost of travellers completing their trips at step $t$. Since inequality can only be measured at the end of an episode, the violation is first computed as

$$v_k = \kappa \max\{0,\ Gini(\bar{C}_k) - \bar{\epsilon}\}. \tag{12}$$

This terminal penalty is then redistributed uniformly across all steps so that its total discounted contribution matches the original episodic violation:

$$c = \frac{v_k}{Z_g}, \qquad Z_g = \sum_{n=0}^{T-1} \gamma^n. \tag{13}$$

Thus, the step-wise constraint signal remains consistent across episodes of different lengths.
Finally, transitions are first stored in an episodic buffer, augmented with the redistributed equity penalty after episode termination, and then transferred to replay memory for training. The Lagrange multiplier is updated according to the observed violation:

$$\lambda_{k+1} = \min\left\{\lambda_{\max}, \max\left\{0, \lambda_k + \lambda_{\mathrm{lr}} \kappa v_k\right\}\right\}. \tag{14}$$

This framework therefore provides a practical way to embed equity directly into simulation-based, dynamic, real-time, network-scale congestion pricing.

## 4 Results and discussion

The evolution of the agent measured by the total travel time of all travellers, is depicted in Figure 3. Figure 3.a, the blue line represents the Total Travel Time (TTT) for each episode, while the orange line shows the moving average over every 10 episodes. Figure 3.b illustrates the evolution of the Lagrange multiplier and the Gini index for the equity-aware agent (Figure 3.a). The multiplier increases and stabilises as convergence is approached, while the Gini index decreases until the inequality tolerance threshold is satisfied ($\epsilon \leq 0.2$). Considering the TTT under the NC scenario is approximately $2.5 \times 10^7$ seconds, it is reduced to around $1.25 \times 10^7$ seconds by the end of training, representing roughly a 50% reduction.

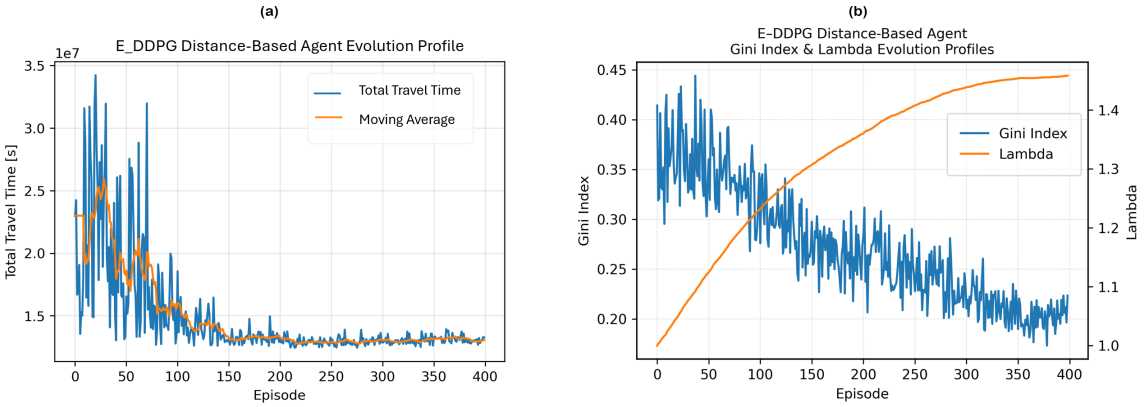



Figure 3: Evolution of the agent based on total travel time.

To assess the fairness and equity of the pricing strategies, travellers are partitioned into sixteen groups based on the quartiles of VoT (used as a proxy for income) and trip length, respectively. For each group, the imposed toll, the welfare change under uniform distribution of total collected revenue are evaluated.
Figure 4 presents the distribution of average toll payments across traveller groups for the proposed agent, illustrated through a heat map. A clear differentiation across trip-length groups is observed under distance-based pricing by linking charges directly to the extent of network usage.
While distance-based pricing appears more usage-based and therefore intuitively fair in the sense that travellers pay proportionally to network utilisation, this characteristic alone does not guarantee equity between individuals. Equity should be assessed based on the distribution of welfare changes across heterogeneous traveller groups, encompassing both toll payers and non-payers within the network. For this purpose, Figure 5 presents the distribution of welfare changes across traveller groups under the assumption that the collected revenue is uniformly redistributed across all travellers.
As shown in Figure 5, the average welfare change is positive across all groups. This indicates that, on average, congestion pricing improves overall welfare across the population. It is important to note that the welfare differences presented in Figure 5 are not directly incorporated into the optimisation objective. Instead, the agent (i) minimises the average travel cost per kilometre

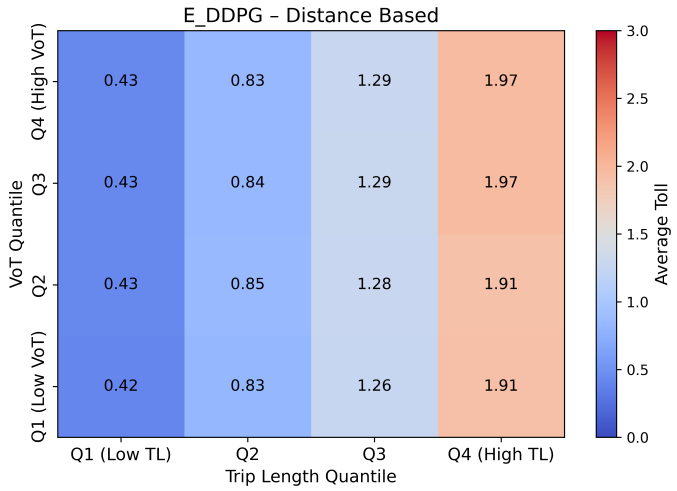



Figure 4: Average toll imposed ($\omega(t)$) on each travellers' groups.

and (ii) constrains the distribution of individual cost per kilometre by enforcing a Gini coefficient threshold below 0.2.

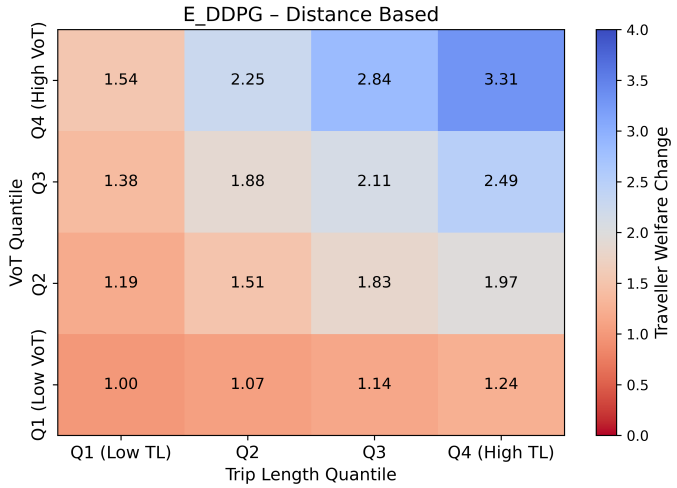



Figure 5: Average traveller welfare change ($\Delta W$) under uniform distribution of collected revenue per VoT $\times$ Trip Length.

Therefore, without reference to the no-control benchmark during training, the agent learns to generate an equitable pricing regime using only real-time system information. This represents a significant achievement and enhances the practical applicability of the framework. The resulting Gini index of cost per kilometre $\bar{C}$ is 0.197, confirming that the imposed equity constraint substantially reduces cost dispersion.

Interestingly, constraining inequality in cost per kilometre not only reduces cost dispersion, as expected, but also leads to a noticeable reduction in the dispersion of welfare changes across travellers. The corresponding Gini index of average welfare change under a uniform redistribution of total collected revenue is 0.199.

This outcome can be explained by the structural relationship between travel cost and welfare. Individual welfare changes are primarily determined by variations in cost components of control and no-control scenarios, including travel time and toll payments. By bounding inequality in marginal cost per kilometre, the optimisation indirectly regularises the distribution of these welfare determinants. In practice, the equity constraint limits extreme toll burdens while accounting for differences in travellers' trip lengths. As a result, time savings and cost adjustments are distributed more evenly across travellers, leading to lower welfare dispersion even though welfare is not directly constrained.

## 5 Conclusions

This study developed a congestion pricing framework based on reinforcement learning, which integrates constrained reinforcement learning with a trip-based 3D-MFD simulation environment. The framework was designed to maximise network outflow while explicitly limiting inequality in travellers' experienced costs. To support this objective, the simulation captures multimodal urban dynamics across cars, buses, and walking, while also accounting for traveller heterogeneity in departure time, trip length, origin–destination patterns, and value of time. The results show that the agent achieves substantial reductions in total travel time while maintaining comparable efficiency to unconstrained pricing approaches. More importantly, by embedding an explicit equity constraint within the optimisation process, the proposed method distributes toll burdens more evenly across travellers and produces more equitable outcomes. Overall, the findings demonstrate that the agent provides a practical and effective framework for real-time, network-scale, and equity-aware congestion pricing.

## References


Aghabayk, K., Parishad, N., & Shiwakoti, N. (2021). Investigation on the impact of walkways slope and pedestrians physical characteristics on pedestrians normal walking and jogging speeds. *Safety Science*, *133*, 105012.

Altman, E. (2021). *Constrained markov decision processes*. Routledge.

Ampountolas, K., Zheng, N., & Geroliminis, N. (2017). Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transportation Research Part B: Methodological*, *104*, 616–637.

Balzer, L., & Leclercq, L. (2022). Modal equilibrium of a tradable credit scheme with a trip-based mfd and logit-based decision-making. *Transportation Research Part C: Emerging Technologies*, *139*, 103642.

Basso, L. J., & Jara-Díaz, S. R. (2012). Integrating congestion pricing, transit subsidies and mode choice. *Transportation Research Part A: Policy and Practice*, *46*(6), 890–900.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. In *Edsger wybe dijkstra: his life, work, and legacy* (pp. 287–290).

Ecola, L., & Light, T. (2009). Equity and congestion pricing: A review of the evidence.

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, *42*(9), 759–770.

Geroliminis, N., Zheng, N., & Ampountolas, K. (2014). A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C: Emerging Technologies*, *42*, 168–181.

Jin, W.-L. (2020). Generalized bathtub model of network trip flows. *Transportation Research Part B: Methodological*, *136*, 138–157.

Li, Y., & Ramezani, M. (2022). Quasi revenue-neutral congestion pricing in cities: Crediting drivers to avoid city centers. *Transportation Research Part C: Emerging Technologies*, *145*, 103932.

Liu, R., Chen, S., Jiang, Y., Seshadri, R., Ben-Akiva, M., & Lima Azevedo, C. (2023). Managing network congestion with a trip-and area-based tradable credit scheme. *Transportmetrica B: Transport Dynamics*, *11*(1), 434–462.

Mariotte, G., Leclercq, L., & Laval, J. A. (2017). Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. *Transportation Research Part B: Methodological*, *101*, 245–267.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Yildirimoglu, M., Ramezani, M., & Amirgholy, M. (2021). Staggered work schedules for congestion mitigation: A morning commute problem. *Transportation Research Part C: Emerging Technologies*, *132*, 103391.

Zheng, N., Waraich, R. A., Axhausen, K. W., & Geroliminis, N. (2012). A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice*, *46*(8), 1291–1303.