Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 95 (2026) 25–32

Transportation Research Procedia

www.elsevier.com/locate/procedia

Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

Reinforcement Learning-Based Freeway Traffic Control Concerning Emissions

Sadullah Goncu[a,c,*], Mehmet Ali Silgu[d,e], Hilmi Berk Celikoglu[b,c]

[a] *Fatih Sultan Mehmet University, Department of Civil Engineering, Istanbul, 34445, Turkey*
[b] *Technical University of Istanbul, Department of Civil Engineering, Istanbul 34469, Turkey*
[c] *Technical University of Istanbul (ITU), ITU ITS Research Lab, Istanbul 34469, Turkey*
[d] *Bartın University, Department of Civil Engineering, Bartın, 74100, Turkey*
[e] *Koc University, Department of Industrial Engineering, Istanbul, 34450, Turkey*

Abstract

This study presents a reinforcement learning based framework involving the integrated use of ramp metering (RM) and variable speed limit (VSL) control towards the ultimate aim of mitigating traffic congestion and emissions. Traditional freeway traffic control strategies often fail to adapt dynamically to evolving traffic conditions, resulting in suboptimal performance. The proposed framework seeks, through simulation, the optimal setting of VSL and RM actions by leveraging RL. The learning-based architecture we have designed is trained and tested using data from a hypothetical freeway network piece and synthetic demand profiles. The performance of the framework is evaluated by considering multiple traffic demand levels and connected and automated vehicle penetration rates.

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025.

Keywords: Reinforced Learning; Freeway Traffic Control; Ramp Metering; Variable Speed Limiting

1. Introduction

Traffic congestion and vehicular exhaust remain two of the foremost challenges for modern freeway operations. Classical countermeasures such as ramp metering (RM) and variable speed limits (VSL) can mitigate congestion and enhance safety, yet their rule-based logic typically reacts slowly to rapidly evolving demand, which results in queue growth, stop-and-go shock-waves, and increases in CO_2 and NO_x emissions (Han et al., 2025).

Recent advances in reinforcement learning (RL) point to a more adaptive alternative. Agents that interact continuously with the traffic environment have reduced travel time by up to 22 % through Q-learning VSL (Zhou et al., 2020;

* Corresponding author. Tel.: +90 212 521 81 00
E-mail address: sgoncu@fsm.edu.tr, sadullahgoncu@gmail.com

2352-1465
Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025.
10.1016/j.trpro.2026.02.004

Wang et al., 2019) and have matched expert ALINEA performance on I-80W with multi-agent RMs (Belletti et al., 2018). Surveys of deep-RL in transport (Haydari and Yılmaz, 2022; Han et al., 2023) concur that such agents, unencumbered by explicit fundamental-diagram assumptions, can internalize nonlinear dynamics and refine their policies online.

Despite these advancements, two critical issues remain unresolved. Only about 7 % of published RL freeway-control studies coordinate more than one actuator, and fewer than 10 % incorporate emissions into the optimization objective (Han et al., 2023). Treating RM or VSL in isolation forfeits potential synergies, while neglecting environmental metrics undermines sustainability goals, concerns that are magnified under high-demand, mixed-traffic conditions with the rising penetration of connected and automated vehicles (CAVs).

To address these gaps, this study introduces an integrated RM + VSL controller, formulated as a dual-objective process and solved with proximal policy optimization (PPO). The agent's rewards are throughput, density stabilization, and environmental cost, where the latter is captured in two complementary ways. One agent variant derives real-time emissions directly from the HBEFA model (Notter et al., 2019), where the second agent relies on proxy indicators, such as speed fluctuations, acceleration, and fuel consumption, which are easier to obtain in field deployment. For comparison, a third agent is trained with a purely traffic flow performance-centric reward (throughput and travel time). Robustness of all three agents is examined across four demand tiers and four CAV penetration rates within a microscopic SUMO test bed. The contributions are twofold. First, it casts the integrated RM and VSL problem as a dual-objective process that simultaneously maximizes throughput and minimizes a vector of pollutant emissions. It solves the problem using a centralized-training, decentralized-execution approach through a PPO architecture, constituting the first demonstration of such an emission-aware RL controller for freeway traffic. Second, it delivers the first systematic evaluation of the practical value of the emission term by comparing three reward designs, traffic-only, direct HBEFA feedback, and a purely kinematic proxy, across four demand tiers and five CAV penetration levels. The results show that the environmental objective provides measurable benefits and reveal that the proxy signal achieves almost the full emission reduction of direct exhaust sensing with minimal extra tuning effort.

The remainder of the paper is organized as follows: Section 2 reviews related work and positions the present study; Section 3 details the proposed RL framework, describes the SUMO test bed and experimental design; Section 4 presents and discusses numerical results; and Section 5 concludes with implementation insights and directions for future research.

2. Literature Review

A growing strand of research now seeks to embed emission-reduction objectives directly inside RL–based freeway controllers. Tang et al. (2021) showed that adding CO_2 and NO_x penalties to a DQN–VSL agent reduced those pollutants by roughly 10 % while preserving throughput, illustrating the feasibility of dual-objective learning. In microscopic simulators such as SUMO, edge-level emission values are readily available. However, outside of simulation, real-time emissions must be inferred from surrogates, such as speed, acceleration, or fuel-flow estimates (Li et al., 2023). Reliable online sensing, therefore, remains a practical challenge.

The earliest RL freeway studies concentrated on a single control actuator. Tabular Q-learning stabilized local ramp queues in Schmidt-Dumont (2018) work, whereas a corridor-wide VSL agent in Zhou et al. (2020) maintained densities near the critical regime without recourse to a fundamental diagram. Deep-RL studies, including distributed Q-learning (DQL) (Wang et al., 2019), deep actor–critic (Wu et al., 2020), and physics-informed RL (Han et al., 2022), reported travel-time gains of 10–22 % on corridors such as I-880 (California, USA), A16, and A13–L (Paris, France). Yet these studies continued to optimize either RM or VSL in isolation and treated emissions as post-analysis indicators rather than explicit objectives. A review of notable studies on RL-based freeway traffic control studies is presented in Table 1.

Driven by climate policy mandates, several researchers have started to penalize emissions in the reward signal (Xu et al., 2020). Nevertheless, the recent meta-analysis by Han et al. (2023) finds that fewer than 10 % of 164 RL traffic-control papers include any pollutant term in the reward, and almost none optimize the full pollutant vector (CO, HC, NO_x, PM).

A second, equally important gap concerns actuator coordination. Only about 7 % of reviewed studies couple RM with VSL, despite decades of evidence that their functional overlap calls for joint optimization. The multi-agent

Table 1: Representative studies on RL-based integrated freeway traffic-control systems, Abbreviations: RM – ramp metering; VSL – variable speed limit; LCC – lane-change control; Fwy – freeway; Urb – urban; R – real; H – hypothetical; S – synthetic; HDV – human-driven vehicle; CAV – connected and automated vehicle; RL – reinforcement learning; Q-L – Q-learning; DRL – deep RL; DQN – deep Q-network; DDQN – double DQN; DQL – distributed Q-learning; AC – actor–critic; TD3 – twin-delayed DDPG; IPMATD3 – improved multi-agent TD3; I2A – imagination-augmented agent; MARL – multi-agent RL; MB-RL – model-based RL; P-RL – physics-informed RL; TL – transfer learning; MPC – model predictive control; Proxy – emissions via proxy variables; Direct – direct emissions in reward.

Reference	Net.	R/H	R/S	Strategy	Ctrl.	RL?	Emis.	Mix
Schmidt-Dumont (2018)	Fwy	H	S	RM+VSL	RL	Yes	–	HDV
Belletti et al. (2018)	Fwy	R	R	RM	DRL	Yes	–	HDV
Wang et al. (2019)	Fwy	R	R	VSL	DQL	Yes	–	HDV
Zhou et al. (2020)	Fwy	R	R	VSL	Q-L	Yes	–	HDV
Xu et al. (2020)	Urb	R	R	VSL	DQN	Yes	Direct	HDV
Wu et al. (2020)	Fwy	R	R	VSL	DRL	Yes	Direct	HDV
Ke et al. (2021)	Fwy	R	S	VSL	DDQN+TL	Yes	–	HDV
Han et al. (2022)	Fwy	H	S	VSL	P-RL	Yes	–	HDV
Wang et al. (2022)	Fwy	R	R	RM+VSL	Deep-AC	Yes	–	HDV
Pan et al. (2021)	Fwy	H	S	RM+VSL+LCC	MB-RL	Yes	Proxy	HDV+CAV
Lu et al. (2023)	Fwy	R	R	VSL	TD3	Yes	–	HDV+CAV
Li and Lasenby (2023)	Fwy	H	S	VSL	I2A	Yes	–	HDV
Han et al. (2024)	Fwy	H	S	VSL	MARL	Yes	–	HDV+CAV
Sun et al. (2024)	Fwy	H	S	RM+VSL	MPC+DRL	Yes	–	HDV
Han et al. (2025)	Fwy	R	R	RM+VSL	IPMATD3	Yes	Direct	HDV+CAV

ALINEA study of Belletti et al. (2018) omitted VSL entirely, whereas the hybrid MB-RL framework of Pan et al. (2021) combined RM, VSL, and lane-change control in a hypothetical CAV-heavy scenario but ignored emissions. The recent IPMATD3 work by Han et al. (2025) is one of the few to integrate RM and VSL with direct emission terms; however, its evaluation is limited to a single demand tier and predominantly human-driven traffic, leaving the robustness in mixed traffic unresolved.

Two large-scale surveys—Haydari and Yılmaz (2022), which reviews deep-RL applications across intelligent-transportation systems, and Han et al. (2023), which focuses on dynamic freeway control—converge on three persistent gaps in the literature: first, only 7 % of RL freeway-control papers coordinate more than one actuator, indicating that integrated schemes remain the exception rather than the rule; second, fewer than 10 % incorporate any emission metric in the reward function, and an even smaller subset treats emissions as a primary optimization objective; and third, simulation-to-field transferability is still largely unexplored because exhaustive on-line exploration is costly and behavioral realism uncertain.

Very recently, Han et al. (2025) proposed an IPMATD3-based integrated-traffic–control scheme that jointly optimizes RM + VSL along an 8 km corridor with three recurrent bottlenecks in mixed HDV–CAV flow (Han et al., 2025). While their multi-agent TD3 with prioritized replay achieves improvements over feedback and single-lever RL controllers, three limitations remain. First, the hybrid reward relies on direct exhaust-emission readings from the SUMO–HBEFA interface, a signal that is unavailable in field deployments. Moreover, no proxy-based alternative is examined. Second, training and evaluation are confined to a single calibrated evening-peak demand trace, so the robustness across different demand regimes remains unknown. Third, a fixed weight vector is adopted for the six reward components, precluding any analysis of throughput–emission trade-offs or sample efficiency under different objective formulations. The present study addresses these gaps by (i) training two emission-aware agents, one with direct HBEFA feedback and one using proxy indicators, and a throughput-only baseline; (ii) testing all agents across four demand tiers and four CAV penetration rates; and (iii) employing a multi-objective PPO that enables systematic weight-sensitivity analyses.

3. Experimental design

This section introduces our test network, simulation environment, agent definitions, test scenarios, training details of the agents, and performance metrics. All experiments are carried out in SUMO, accessed through the Python TraCI API (Lopez et al., 2018). The test network is a hypothetical 4 km, three-lane freeway corridor divided into four 1 km segments (Fig. 1). Each segment has a single on-ramp equipped with a ramp-metering signal, and the first segment in the network is the VSL zone whose advisory speeds are displayed on variable-message signs (VMS). HDVs follow the IDM Treiber et al. (2000), while CAVs' longitudinal driving behavior is modeled through the CACC car-following model (Milanés and Shladover, 2014). The RM and VSL control actions are updated every 15s. The simulation du-

rations are set to 10800s with a 0.1s timestep. The Proximal Policy Optimization (PPO) algorithm (Schulman et al., 2017) is employed to train the RL agents, chosen for its sampling efficiency and stability in continuous control problems. The RL agent's action space consists of discrete VSL values (ranging from 50 to 120 km/h) and ramp metering rates (5 to 15 seconds red time). The state space includes lane occupancy, queue lengths, and emission levels (depending on the RL-agent). Testing is performed by utilizing synthetic demand profiles and varying the penetration rates of CAVs to evaluate the generalization ability of the trained agent. Training demands sampled uniformly from three demand tiers—Low (7 200 / 3 600 veh), Medium (12 960 / 4 320 veh), and High (16 200 / 5 400 veh) for mainline/on-ramp flows—at 0 % CAV penetration. Three PPO variants are evaluated. **$\mathbf{PPO_{TTT}}$** maximizes mainline throughput while minimizing both total travel time and density. **$\mathbf{PPO_{DIR}}$** maximizes mainline throughput and minimizes total travel time and density, and additionally minimizes the sum of direct HBEFA pollutants (CO_2, CO, HC, NO_x).**$\mathbf{PPO_{PRX}}$** maximizes mainline throughput and minimizes total travel time and density, and minimizes the emissions through proxy indicators consisting of speed variance, acceleration variance, and instantaneous fuel rate.

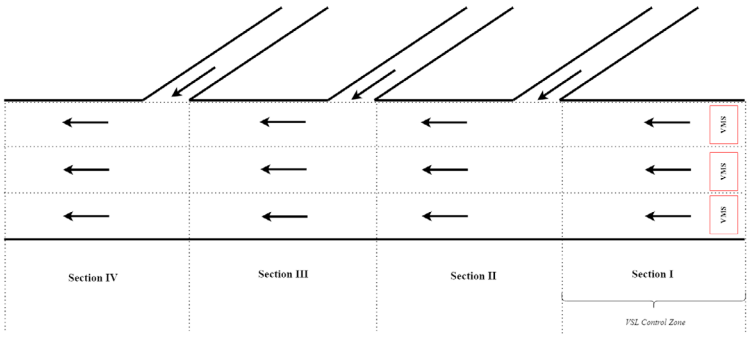

Fig. 1: Four-segment (4 km) test corridor with one on-ramp per segment and a variable-speed-limit zone in Section I

Let v, a, and f denote the instantaneous speed, longitudinal acceleration, and fuel rate of a vehicle in the VSL zone. Each proxy is normalized by the episode-wide 5th–95th percentile range to obtain dimensionless values $\hat{\sigma}_v = (\sigma_v - \sigma_v^{5\%})/(\sigma_v^{95\%} - \sigma_v^{5\%})$, $\hat{\sigma}_a$ and $\hat{f}$ analogously. The composite proxy term in the reward is

$$r_{\text{proxy}} = -(w_v \hat{\sigma}_v + w_a \hat{\sigma}_a + w_f \hat{f}) \tag{1}$$

with weights $w_v = 0.4$, $w_a = 0.3$, $w_f = 0.3$. Weights were selected by a three-point grid-search to minimize the mean absolute error between r_{proxy} and direct HBEFA CO_2 output on 50 validation episodes. This normalization keeps each sub-term in $[0, 1]$ and prevents the high variance of fuel-rate from dominating the reward.

After the training of the agents, five control scenarios have been defined for testing purposes, which are no control, ALINEA RM (Papageorgiou et al., 1997), and one for each of the RL-agents. Additionally, the demand level for testing is higher than the high-tier level in the training procedure (21,600 / 5,400 veh). For each control scenario, the market penetration rate (MPR) of CAVs has been increased from 0-100% with 25% increments. To account for stochasticity, the same seed has been used in testing scenarios. Furthermore, to show the robustness of the RL-agents, additional experiments have been conducted, where we utilize the medium-tier level from the training procedure, under five MPR ranges. For PPO_{DIR}, an additional one-shot sensitivity test has been conducted where analysis varies the pollutant weight $\lambda_{\text{emis}} \in \{0.25, 0.50, 0.75, 1.00\}$ at high demand and 50 % MPR, thereby illustrating the throughput–emission trade-off. As for the performance metrics, throughput, total travel times (TTT), queue lengths at on-ramps, and emission results have been comparatively evaluated across controllers.

4. Results and Discussion

This section presents the traffic flow performance evaluations, robustness results, and sensitivity analysis findings.

4.1. Traffic Flow Performance Evaluations

Figure 2 presents the total travel time (TTT, left plot) and total throughput (right plot) as functions of the MPR. Results refer to the high-demand test and the five control scenarios defined in Section 3. Across all controllers, throughput rises with MPR. The additional CAV capacity outweighs the minor efficiency loss caused by platoon build-up in the merge areas. The two emission-aware agents deliver the highest flow-rates, PPO_{DIR} outperforms the no-control case by 6–7 % at 50–75 % MPR, while PPO_{PRX} maintains the lead at full penetration. The (PPO_{TTT}) agent results in marginal throughput gain (< 0.5% at 100 % MPR) for noticeably higher congestion in the merge sections. ALINEA increases throughput relative to No-control only beyond 50 % MPR, reflecting its inability to smooth mainline speed without coordinated VSL. TTT results are complementary to the throughput results pattern. For MPR up to 50 % the RL controllers reduce travel time by 3–5 % with respect to the baseline by preventing queue spillback from the on-ramps. At 75–100 % penetration, the system approaches capacity, and TTT increases for every controller. PPO_{PRX} retains a slight advantage (3 %) over No-control, whereas PPO_{TTT} incurs the largest penalty because its higher flow is realized at the cost of denser traffic in Segment 4. ALINEA performs best at the low MPRs (i.e., 0 % MPR) but is outperformed by the integrated RL controllers once CAV penetration exceeds 25 %. Although ALINEA achieves the shortest TTT at 0% MPR, its advantage is confined to a single operating point and stems from aggressive queue-holding without any mainline speed smoothing.The RL controllers, in contrast, offer one policy for the entire 0–100 % MPR range and still match ALINEA's emissions while delivering markedly higher throughput as soon as CAVs exceed 25 %MPR. In summary, the integrated RM+VSL agents improve corridor throughput over the entire penetration range and do so without degrading travel time at low–medium MPR. Figure 3 presents the emission and fuel consumption results for the high-demand experiment. The left plot aggregates the five pollutants (CO_2, CO, HC, NO_x and $PM_{2.5}$) into a single metric in terms of per vehicle, while the right plot reports total fuel consumption.

At zero MPR the emission-aware agents already yield a tangible improvement. PPO_{DIR} reduces the composite emission index by approximately 15% relative to the no-control baseline, closely followed by PPO_{PRX}. The PPO_{TTT} agent offers a minor reduction (12%), reflecting its indirect influence on speed variance. ALINEA attains the lowest emissions and fuel consumption at this penetration, but at the cost of much lower throughput compared to other agents.

As MPR increases, emissions per vehicle rise for every controller, which is an expected outcome of the higher average speeds obtained with larger CAV MPR, yet the controller ranking remains largely intact. Up to 75% MPR, both emission-aware agents track one another and maintain a gap of 8–10% to PPO_{TTT}. At full penetration, that gap narrows where the direct-feedback agent and its proxy-based counterpart converge to virtually identical values, indicating that the proxy bundle captures the essential dynamics once vehicle heterogeneity is removed. ALINEA's edge diminishes beyond 50% MPR because the absence of VSL allows higher mainline speeds and hence larger fuel use. Nevertheless, it remains the lowest emission option at 100% penetration, about 18% below the no-control case on both indicators, while having a lower throughput. Fuel-consumption trends mirror the pollutant curves. The two emission-aware agents consume about 6% less fuel than PPO_{TTT} at medium penetration and 3% less at full penetration. The residual difference between PPO_{DIR} and PPO_{PRX} never exceeds 1%, confirming that the proxy method can approximate direct HBEFA measurements to some degree. In summary, incorporating emissions in the reward, whether through direct HBEFA output or proxy indicators, yields consistent environmental gains without sacrificing the mobility benefits. The proxy design approaches the direct-feedback performance across all penetration levels, supporting its use in field implementations where real-time exhaust sensing is not yet feasible.

4.2. Robustness and Sensitivity Results

Table 2 reports the medium-demand results (12 960/4 320 veh) for the five control scenarios over the MPR range. The simulation duration for robustness simulations is 3600s. Total throughput, TTT, and the emission results are presented for each MPR. Despite the substantial drop in flow relative to the high-demand^{+} experiments discussed in Section 4.1, the ranking of controllers is preserved.

The medium-demand robustness analysis confirms that the relative behaviour of all controllers is robust to a 20 % reduction in traffic volume. Both emission-aware agents, PPO_{DIR} and PPO_{PRX}, preserve the performance lead they held under high demand. Their travel-time and emission advantages over ALINEA remain within the same order of magnitude, and no reversal in ranking is observed at any MPR. Even the traffic-centered PPO_{TTT} sustains its pattern, albeit at the cost of higher emissions, indicating that its policy does not over-fit to the congestion dynamics

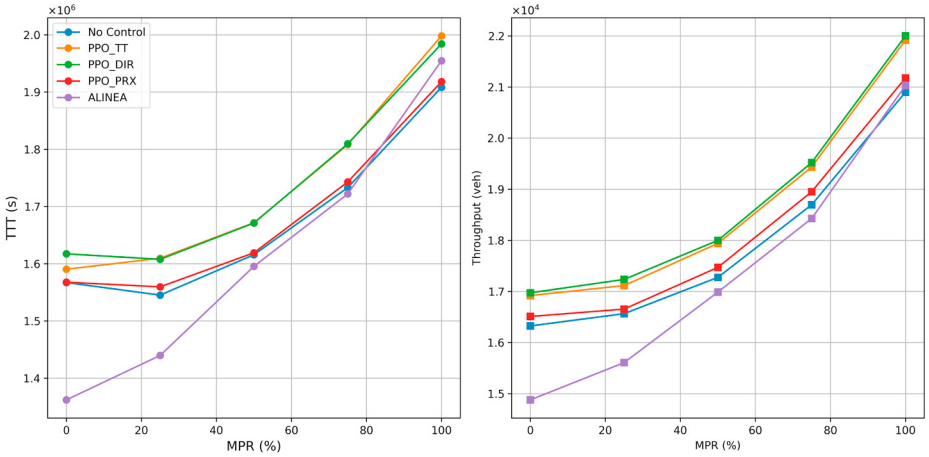

Fig. 2: Effect of MPR on (a) TTT and (b) corridor throughput under the High-demand scenario (21 600 veh/3h mainline, 5 400 veh/3h on-ramps).

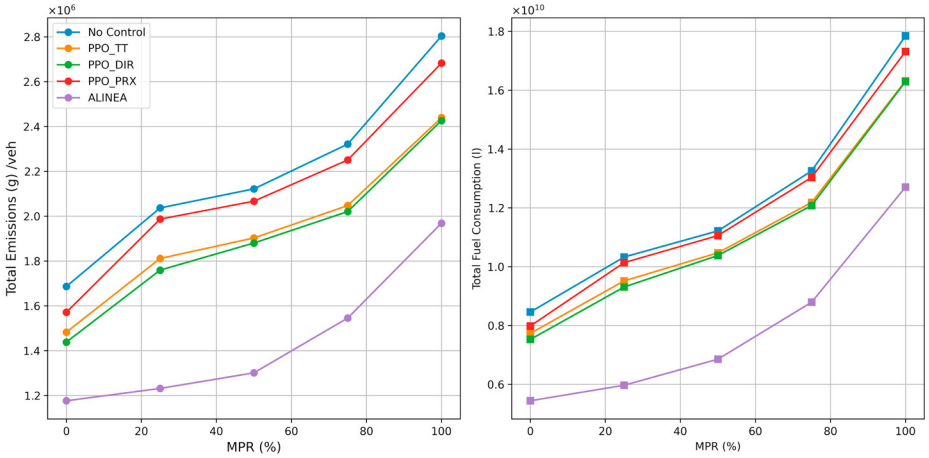

Fig. 3: Environmental performance for the same demand: (a) composite pollutant mass per vehicle; (b) total fuel consumption.

specific to the training high tier. ALINEA continues to excel only at the extreme case of 0 % CAVs, once a modest 25 % penetration is introduced, the integrated RM + VSL controllers again outperform, showing that their learned coordination strategy generalizes to lighter demand without parameter retuning. Figure 4 visualizes the results of the sensitivity analysis of the PPO_{DIR} controller for four reduced pollutant weights (λ = 0.25, 0.50, 0.75, 1.00) under the benchmark scenario (medium-demand, 50 % MPR). With respect to the traffic-only baseline ($\lambda = 0$), a modest weight of $\lambda = 0.25$ reduces the total emissions per vehicle by 2.8 % and lowers TTT by 5.0 %, at a throughput loss below 1 %. Raising the weight to $\lambda = 0.50$ produces a further 3.5 % emission reductions and a 10 % drop in TTT, while throughput falls by only about 2%. Beyond that point the frontier flattens, where $\lambda = 0.75$ yields no additional environmental benefit and slightly increases fuel consumption, whereas $\lambda = 1.00$ achieves the largest emission reduction (11.6 %) but delivers only a marginal improvement over $\lambda = 0.50$ while incurring a 3 % throughput penalty. The Pareto front, therefore, lies near $\lambda \approx 0.5$, exactly where the PPO_{PRX} controller results in the previous section, indicating that proxy indicators capture almost the full environmental benefit without online exhaust measurements.

Table 2: Robustness analysis results

MPR	Controller	Throughput [veh]	TTT [s]	Total Em/veh [g]	Fuel [mL]
0	PPO_{TTT}	5 383	2.44×10^6	1.39×10^6	2.32×10^9
25	PPO_{TTT}	5 428	2.51×10^6	1.51×10^6	2.53×10^9
50	PPO_{TTT}	5 719	2.59×10^6	1.48×10^6	2.61×10^9
75	PPO_{TTT}	6 183	2.66×10^6	1.40×10^6	2.67×10^9
100	PPO_{TTT}	6 911	2.04×10^6	1.18×10^6	2.53×10^9
0	PPO_{DIR}	5 598	2.45×10^6	1.72×10^6	2.34×10^9
25	PPO_{DIR}	5 808	2.53×10^6	1.82×10^6	2.53×10^9
50	PPO_{DIR}	5 891	2.60×10^6	1.44×10^6	2.62×10^9
75	PPO_{DIR}	6 245	2.67×10^6	1.39×10^6	2.68×10^9
100	PPO_{DIR}	6 980	2.05×10^6	1.17×10^6	2.53×10^9
0	PPO_{PRX}	5 291	2.60×10^6	1.48×10^6	2.42×10^9
25	PPO_{PRX}	5 330	2.72×10^6	1.70×10^6	2.78×10^9
50	PPO_{PRX}	5 619	2.88×10^6	1.65×10^6	2.86×10^9
75	PPO_{PRX}	6 083	3.03×10^6	1.59×10^6	2.98×10^9
100	PPO_{PRX}	6 761	2.50×10^6	1.36×10^6	2.84×10^9
0	AL	5 072	1.32×10^6	1.15×10^6	1.81×10^9
25	AL	5 128	1.32×10^6	1.19×10^6	1.90×10^9
50	AL	5 266	1.32×10^6	1.23×10^6	2.01×10^9
75	AL	5 473	1.33×10^6	1.29×10^6	2.19×10^9
100	AL	5 775	1.36×10^6	1.45×10^6	2.59×10^9

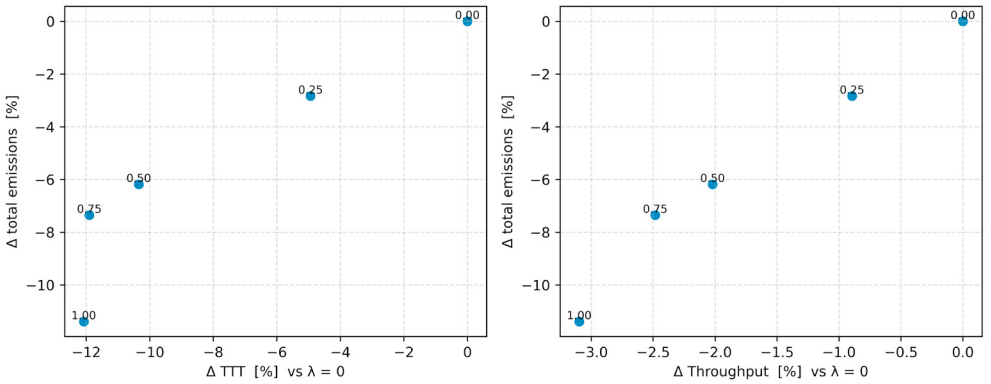

Fig. 4: Sensitivity of the PPO_{DIR} to the emission reward weight λ (medium-demand, 50 % MPR). Left: relative change in total travel time versus relative change in composite emissions, both normalized to the traffic-only baseline ($\lambda = 0$). Right: relative change in corridor throughput versus the same emission metric. Each marker is labeled by its λ value.

5. Conclusion

This paper introduced an RL framework that coordinates RM and VSL while incorporating multi-pollutant emissions into the control objective. Experiments on a microscopic 4 km test corridor showed that the two emission-aware agents, one relying on direct HBEFA feedback, the other on low-cost kinematic proxies, consistently raised throughput by about 10% relative to the ALINEA scenario at high demand and high CAV MPR, yet reduced the emission/veh by up to 20% and held total travel time within 3% of ALINEA. Moreover, at 20% demand, the ranking of controllers and the size of the benefits remained virtually unchanged, indicating that the learned policies generalize beyond the training regime. A one-shot sensitivity analysis further revealed a clear Pareto front near a pollutant weight of 0.5, where an additional 10% emission reduction is obtained for only a modest 2-3% loss in throughput. In a synthetic corridor tuned to urban-motorway density and spill-back patterns, domain-randomized PPO agents (±20% car-following parameters, 3% detector noise, ±15% demand) learn from relative, aggregated signals and jointly reduce congestion and emissions. Both the direct-emission and proxy variants approach the traffic optimum, with proxies offering the more deployable path. Future work will validate transferability in hardware-in-the-loop tests, retrain with real detector traces, add adaptive weight scheduling, and extend control to lane-change and queue-override coordination, advancing scalable, emission-aware RL for freeway management.

References

Belletti, F., Haziza, D., Gomes, G., Bayen, A.M., 2018. Expert level control of ramp metering based on multi-task deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems 19, 1198–1207. doi:10.1109/TITS.2017.2725912.

Han, L., Zhang, L., Guo, W., 2024. Multi-agent deep reinforcement learning for multi-lane freeways: Differential variable speed limit control in mixed traffic environment. Transportation Research Record doi:10.1177/03611981241230524.

Han, L., Zhang, L., Pan, H., 2025. Improved multi-agent deep reinforcement learning-based integrated control for mixed traffic flow in a freeway corridor with multiple bottlenecks. Transportation Research Part C: Emerging Technologies 174, 105077. doi:10.1016/j.trc.2025.105077.

Han, Y., Hegyi, A., Zhang, L., He, Z., Chung, E., Liu, P., 2022. A new reinforcement learning-based variable speed limit control approach to improve traffic efficiency against freeway jam waves. Transportation Research Part C 144, 103900. doi:10.1016/j.trc.2022.103900.

Han, Y., Wang, M., Leclercq, L., 2023. Leveraging reinforcement learning for dynamic traffic control: A survey and challenges for field implementation. Communications in Transportation Research 3, 100104. doi:10.1016/j.commtr.2023.100104.

Haydari, A., Yılmaz, Y., 2022. Deep reinforcement learning for intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems 23, 11–32. doi:10.1109/TITS.2020.3008612.

Ke, Z., Li, Z., Cao, Z., Liu, P., 2021. Enhancing transferability of deep reinforcement learning-based variable speed limit control using transfer learning. IEEE Transactions on Intelligent Transportation Systems 22, 4684–4696. doi:10.1109/TITS.2020.2990598.

Li, D., Lasenby, J., 2023. Imagination-augmented reinforcement learning framework for variable speed limit control. IEEE Transactions on Intelligent Transportation Systems 25, 1384–1393.

Li, L., Zhu, R., Wu, S., Ding, W., Xu, M., Lu, J., 2023. Adaptive multi-agent deep mixed reinforcement learning for traffic light control. IEEE Transactions on Vehicular Technology 73, 1803–1816.

Lopez, P.A., Wiessner, E., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flotterod, Y.P., Hilbrich, R., Lucken, L., Rummel, J., Wagner, P., 2018. Microscopic traffic simulation using sumo, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2575–2582. doi:10.1109/ITSC.2018.8569938.

Lu, W., Yi, Z., Gu, Y., Rui, Y., Ran, B., 2023. Td3lvsl: A lane-level variable speed limit approach based on twin delayed deep deterministic policy gradient in a connected automated vehicle environment. Transportation Research Part C 153, 104221. doi:10.1016/j.trc.2023.104221.

Milanés, V., Shladover, S.E., 2014. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. Transportation Research Part C: Emerging Technologies 48, 285–300.

Notter, B., Keller, M., Cox, B., 2019. Handbook emission factors for road transport 4.1. Quick Ref. Bern, Germany 28.

Pan, T., Guo, R., Lam, W.H.K., Zhong, R., Wang, W., He, B., 2021. Integrated optimal control strategies for freeway traffic mixed with connected automated vehicles: A model-based reinforcement learning approach. Transportation Research Part C 123, 102987. doi:10.1016/j.trc.2021.102987.

Papageorgiou, M., Hadj-Salem, H., Middelham, F., 1997. Alinea local ramp metering: Summary of field results. Transportation research record 1603, 90–98.

Schmidt-Dumont, T., 2018. Reinforcement learning for the control of traffic flow on highways. Ph.D. thesis. Stellenbosch University. URL: https://www.vuuren.co.za/Theses/Schmidt-DumontPhD.pdf.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 .

Sun, D., Jamshidnejad, A., Schutter, B.D., 2024. A novel framework combining mpc and deep reinforcement learning with application to freeway traffic control. IEEE Transactions on Intelligent Transportation Systems 25, 6756–6768. doi:10.1109/TITS.2023.3342651.

Tang, X., Chen, J., Liu, T., Qin, Y., Cao, D., 2021. Distributed deep reinforcement learning-based energy and emission management strategy for hybrid electric vehicles. IEEE Transactions on Vehicular Technology 70, 9922–9934.

Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Physical review E 62, 1805.

Wang, C., Xu, Y., Zhang, J., Ran, B., 2022. Integrated traffic control for freeway recurrent bottleneck based on deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems 23, 15522–15534. doi:10.1109/TITS.2022.3141730.

Wang, C., Zhang, J., Xu, L., Li, L., Ran, B., 2019. A new solution for freeway congestion: Cooperative speed limit control using distributed reinforcement learning. IEEE Access 7, 41947–41957.

Wu, Y., Tan, H., Qin, L., Ran, B., 2020. Differential variable speed limits control for freeway recurrent bottlenecks via deep actor-critic algorithm. Transportation Research Part C: Emerging Technologies 117, 102649. doi:10.1016/j.trc.2020.102649.

Xu, Z., Cao, Y., Kang, Y., Zhao, Z., 2020. Vehicle emission control on road with temporal traffic information using deep reinforcement learning, in: IFAC-PapersOnLine, pp. 14960–14965. doi:10.1016/j.ifacol.2020.12.1988.

Zhou, W., Yang, M., Lee, M., Zhang, L., 2020. Q-learning-based coordinated variable speed limit and hard shoulder running control strategy to reduce travel time at freeway corridor. Transportation Research Record 2674, 915–925.