SUMO User Conference 2026

TIBconferencesession

TIB-OP will set DOI with \TIBdoi

Published: 1970-01-01

Energy Optimized Green Light Assist in Varying Traffic Scenarios Using Reinforcement Learning

Johan L. Kolms[1], Kai M. Blum[1], Michael Thane[1], Tamás Kurczveil[1] & Dirk J. Lehmann

[1]Ostfalia University of Applied Science

Abstract: Optimizing traffic flow at signalized intersections is a central challenge in road networks. Existing approaches address this problem through measures such as adaptive traffic signal control to decrease travel time and reduce vehicle energy consumption. A complementary approach is the optimization of vehicle-level driving behaviour, which directly influence queue formation and stop-and-go dynamics. This paper evaluates deterministic and reinforcement learning approaches for recommending an approach speed towards a traffic light in a simple SUMO scenario. Using energy consumption and travel time as primary metrics, we compare both methods across traffic volumes. Baseline runs with SUMO's default driving behaviour were conducted as reference, followed by experiments with two, six, and 20 vehicles. We achieve a decrease of energy consumption by 22.83 over the distance travelled in a scenario with 20 vehicles while roughly maintaining the reference travel time (++1.73). The reinforcement learning agent demonstrated improved consistency and reduced deviation compared to the baseline runs without any external control, especially in high traffic volume scenarios.

Keywords: Traffic Signal Optimization, SUMO, GLOSA, Reinforcement Learning, Energy-Efficient Driving, Connected and Automated Vehicles

1 Introduction

Signalized intersections are major bottlenecks in road networks. Stop-and-go dynamics caused by red phases and queue formation increase travel time variability and lead to unnecessary energy consumption, especially for electric vehicles where frequent acceleration events dominate the consumption profile. While many approaches mitigate these effects via adaptive traffic signal control, a complementary direction is to influence vehicle behavior directly by recommending an approach speed that reduces unnecessary braking and improves phase utilization.

Speed advisory strategies require only limited intervention: instead of controlling the signal plan or assuming full vehicle autonomy, an advisory system can propose a target speed that aligns vehicle arrivals with green phases. Such recommendations can be communicated to connected vehicles via vehicle-to-infrastructure (V2I/V2X) messages or displayed to human drivers in driver assistance systems. However, the effectiveness of speed recommendations depends on traffic density, interactions with other vehicles, and variability in signal timing. In particular, deterministic planning methods such as manually programmed algorithms can perform well in isolated settings but may degrade under disturbances caused by heterogeneous traffic.

In this work, we evaluate vehicle-level speed recommendation strategies in microscopic traffic simulation using SUMO. We use a minimal but controlled scenario consisting of a single-lane, one-way road segment with one signalized intersection and randomized phase durations. Within this setting, we compare (i) a **deterministic speed planning baseline** that targets a predicted green phase and (ii) a **reinforcement learning (RL) agent** that learns adaptive speed recommendations from vehicle- and signal-level observations. Both approaches are evaluated against SUMO's default driving behavior across multiple traffic volumes (only the agent vehicle, two, six and 20 vehicles total), using energy consumption () and average speed () as primary metrics.

This paper investigates under which traffic conditions deterministic speed advisory strategies becomes insufficient, and when adaptive reinforcement learning–based approaches provide additional benefits in terms of energy efficiency while maintaining travel time.

The main contributions of this paper are:

- A reproducible SUMO scenario to benchmark speed advisory strategies at a signalized intersection under varying traffic volumes and changing signal timing.
- A deterministic phase-alignment algorithm for computing a target approach speed based on current signal state and estimated arrival time.
- A reinforcement learning formulation for adaptive speed recommendation that incorporates lead-vehicle interaction and queue proxy information.
- An empirical comparison showing that speed recommendations can reduce energy consumption with only minor travel time impact, with the RL agent providing the largest gains in high-traffic settings.

The remainder of this paper is structured as follows. *Section 2* reviews related work on signal control and vehicle-level speed optimization. *Section 3* describes the SUMO scenario, observation and action spaces, and reward design. *Section 4* reports experimental results, followed by a discussion of limitations and implications in Section 5.

2 Related Work

Research on improving performance at signalized intersections broadly follows two complementary directions: *infrastructure-level* optimization (e.g., adaptive signal control) and *vehicle-level* control or advisory (eco-driving, speed guidance). Since this paper targets vehicle behavior to mitigate stop-and-go dynamics without modifying signal plans, we focus on prior work on speed advisory, eco-approach and departure (EAD), and learning-based driving policies.

2.1 Speed advisory and eco-approach at signalized intersections

Green Light Optimal Speed Advisory (GLOSA) and related eco-approach and departure (EAD) methods use signal phase and timing (SPaT) information to recommend approach speeds that reduce unnecessary stopping and aggressive acceleration. A broad view of this literature is provided by Mellegård and Reichenberg, who map Day 1 C-ITS GLOSA research and highlight the diversity of assumptions regarding information availability, driver/vehicle compliance, and traffic conditions, which complicates direct comparison of reported benefits across studies [1].

A common extension of speed advisory is to coordinate it with signal control. Erdmann combines adaptive junction control with simultaneous GLOSA at a single junction using V2I information and dynamic programming, illustrating that larger system-level benefits are achievable when the infrastructure and vehicle advice are optimized jointly rather than treating the signal plan as exogenous [2]. This line of work motivates a clear separation between (i) *joint* signal–vehicle optimization and (ii) *advisory-only* approaches that operate under a fixed (or externally determined) signal plan. Our work focuses on the latter setting to reflect a low-intrusion deployment model in which only approach speed recommendations are issued and no modification of signal control is assumed. This approach allows for integration in current vehicle models with no automated driving capabilities as well as capability with fully autonomous vehicles in the longer term.

A central practical limitation of SPaT-based planning in advisory-only settings is that the approach is often constrained by surrounding traffic. Ye et al. address this by explicitly forecasting preceding-vehicle speeds and incorporating these constraints into prediction-based EAD at signalized intersections, showing that interaction-aware planning is required when leader dynamics dominate achievable trajectories [3]. A second, orthogonal limitation is uncertainty in signal timing itself. Mahler and Vahidi propose an optimal velocity-planning scheme that accounts for probabilistic traffic-signal timing, underlining that plans based on deterministic phase durations can be suboptimal when phase switching is not perfectly predictable [4]. For electric vehicles specifically, Simchon and Rabinovici formulate and implement a real-time Dynamic-GLOSA method for EVs over multi-segment routes, emphasizing the joint energy–travel-time trade-off and the importance of computational tractability for online advisory [5].

These works collectively motivate two aspects that shape our problem setting. First, even strong model-based advisory formulations must contend with (i) leader constraints and queueing, the latter is especially important with higher traffic volumes, on the approach [3] and (ii) variable or uncertain signal timing [4]. Second, joint signal–vehicle optimization can further improve outcomes but changes the deployment assumptions and requires infrastructure control authority [2]. In this paper, we therefore evaluate advisory policies under different phase durations for each approach and systematically varied interaction levels (different numbers of other vehicles), while restricting the intervention to a single recommended approach speed without signal-plan modification.

2.2 Learning-based speed advisory and eco-driving

Learning-based controllers have been explored as an alternative to purely model-based planning when the environment exhibits stochasticity and complex interactions. Jayawardana and Wu learn eco-driving strategies at signalized intersections and demonstrate that reinforcement learning can discover efficient behaviors that reduce fuel use while maintaining acceptable travel time, providing evidence that data-driven policies can compete with analytical approaches in this domain [6]. More recent work has expanded RL-GLOSA formulations beyond the recommended speed itself. Xu et al. propose adaptive frequency GLOSA, where deep reinforcement learning determines not only advisory content but also when advisories should be issued, targeting stability and reduced overreaction under changing traffic conditions [7].

Hybrid methods aim to combine the robustness of learning with the structure of model-based eco-driving. Bai et al. propose a hybrid reinforcement learning strategy for connected and automated vehicles at signalized intersections, explicitly addressing uncertainty while retaining an efficiency-oriented control structure [8]. Closely aligned with the interaction focus of our work, Ren et al. incorporate on-the-fly queue dissipation estimation and lane-merging disturbances into deep RL eco-driving at signalized intersections, highlighting that queue evolution and disturbance modeling are pivotal for performance in traffic-heavy settings [9]. Taken together, these studies suggest that RL is particularly valuable when performance is limited by factors that are difficult to capture with a single-shot phase-alignment rule (e.g., leader-induced constraints and time-varying queue dynamics), motivating controlled comparisons between transparent deterministic advisories and learned policies under varying interaction intensity.

2.3 Simulation and energy modeling in SUMO

Microscopic simulation is widely used to evaluate speed advisory systems because it enables repeatable comparisons under controlled information and compliance assumptions. For EV-focused studies, the credibility of energy outcomes depends on the energy consumption model used. Behrisch et al. compare and parameterize electrical energy consumption models in SUMO, underscoring that results can vary with modeling and calibration choices and should therefore be reported with vehicle and consistent configuration [10]. Within the SUMO community, Halbach presents a GLOSA-focused analysis in SUMO that studies impacts on speed distributions and emissions, providing further evidence that advisory-style speed changes can measurably influence both microscopic driving behavior and aggregate outcomes in simulation [11]. Building on these foundations, our study uses SUMO to provide a minimal, reproducible benchmark for comparing deterministic and reinforcement learning speed recommendations at a single signalized intersection across traffic volumes, using energy consumption and travel time as primary metrics.
While several larger studies leverage toolchains such as Flow [12] or SUMO-RL [13], we did not adopt those wrappers in this work. Instead, we use a custom environment and agent implementations developed from scratch.

3 Methods

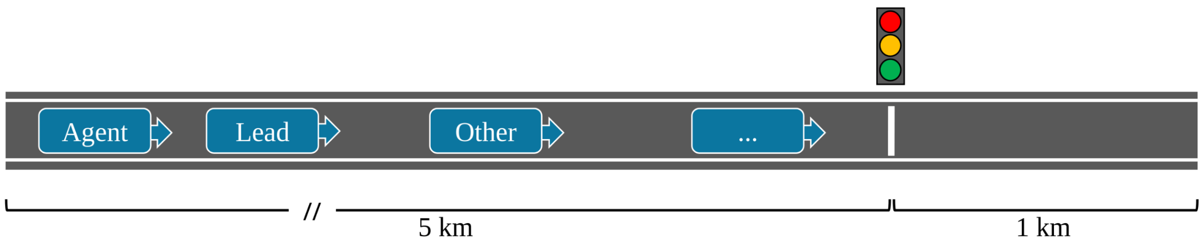
In this paper, two scenarios will be discussed separately. First, a SUMO [14] environment consisting of only one long and straight road with vehicles following it, where one vehicle follows speed suggestions from a deterministic algorithm and a traffic light is located at the end of that street. Second, the same SUMO environment but with the vehicle being controlled by a reinforcement learning agent.

A deterministic approach was first implemented to address scenarios with low traffic volumes. This algorithm was optimized for situations in which no other vehicles are present to interfere with its behaviour. Since this represents a limitation for real-world applications, a reinforcement learning agent was subsequently integrated. Several additional traffic flows were defined for the training episodes. However, to maintain comparability between the different approaches, both training and testing were conducted across all scenarios.

3.1 Scenario

The SUMO environment shown in Figure 1 was designed with simplicity in mind. It consists of a single-lane, one-way road with a total length of six kilometres. A traffic light is placed five kilometres from the start of the road. The remaining one kilometre downstream provides sufficient distance for vehicles to accelerate back to their desired speed after passing the signal. We distinguish between scenarios with only the agent vehicle driving and with two, six and 20 vehicles total, including the agent vehicle. For better reproducibility we opted not to use SUMO Flows. Instead, vehicles are manually spawned for each scenario. In each case where other vehicles coexist, the agent vehicle is the last vehicle as following vehicles would have no effect on decisions made in this scenario. The vehicle directly in front of the agent vehicle is designated as the lead vehicle. For all scenarios the $Energy/default$ emissions model is used. It is also assumed that all human drivers follow any given speed advice completely.

This setup is more representative of a suburban or peri-urban corridor than a dense urban scenario. In particular, a five-kilometre approach to a single signal is atypical for inner-city networks, where intersections occur at much shorter spacing. Likewise, the maximum speed of 100 aligns more closely with rural-road conditions than with typical city speed limits. Nevertheless, we considered it useful to start with this larger-scale, low-complexity setting as a gradual entry point to the problem: it provides a controlled environment in which vehicle-level adaptations can be implemented and evaluated at a meaningful scale before moving towards denser, more complex urban networks. The long stretch of road before the traffic light allows for vehicles to group up because some want to drive faster than others. This results in vehicles arriving in batches at the traffic light and further increases the level of difficulty to approximate an optimal speed to approach the traffic light, as it may not be guaranteed that all vehicles will pass the light during a single green phase.

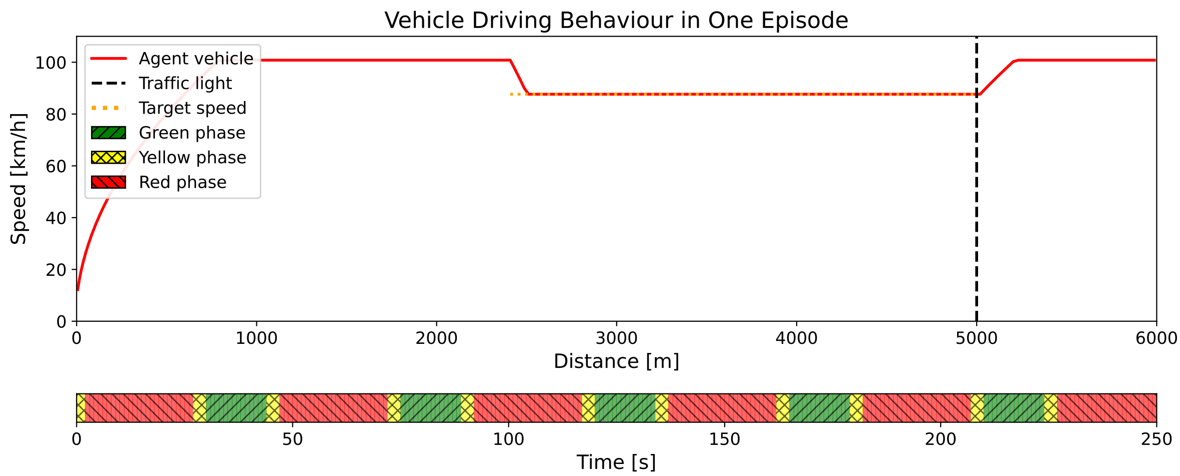


Vehicles positioned along a 5km road approach a traffic light, highlighting the agent, lead, and surrounding traffic.

Figure 1. *One-way street with a single lane and several vehicles driving towards a traffic light.*

The length of both the red and green phase is randomized within a given range for every approach by the agent vehicle. This makes the result of the actions of the reinforcement learning agent more realistic and useful because some traffic lights may

even change phase lengths for every cycle. The time range in which the randomization occurs is preset as $[10..20]$ seconds for the green phase and $[25..35]$ seconds for the red phase. The yellow phases are fixed at a length of 3 seconds. Each phase colour length is chosen independently to allow for real-world variation and a less predictable pattern.
Additionally, the road was created with just one lane to prevent vehicles from overtaking. This would be an interesting dynamic to observe but stretches outside the scope for this work. Using only one lane guarantees that vehicles have to slow down for other slower vehicles, thereby creating another level of difficulty for both algorithms in predicting whether all vehicles will pass the light within a single green phase.
As soon as the traffic light is passed, all control is given back to SUMO as this part of the road is solely existent for acceleration to the previously desired speed.



Speed profiles of the agent vehicle over distance, shown alongside traffic-light phases during a single driving episode.

Figure 2. *Representative episode of a single vehicle following the speed recommendation of the deterministic algorithm with no other interfering vehicles present*

3.2 Deterministic Approach

A deterministic speed-planning algorithm was implemented to serve as a transparent baseline for comparison with the learning-based controller. The approach computes a single target speed that aims to align the vehicle's arrival at the intersection with a green signal phase, assuming that the vehicle maintains a constant speed after the recommendation is issued.

The algorithm operates in a simplified SUMO scenario consisting of a straight, single-lane road of 6 length with one signalized intersection located after 5. This configuration provides sufficient distance for the vehicle to adjust its speed well in advance of the traffic light and to recover to its desired speed on the final road segment after the intersection. To reflect realistic driving conditions, vehicles are spawned with randomized initial speeds and starting positions. Information about the traffic light, including the current phase, the remaining time until the next phase switch, and the durations of all upcoming phases (which are randomized for each training episode), is obtained via the TraCI interface at a distance between 2 and 4 before the intersection.

Based on the current distance d to the traffic light and the vehicle's current speed v, a nominal arrival time is estimated under the assumption of constant speed. Starting from the current signal phase, the algorithm then advances through the sequence of traffic light phases by cumulatively adding their durations until the simulated signal time exceeds the estimated arrival time. This procedure determines the signal phase that the vehicle would encounter if no speed adjustment were made. If this predicted arrival phase is green, the current speed is retained.

If the predicted arrival phase is not green, the algorithm instead targets one following the red phase after the original arrival time. The time remaining until the beginning of this target green phase is computed, and the recommended speed is obtained by dividing the remaining distance d by this time interval. The resulting speed is clipped to the allowed speed range of the scenario. In effect, the vehicle is slowed down just enough to reach the stop line at the start of a feasible green phase, thereby avoiding a full stop at the signal whenever possible.

This deterministic approach does not explicitly account for interactions with leading vehicles or queue formation and therefore represents an optimistic baseline that is expected to perform best in low-traffic conditions. In Figure 2 an example trial (of the overall 1000 trials per traffic flow pattern and control method) can be seen. The figure shows how the speed is adapted with respect to the distance of the traffic light.

3.3 Reinforcement Learning Approach

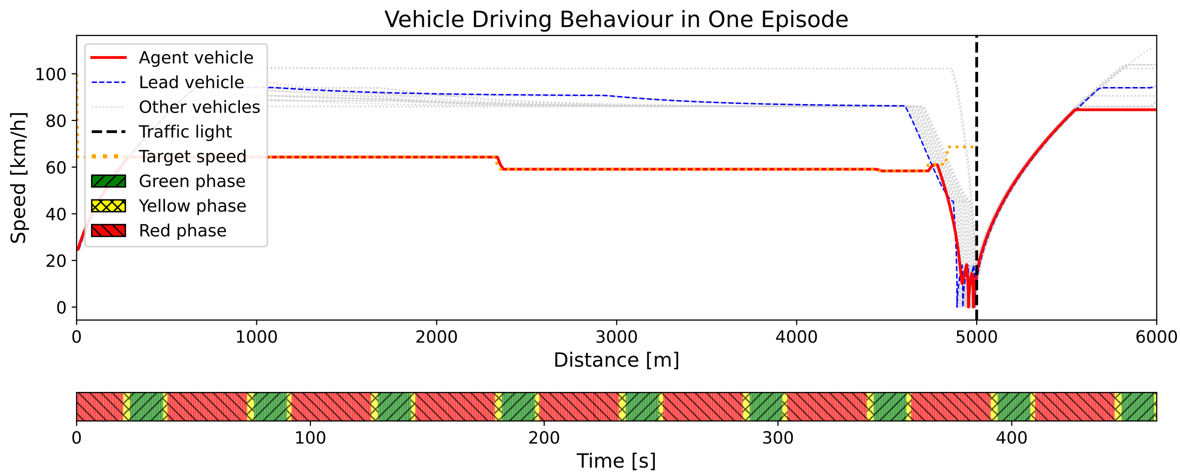
We train a reinforcement learning controller that produces speed recommendations for a single agent vehicle within the SUMO simulation. The task is formulated as a Markov Decision Process in which the simulator advances in discrete time steps of one second (one SUMO tick). At each step t, the agent observes the current traffic situation s_t, selects an action a_t, and receives a scalar reward r_t. Training is performed using Proximal Policy Optimization (PPO)[15] as implemented in Stable-Baselines3 [16] with an MLP-based actor–critic policy in scenarios with surrounding traffic. In Figure 3 we show an example trial of the reinforcement learning agent in a scenario with 20 vehicles. Next, we define the states, actions, and reward function used in our reinforcement learning algorithm.

3.3.1 State

The state is represented by a fixed-length vector

$$s_t = \left(v_t,\ d_t,\ \phi_t,\ \tau_t^{\text{green}},\ \tau_t^{\text{red}},\ v_t^{\text{lead}},\ d_t^{\text{lead}},\ E_t,\ N_t\right), \tag{1}$$

where v_t is the agent speed, d_t is the remaining distance to the traffic light, $\phi_t \in \{0,1,2,3\}$ encodes the current signal phase, and τ_t^{green} and τ_t^{red} denote the remaining time until the next green and red phase, respectively. To capture traffic interaction, v_t^{lead} and d_t^{lead} denote the speed and gap of the closest leading vehicle; if no leader exists, arbitrarily inconsequential values are used as a placeholder. The term E_t denotes the cumulative electrical energy consumed by the agent vehicle (Wh), and N_t denotes the number of vehicles ahead on the same road segment, serving as a coarse proxy for queue length.



Speed profiles of agent, lead, and other vehicles over distance, shown alongside traffic-light phases during a single driving episode.

Figure 3. *Representative episode of 20 vehicles driving on the road with the agent vehicle following the speed recommendation of the RL agent*

3.3.2 Actions

The action space is one-dimensional and continuous,

$$a_t \in [-1, 1]. \tag{2}$$

Actions are mapped to a target speed recommendation in via

$$v_{\text{kph}}^{\text{rec}}(a_t) = 50(a_t + 1), \tag{3}$$

which corresponds to the interval $[0, 100]$. The recommended speed in is

$$v_t^{\text{rec}} = \frac{v_{\text{kph}}^{\text{rec}}(a_t)}{3.6} = \frac{50}{3.6}(a_t + 1). \tag{4}$$

The recommendation is applied through the simulator interface; acceleration and deceleration remain bounded by SUMO's vehicle dynamics such that no instantaneous speed changes occur.

3.3.3 Reward Function

The reward r_t is computed at each step as the sum of three components,

$$r_t = r_t^{\text{time}} + r_t^{\text{energy}} + r_t^{\text{smooth}}. \tag{5}$$

The time component encourages progress by comparing the estimated remaining travel time to a reference duration $T_{\text{ref}} = 200$. Let

$$\hat{T}_t = \begin{cases} \dfrac{d_t^{\text{end}}}{v_t}, & v_t \geq 1, \\ 1000, & v_t < 1, \end{cases} \tag{6}$$

where d_t^{end} [] denotes the remaining distance to the end of the episode. The time reward is then

$$r_t^{\text{time}} = \begin{cases} 1, & \hat{T}_t \leq T_{\text{ref}}, \\ -\dfrac{\hat{T}_t - T_{\text{ref}}}{T_{\text{ref}}}, & \hat{T}_t > T_{\text{ref}}. \end{cases} \tag{7}$$

The energy component compares the current consumption rate C_t [] to a reference level $C_{\text{ref}} = 120$,

$$r_t^{\text{energy}} = \begin{cases} -\dfrac{C_t - C_{\text{ref}}}{60}, & C_t > C_{\text{ref}}, \\ \dfrac{C_{\text{ref}} - C_t}{C_{\text{ref}}}, & C_t \leq C_{\text{ref}}. \end{cases} \tag{8}$$

Finally, a smoothness penalty discourages abrupt speed adjustments. Let Δv_t [] denote the magnitude of the applied speed change. Then

$$r_t^{\text{smooth}} = \begin{cases} -\dfrac{\Delta v_t}{25}, & \Delta v_t > 1, \\ 0, & \Delta v_t \leq 1. \end{cases} \tag{9}$$

Together, these terms promote efficient driving with minimal delay while penalizing aggressive speed changes that would reduce comfort.

3.4 Evaluation Metrics

To evaluate the effectiveness of both the deterministic approach and the reinforcement learning agent, we consider travel time and energy consumption as primary performance metrics. All metrics are computed over the full lifetime of the agent vehicle, from its spawn position until it leaves the simulated road segment.

Travel time is measured as the total number of simulation ticks required to complete the episode. To account for variations in spawn position across episodes, travel time is normalized by the driven distance and reported as ticks per kilometer. Lower values indicate higher average speeds and fewer delays caused by stopping or queuing at the traffic light. Since ticks per kilometer is an unconventional unit, we later convert the travel time measurements to an average speed measurement in for more intuitive comparisons.

Energy efficiency is measured as cumulative electrical energy consumption normalized by the driven distance and reported in watt-hours per kilometer (). Lower energy consumption corresponds to lower average speeds or smoother driving behavior with fewer acceleration and deceleration events, particularly avoiding full stops at red signal phases. All vehicles in the scenario are electric and support regenerative braking, allowing kinetic energy to be partially recovered during deceleration. While recuperated energy is reflected in the cumulative energy consumption value, negative energy usage is not explicitly rewarded. Nevertheless, full stops are expected to increase travel time even when partial energy recovery is possible.

Together, these metrics capture the fundamental **trade-off between efficiency and travel time** that underlies speed recommendation strategies at signalized intersections.

3.5 Statistics

Normality was assessed using the Shapiro–Wilk test. As normality could not be assumed, non-parametric methods were employed. Differences between independent groups were evaluated using the Mann–Whitney U test. Correlations were analyzed using Spearman's rank correlation coefficient (ρ). To control the error rate across multiple comparisons, p-values were adjusted using the Holm–Bonferroni correction.

4 Results

We report results for the baseline (SUMO default driving), the deterministic speed-planning strategy, and the reinforcement learning (RL) agent across four traffic volumes (1, 2, 6, and 20 vehicles total). Each configuration was evaluated over $n = 1000$ episodes. Performance is assessed using (i) energy consumption normalized by distance () and (ii) average speed (). Table 1 reports means and standard deviations. Figure 4 summarizes distributions, Figure 5 shows histograms, and Figure 6 summarizes the energy–speed relationship.

4.1 Performance and distributional comparison across traffic volumes

Table 1 reports the mean energy consumption and average speed for each method at 1, 2, 6, and 20 vehicles total. Figure 4 summarizes the corresponding distributions, and Figure 5 shows the histograms over $n = 1000$ episodes per configuration.

Table 1. *Mean and standard deviation for energy consumption (left) and average speed (right).*

Energy consumption []		
Scenario	**Mean**	**Std**
Baseline, 1 vehicle	154.4	21.6
Baseline, 2 vehicles	144.0	17.8
Baseline, 6 vehicles	131.7	13.7
Baseline, 20 vehicles	**127.0**	11.0
Deterministic, 1 vehicle	132.7	19.1
Deterministic, 2 vehicles	124.7	15.1
Deterministic, 6 vehicles	120.0	12.7
Deterministic, 20 vehicles	**119.7**	11.2
RL Agent, 1 vehicle	141.7	12.4
RL Agent, 2 vehicles	116.0	8.6
RL Agent, 6 vehicles	111.6	8.5
RL Agent, 20 vehicles	**98.0**	5.3

Average speed []		
Scenario	**Mean**	**Std**
Baseline, 1 vehicle	87.4	11.5
Baseline, 2 vehicles	80.5	8.8
Baseline, 6 vehicles	70.2	6.3
Baseline, 20 vehicles	**51.7**	7.1
Deterministic, 1 vehicle	84.6	10.1
Deterministic, 2 vehicle	76.8	6.7
Deterministic, 6 vehicle	68.9	5.3
Deterministic, 20 vehicles	**51.7**	6.5
RL Agent, 1 vehicle	83.8	9.3
RL Agent, 2 vehicle	65.3	4.8
RL Agent, 6 vehicle	62.2	4.6
RL Agent, 20 vehicle	**50.8**	6.1

Side-by-side tables comparing mean and standard deviation of energy consumption and average speed across baseline, deterministic, and RL scenarios.

Across all traffic volumes, the RL agent achieves the lowest mean energy consumption, with the largest improvement observed in the 20-vehicle condition. While the deterministic approach reduces energy usage compared to the baseline in low-traffic scenarios, its gains saturate under higher interaction levels. In contrast, the RL agent continues to reduce mean energy consumption as traffic density increases.

The distributional analysis in Figure 4 confirms these trends. For energy consumption, the RL agent exhibits significantly lower distributions compared to both baseline and deterministic approaches in all traffic conditions (Holm–Bonferroni corrected $p < 0.001$). The separation is most pronounced in the 20-vehicle scenario, where the RL policy shifts the distribution substantially toward lower values and reduces high-energy outliers.

For average speed, differences depend on traffic density. In the 2- and 6-vehicle conditions, the RL agent shows significantly lower speeds compared to both baseline and deterministic control (corrected $p < 0.001$), reflecting a more conservative strategy in moderately constrained regimes. In the 20-vehicle scenario, speed differences remain statistically significant but are comparatively small ($p_{\text{corr}} = 4.36 \times 10^{-4}$ for baseline vs. RL), indicating that the observed energy improvements are achieved with only minor travel-time penalties.

4.2 Energy–speed relationship

Figure 6 visualizes the relationship between energy consumption and average speed for all methods and traffic volumes. The baseline correlations are $\rho = 0.50$ ($p < 0.0005$) at 1 vehicle, $\rho = 0.42$ ($p < 0.0005$) at 2 vehicles, $\rho = 0.31$ ($p < 0.0005$) at 6 vehicles, and $\rho = -0.09$ ($p = 0.027$) at 20 vehicles total.

For the deterministic approach, correlations are $\rho = 0.78$ ($p < 0.0005$) at 1 vehicle, $\rho = 0.70$ ($p < 0.0005$) at 2 vehicles, $\rho = 0.42$ ($p < 0.0005$) at 6 vehicles, and $\rho = -0.09$ ($p = 0.0033$) at 20 vehicles total.

For the RL agent, the correlation is $\rho = 0.02$ ($p = 0.5343$) at 1 vehicle, $\rho = -0.21$ ($p < 0.0005$) at 2 vehicles, $\rho = 0.08$ ($p = 0.0128$) at 6 vehicles, and $\rho = -0.14$ ($p < 0.0005$) at 20 vehicles total.

5 Discussion

This section interprets the empirical findings with respect to the energy–time trade-off and the influence of traffic density on advisory effectiveness. We discuss how deterministic and learning-based strategies differ across interaction regimes and what this implies for practical deployment.

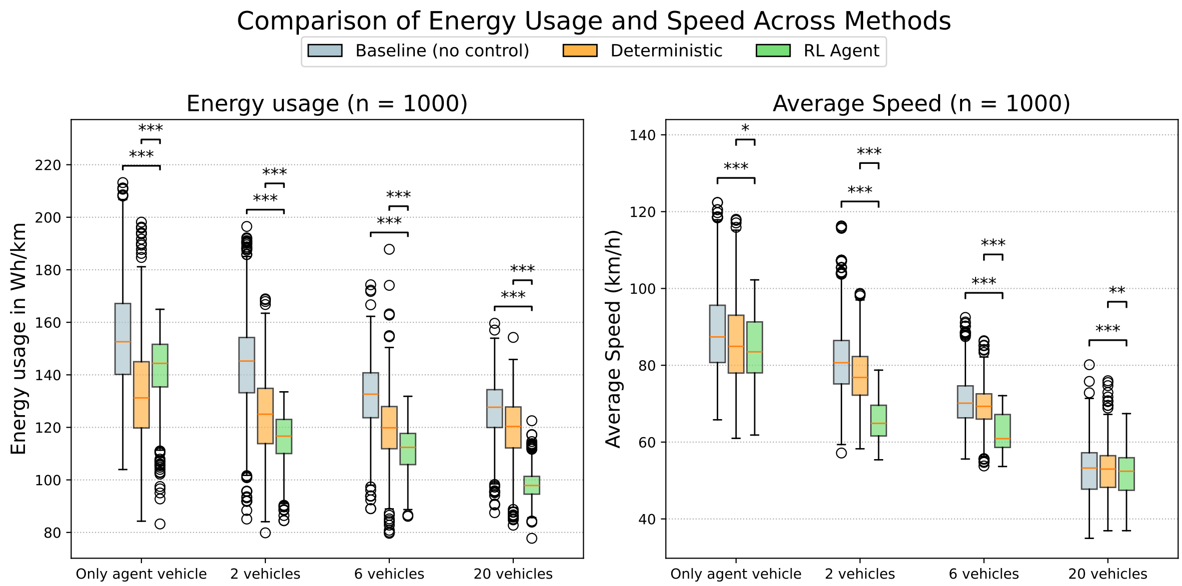
5.1 Finding 1: Energy reductions are consistent across traffic volumes

Across all traffic volumes (1, 2, 6, and 20 vehicles total), both advisory strategies reduce mean energy consumption relative to the baseline. The deterministic strategy reduces mean energy from 154.4 to 132.7 in the 1-vehicle case, and from 127.0 to 119.7 in the 20-vehicle case. The RL agent reduces mean energy to 141.7 (1 vehicle) and to 98.0 (20 vehicles total). This corresponds to a reduction of 22.83 compared to the baseline. It establishes that speed advice can shift the energy distribution downward (Figure 4) relative to unassisted driving.

5.2 Finding 2: Travel-time effects depend on traffic volume and method

Mean speed decreases with traffic volume for all methods. In the baseline, mean speed drops from 87.4 (1 vehicle) to 51.7 (20 vehicles total). The deterministic strategy remains close to the baseline at each traffic volume, including identical mean speed of 51.7 at 20 vehicles total. The RL agent shows larger speed reductions in the 2- and 6-vehicle conditions (65.3 and 62.2) relative to the baseline (80.5 and 70.2), while remaining close at 20 vehicles total (50.8 vs. 51.7).

Interpreted as an energy–time trade-off, the deterministic controller achieves energy reductions while largely preserving mean speed, whereas the RL policy's energy reductions in moderate traffic coincide with lower mean speeds.

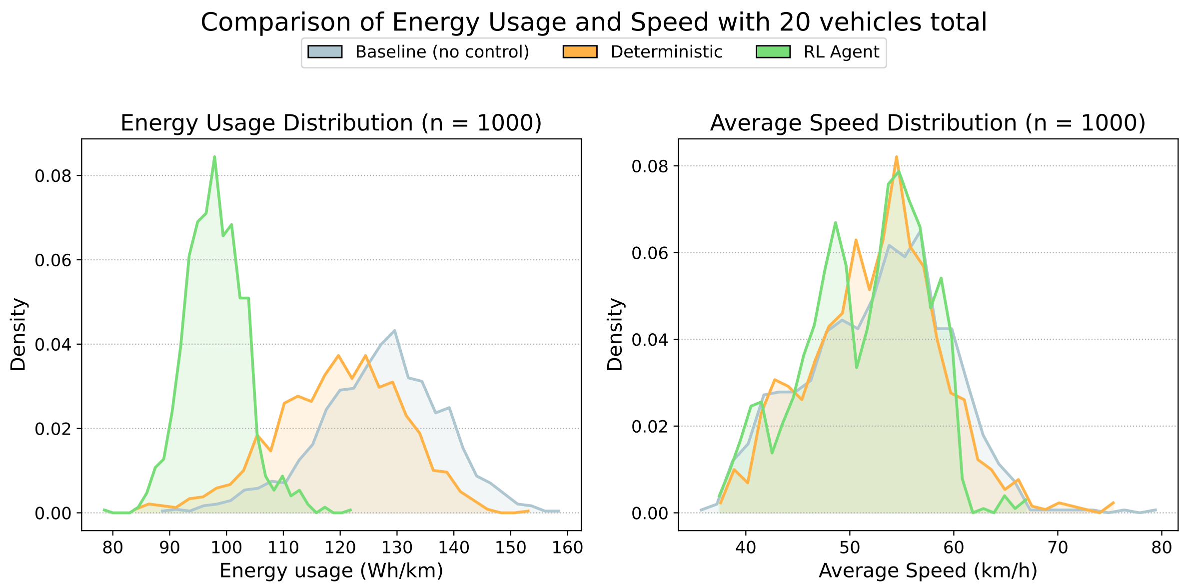


Box plots comparing energy usage and average speed across baseline, deterministic, and RL methods for different traffic densities.

Figure 4. *Distributions of energy consumption and average speed across methods and traffic volumes ($n = 1000$ per configuration and *** for $p < 0.001$, ** for $p < 0.01$ and * for $p < 0.05$)*

5.3 Finding 3: The RL policy improves consistency and reduces tail events

The RL agent shows narrower dispersion and fewer extreme high-energy outcomes relative to baseline driving (Figures 4–5).

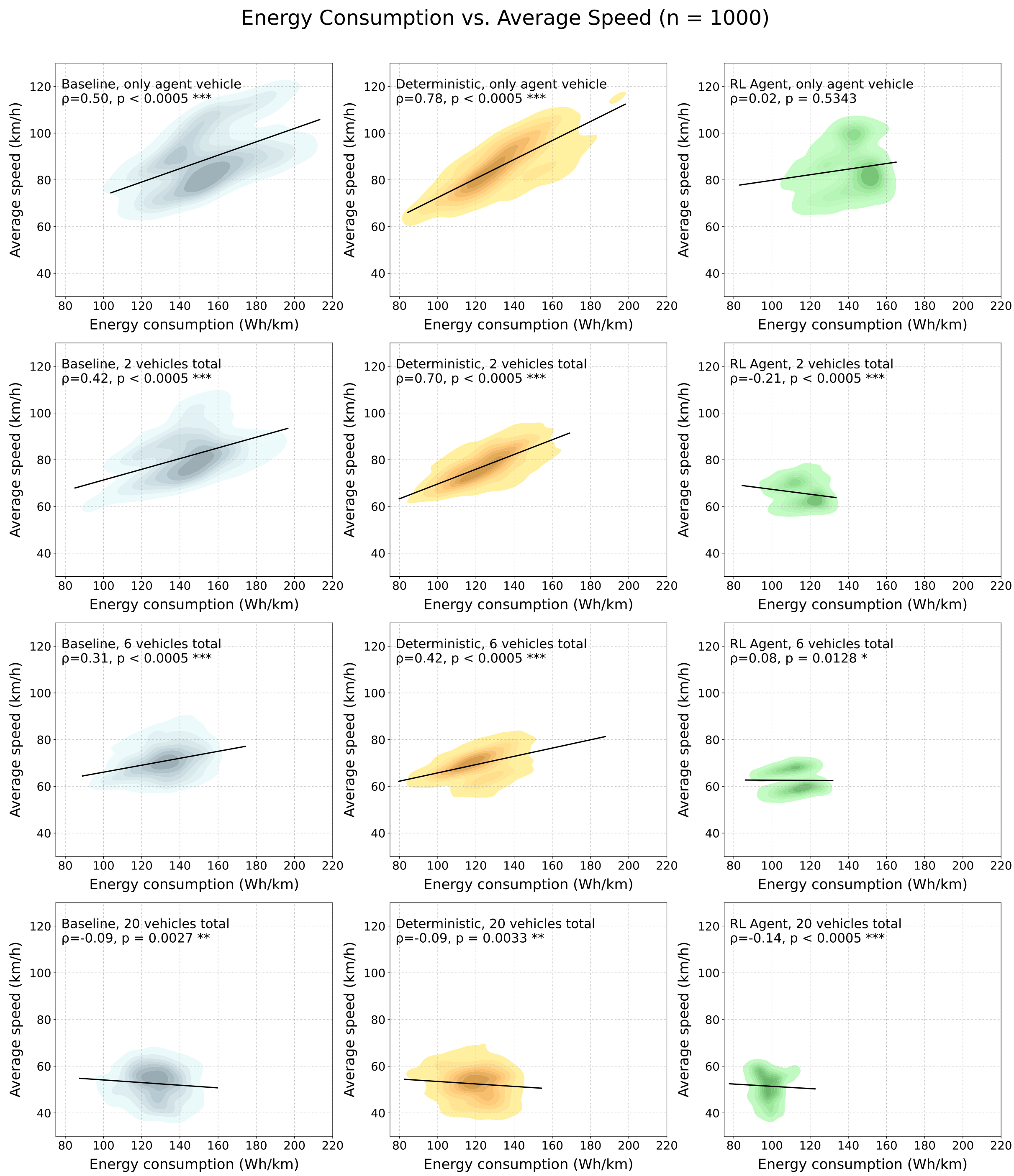


Density histograms comparing energy-usage and average-speed distributions for 20-vehicle simulations under baseline, deterministic, and RL-controlled conditions.

Figure 5. *Histograms of energy consumption and average speed across $n = 1000$ runs.*

In particular, the maximum observed energy consumption decreases from 164.9 (only agent vehicle driving) to 133.5 as soon as any other vehicles are present, while minimum values remain comparable across methods (around 85). From an operational perspective, reduced variance implies more predictable energy outcomes, which is relevant for driver acceptance and system-level planning.

5.4 Finding 4: Correlation patterns indicate regime shifts with traffic density



Scatter plots showing the relationship between average speed and energy consumption across baseline, deterministic, and RL-agent conditions with varying traffic levels.

Figure 6. *Correlation of energy consumption and average speed for three control methods across all traffic scenarios.*

The baseline and deterministic approaches show positive correlations between energy and speed in low traffic (1–2 vehicles), with $\rho = 0.50$ and $\rho = 0.42$ for the baseline and $\rho = 0.78$ and $\rho = 0.70$ for the deterministic strategy. At 20 vehicles total, both show slightly negative correlations ($\rho = -0.09$ for both), indicating a different operating regime in dense traffic. The RL agent shows a distinct pattern, with near-zero correlation at 1 vehicle ($\rho = 0.02$), negative correlation at 2 vehicles ($\rho = -0.21$), a weak positive correlation at 6 vehicles ($\rho = 0.08$), and negative correlation at 20 vehicles ($\rho = -0.14$) (Figure 6).

A plausible interpretation is that, as traffic becomes denser, variability in average speed increasingly reflects interaction effects (leader constraints, queue progression) rather than only acceleration choice, which changes the energy–speed association. The RL agent's sign changes across traffic volumes suggest that its policy adapts its driving style depending on the degree of constraint.

5.5 Implication: When to prefer deterministic vs. learning-based advice

In terms of method selection, the deterministic strategy offers energy reductions with minimal change in mean speed across traffic volumes, which supports use cases where transparency and stable travel-time behavior are primary requirements. The RL agent delivers the largest mean energy reduction in dense traffic (down to 98.0 at 20 vehicles total) and reduces high-energy outliers, which supports use cases where robustness and tail-risk reduction are important. In moderate traffic (2–6 vehicles total), the RL agent's lower mean speeds (65.3, 62.2) indicate a stronger energy–time trade-off than the deterministic approach, which should be considered when prioritizing throughput.

6 Limitations

The study was conducted in a simplified single-lane, single-intersection SUMO scenario to enable controlled comparison of advisory strategies. Network-level effects such as multi-intersection coordination, spillback, and lane-changing dynamics were not considered, and no multi-lane interactions or overtaking behavior were modeled.

All vehicles were modeled as electric and assumed to fully comply with recommended speeds. Real-world deployments would involve heterogeneous vehicle types and partial driver compliance, which may influence both energy savings and stability. Human driving variability, such as delayed reactions or imperfect adherence, was not represented.

The scenario models a long single-lane approach (5 km) to a single signalized intersection, which is more representative of peri-urban or suburban corridors than dense urban grids. While higher vehicle counts were used to simulate increased interaction intensity, the spatial structure does not reflect short intersection spacing typical for inner-city traffic. Consequently, the observed performance under "dense" traffic should be interpreted as high-interaction conditions within a corridor setting rather than a full urban network. Moreover, only the agent vehicle was actively controlled, preventing assessment of coordinated multi-vehicle effects.

Additionally, no speed limits were imposed upstream of the traffic light, as would typically be the case in real-world scenarios. This likely affects the significance of the results, as the overall energy savings potential would be lower under conditions with lower speed.

Traffic was generated using a manual instead of a SUMO flow configuration, possibly limiting variability in arrival patterns and excluding more flexible or stochastic flow conditions.

Despite these simplifications, the controlled setup allows isolation of interaction effects between traffic density and advisory strategy.

7 Future Work

While the present study provides a controlled comparison of deterministic and reinforcement learning–based speed advisory strategies, several extensions are necessary to assess real-world applicability.

First, the current evaluation is restricted to a single-lane, single-intersection scenario. Future work will extend the framework to multi-intersection corridors and urban networks with heterogeneous signal spacing, where coordination effects and spillback phenomena may further influence advisory performance.

Second, the present setup assumes homogeneous electric vehicles and full compliance with recommended speeds. Incorporating heterogeneous vehicle types (e.g., internal combustion engines, varying acceleration capabilities) and partial compliance models would allow investigation of mixed-traffic environments and driver acceptance.

Third, reward design was fixed across traffic volumes. Adaptive or multi-objective formulations that dynamically reweight energy and travel-time components could provide better context-aware behavior. Additionally, ablation studies on the state representation may clarify which traffic features are most critical for policy performance.

Finally, transferability and robustness remain open questions. Evaluating policy generalization across unseen traffic densities, signal timing distributions, and stochastic disturbances will be essential for deployment-oriented validation.

8 Conclusion

This paper evaluated deterministic and reinforcement learning–based speed recommendation strategies for approaching a signalized intersection in a microscopic SUMO simulation. Both approaches demonstrate that modest reductions in driving speed upstream of a traffic light can lead to significant energy savings with only minor impacts on travel time.

The deterministic speed-planning method proves effective in low-traffic scenarios, where signal timing is the dominant factor and vehicle interactions are limited. In contrast, the reinforcement learning agent provides the largest gains in high-traffic conditions, achieving substantial energy reductions while maintaining nearly identical travel times to unassisted driving. These results suggest that adaptive, interaction-aware policies are particularly valuable when queueing and leader constraints dominate vehicle behavior.

Taken together, the findings support a context-dependent deployment strategy: deterministic speed advisories offer a lightweight and interpretable solution for sparse traffic, while learning-based approaches are better suited for congested environments where adaptability and robustness are required. Future work will extend the evaluation to multi-intersection corridors, heterogeneous vehicle fleets, and mixed compliance settings, and will explore online adaptation and policy transfer across traffic scenarios.

Data Availability Statement

The implementation and datasets can be made available for research purposes upon request.

Author Contributions

Johan Kolms: Conceptualization, Methodology, Software, Investigation, Visualization, Writing: Original Draft, Review and Editing
Michael Thane: Data Curation, Formal Analysis, Visualization, Writing: Review and Editing
Kai Blum: Data Curation, Formal Analysis, Visualization, Writing: Review and Editing
Tamás Kurczveil: Project Administration, Conceptualization, Supervision, Writing: Review and Editing
Dirk Lehmann: Supervision, Conceptualization, Writing: Review and Editing

AI tool usage statement

During the preparation of this work, the authors used Microsoft Copilot for accessible alt text generation for visuals, for grammar checking and to improve the coherence of selected paragraphs. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work received no external funding.

Competing interests

The authors declare that they have no competing interests.

References

[1] N. Mellegård and F. Reichenberg, "The day 1 c-its application green light optimal speed advisory—a mapping study," *Transportation Research Procedia*, vol. 49, pp. 170–182, 2020, Facing the complexity of transport models and innovative developments in sustainable mobility - Selected Proceedings of the 47th European Transport Conference, ETC 2019, ISSN: 2352-1465. DOI: https://doi.org/10.1016/j.trpro.2020.09.015.

[2] J. Erdmann, "Combining adaptive junction control with simultaneous green-light-optimal-speed-advisory," in *2013 IEEE 5th International Symposium on Wireless Vehicular Communications (WiVeC)*, 2013, pp. 1–5. DOI: 10.1109/wivec.2013.6698230.

[3] F. Ye, P. Hao, X. Qi, G. Wu, K. Boriboonsomsin, and M. J. Barth, "Prediction-based eco-approach and departure at signalized intersections with speed forecasting on preceding vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1378–1389, 2019. DOI: 10.1109/TITS.2018.2856809.

[4] G. Mahler and A. Vahidi, "An optimal velocity-planning scheme for vehicle energy efficiency through probabilistic prediction of traffic-signal timing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2516–2523, 2014. DOI: 10.1109/TITS.2014.2319306.

[5] L. Simchon and R. Rabinovici, "Real-time implementation of green light optimal speed advisory for electric vehicles," *Vehicles*, vol. 2, no. 1, pp. 35–54, 2020. [Online]. Available: https://www.mdpi.com/2624-8921/2/1/3.

[6] V. Jayawardana and C. Wu, "Learning eco-driving strategies at signalized intersections," in *2022 European Control Conference (ECC)*, 2022, pp. 383–390. DOI: 10.23919/ECC55457.2022.9838000.

[7] M. Xu, D. Zuo, and J. Li, "Adaptive frequency green light optimal speed advisory based on deep reinforcement learning," *Journal of Transportation Engineering, Part A: Systems*, vol. 150, no. 10, p. 04 024 059, 2024. DOI: 10.1061/JTEPBS.TEENG-8392.

[8] Z. Bai, P. Hao, W. ShangGuan, B. Cai, and M. J. Barth, "Hybrid reinforcement learning-based eco-driving strategy for connected and automated vehicles at signalized intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 850–15 863, 2022. DOI: 10.1109/TITS.2022.3145798.

[9] X. Ren, C. S. Lai, G. Taylor, and Y. Yuan, "Eco-driving with deep reinforcement learning at signalized intersections considering on-the-fly queue dissipation estimation and lane-merging disturbances," *IEEE Open Journal of Vehicular Technology*, vol. 6, pp. 2789–2803, 2025. DOI: 10.1109/OJVT.2025.3618855.

[10] M. Behrisch, Y.-P. Flötteröd, and P. Wagner, "Comparing and parameterizing the electrical energy consumption models in sumo," *SUMO Conference Proceedings*, vol. 5, pp. 1–16, Jul. 2024. DOI: 10.52825/scp.v5i.1012. [Online]. Available: https://www.tib-op.org/ojs/index.php/scp/article/view/1012.

[11] M. Halbach, "Green light optimal speed advisory with reduced default speed - analysis of speed distribution and emissions of road users," in *SUMO User Conference 2025*, Mai 2025. [Online]. Available: https://elib.dlr.de/214276/.

[12] C. Wu, A. R. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: Architecture and benchmarking for reinforcement learning in traffic control," *arXiv preprint arXiv:1710.05465*, 2017.

[13] L. N. Alegre, *SUMO-RL*, https://github.com/LucasAlegre/sumo-rl, 2019.

[14] P. Alvarez Lopez, M. Behrisch, L. Bieker-Walz, *et al.*, "Microscopic traffic simulation using sumo," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, Nov. 2018, pp. 2575–2582. [Online]. Available: https://elib.dlr.de/127994/.

[15] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017. [Online]. Available: https://arxiv.org/abs/1707.06347.

[16] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1–8, 2021. [Online]. Available: http://jmlr.org/papers/v22/20-1364.html.