Impact of Deep RL-based Traffic Signal Control on Air Quality

Ammar Haydari *, Michael Zhang †, Chen-Nee Chuah *, Dipak Ghosal ‡
†Department of Civil and Environmental Engineering
*Department of Electrical and Computer Engineering
‡Department of Computer Science
University of California, Davis, California, USA
Email: {ahaydari, hmzhang, chuah, dghosal}@ucdavis.edu

***Abstract*—One major source of air pollution is automobile emissions in urban areas. Although hybrid and fully electric vehicles are started to gain popularity, the majority of vehicles are still fuel-based. With the rapid advancement of artificial intelligence (AI) and automation based controllers, there have been numerous studies applying such learning-based techniques to Intelligent Transportation Systems (ITS). Combining deep neural networks with reinforcement learning (RL) models called DRL has shown promising results when applied to urban Traffic Signal Control (TSC) for adaptive adjustment of traffic light schedules. Centralized and decentralized DRL-based controller models are proposed in literature to optimize the total system travel time. However, the associated impact of such learning-based TSCs to the air quality remains unexplored. In this paper, we examine the impact of DRL-based TSCs on the environment in terms of fuel consumption and CO2 emission. We studied a major DRL approach called advantage actor-critic (A2C) using multi-agent settings on a synthetic multi-intersection network and on a real traffic network of San Francisco downtown with 24 hours traffic dataset. Our initial results indicate that learning based DRL methods achieved the lowest air pollution level on synthetic networks even with a simple delay-based reward function. However, DRL-based TSC performs slightly worse than rule-based adaptive TSCs (max-pressure control) in the San Francisco network.**

***Index Terms*—Deep reinforcement learning, Intelligent transportation systems, Traffic signal control, Multi-agent systems, Deep learning.**

I. Introduction

Air pollution becomes a very problematic issue in urban areas due to the rise of the number of motor vehicles. In the US, transportation accounts for the 28% of greenhouse gas emissions in which 97.2% of source of emission is CO_2 via consumption of fuels [1]. Vehicular emission depends on several circumstances such as traffic condition, vehicle characteristics, and driver behaviors. Traffic intersections play a key role in managing mobile air pollution since frequent vehicles' speed changes and stop-and-go traffic result in increased fuel consumption and CO_2 emissions.

Machine learning-based control mechanisms in intelligent transportation systems (ITS), such as traffic signal control (TSC) systems, take action based on real-time data from the environment for online updating. Today, popular learning-based controller approaches combine deep neural networks (DNN) with RL, referred to as DRL, in which policy estimation is performed by DNNs. One good example application of such methods in ITS is developing the optimal traffic signal schedules. In general, learning-based TSCs perform better than standard dynamic TSCs in terms of delay and throughput for multi-intersection settings [2]. However, it remains an open research question how such learning based TSCs affect local mobile emissions near the surface streets.

In this context, we investigate the emission and fuel consumption produced by DRL controlled intersections. To assess the impact of such controllers in terms of the emissions, we consider policy-gradient-based advantage actor-critic (A2C) DRL algorithm with multi-agent settings and simulate the following: (i) grid-like 4-intersection TSC scenario, and (ii) the San Francisco Downtown road network. We run all our experiments on the SUMO traffic simulator where pollutant emission and fuel consumption models are derived from the HBEFA application database [3]. SUMO collects fuel consumption and pollutant emission results from each vehicle individually based on the speed and acceleration parameters. These emission statistics are examined using different type of traffic network settings.

The contributions of this paper are as follows:

- We quantify fuel consumption and CO_2 emission rates with multi-agent DRL controller methods using a simple delay-based penalty function. Our results show that the pollution levels are highly correlated with the total travel time in intersections and reducing the travel times spent in intersection also lowers the CO_2 emission.
- In addition to simulation study of a synthetic 2x2 grid network (Fig. 2), we train and test our DRL controller on the San Francisco downtown network with real data consisting of 24 hours traffic flow. To study the effect of peak and off-peak hours, we also perform different trace-driven simulations with 3 hours in the morning and 3 hours in the afternoon traffic flow, respectively.
- Although DRL-based TSCs perform the best on the synthetic network, they do not outperform the rule-based TSC method (max pressure control) in the San Francisco downtown network in terms of CO2 emissions and fuel

The authors acknowledge the financial support from CTECH, a USDOT University Transportation Center.

consumption. DRL-based TSCs outperforms both fixed-time and queue-based vehicle-actuated TSCs.

The rest of the paper is organized as follows. Section II discusses related work while Section III provides background for DRL learning agents and TSC settings. We discuss our simulation results in Section IV. Section V concludes the paper.

II. Literature Review

Learning-based TSC control mechanisms have good performance compared to classic TSC approaches. One such approach leverages different DNN settings, RL settings and traffic network structures referred to as DRL [2]. In general, the performance of learning based TSCs are better than standard TSC controllers in terms of delay and total waiting time [4]. Existing DRL based TSC approaches may differ from one another in terms of problem definitions [5], neural network structures [6] and applied algorithms [7]. While some studies control multiple intersections with a centralized agent [8], some others assign different agents for different intersections with multi-agent models [9].

Emission and fuel consumption increases in the urban areas due to high load of traffic and congestion [10]. Authors in [11] evaluates the impact of TSCs on air pollution based on VT-Micro microscopic fuel and emission estimation model [12]. Another team studies the effects of coordinated and non-coordinated TSC on emission rates with different emission models [13]. The work in [14] examines the roundabout effects on air pollution on a microscopic traffic simulator by comparing the results with standard fixed-time TSCs. A recent review discusses impact of different traffic management systems such as lane management, speed management and traffic flow control strategies on air pollution [10].

There are not many studies investigating the effects of learning based TSCs on air quality. In this paper, we examine the effects of learning based TSCs on CO_2 emission and fuel consumption on both a synthetic network and the San Francisco downtown network with the SUMO microscopic traffic simulator.

III. DRL-based Traffic Signal Controllers

A. Deep Reinforcement Learning

Reinforcement learning (RL) is a trial-and-error based learning algorithm where agent interacts with the environment and takes action to maximize cumulative reward. Mathematical formulation of RL is based on Markov Decision Process (MDP). In general, an RL agent interacts with the environment and receives a numerical reward (or penalty if it is negative). Continuously observing the environment called state s_t, receiving feedback from the environment r_t and taking action a_t, an RL agent learns an action policy which defines how to behave by computing action value function $Q(s_t, a_t)$ after each iteration [15]. Through linear approximation, DNN can estimate this function easily. Controlling RL agents with DNN-based function approximations is called DRL [16].

1) Advantage Actor-Critic DRL: In a general DRL model, DNNs extract the features from data with multi-layered neural networks [16]. Actor-critic-based DRL models consist of policy estimation and value function estimation algorithms applying to an advantage function (Fig 1). Instead of estimating the Q-value function only with a single learner neural network, the A2C approach estimates the policy function with critic network and Q-value function with an actor network. Since the policy gradient-based policy estimation algorithms are also not effective in large-scale applications due to high variance of the policy estimation, a general solution is to combine policy and value functions with an advantage function using two individual estimators, where the agent's behaviour is controlled by policy and the actions are balanced with Q-value functions. A2C actor-critic models update both actor and critic networks synchronously. There are several synchronous and asynchronous actor critic models in literature [17]. Asynchronous advantage actor-critic (A3C) models estimate both actor and critic networks in parallel asynchronously, which increases the computation time. Since there is not much performance difference between synchronous and asynchronous actor-critic models, we used synchronous actor-actor critic method know as A2C. The learning is stabilized with experience replay memory [15], which stores the experiments in replay memory and samples experiments randomly from memory for policy and value function estimation. Such experience replay models are good for preventing agents from getting stuck in a local optimal point.

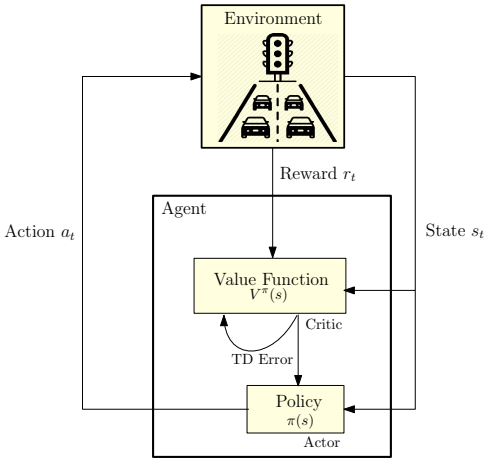


Fig. 1. Actor Critic RL model for a TSC

B. Deep Reinforcement Learning for TSC

In this work, the states of A2C agents are value vectors for each incoming lane of intersection. For one intersection, we created two value vectors for each lane: one is average speed and the other is total number of vehicles. Position and speed of each vehicle can be collected from individual vehicles via vehicle-to-infrastructure (V2I) communication to calculate the average speed and number of vehicles. Using the formed state input, the DRL agent in TSC selects a green

phase from among possible green phases: North-South Green, East-West Green, North-South Advance Left Green, East-West Advance Left Green. Each selected green phase is executed after a yellow phase transition. With the objective of maximizing cumulative reward, a scalar reward is computed for penalizing or rewarding each taken action. There are several reward/penalty definitions for TSC settings such as vehicle waiting time, cumulative delay, and queue length. Although there are more complicated reward designs in literature, the authors in [8] demonstrated that in general, simpler state and reward definitions are superior to the complex reward functions. For this reason, in our DRL-based TSCs, we choose a simpler reward function namely the change in the waiting time at an intersection for one green phase.

For DRL models, designing a DNN structure for better performance is another critical step. In this paper, we used multi-layer perceptron with 5 layers for both actor and critic, with "relu" and "softmax" activation functions for policy estimations of learning agents. In multi-agent RL settings, interaction with the nearest agents is necessary to reach a global optimum. In our experiments, each agent updates its policy by including the current traffic condition of neighbor TSCs as well to decrease the overall traffic delay. The global state is found with concatenation of the local states of neighboring intersections and the reward is generated by summing the local rewards of neighboring intersections.

C. Fuel Consumption and Emission Models

There are several vehicle acceleration-based emission estimation models such as HBEFA [18], MODEM [19]. We adopted the HBEFA emission estimation model in our experiments, which is widely used in Europe, with SUMO traffic simulator providing a variety of tools for collecting statistics from the simulation. The parameter called relative positive acceleration (RPA) is a key component determining the emission rates for driving cycles. RPA value is calculated using the equation:

$$RPA = \frac{1}{\lambda} \sum a_i * v_i * \Delta t \tag{1}$$

where λ is the total traveled distance, a_i is positive acceleration value, v_i is speed for the sample i, and Δt is the time interval between sample i and $i-1$.

The latest version of HBEFA is v4.1 released in August 2019. Although HBEFA includes a large amount of source data for different sort of pollutants, SUMO only allows its users to simulate a few of them such as fuel consumption, CO_2, CO, HC. In our experiments, we only measured the rates of fuel consumption and CO_2 since we know that 97.2% of emission in traffic is only CO_2. SUMO also enables the use of different vehicle classes for simulating such parameters. Some of them are passenger cars, buses, heavy duty vehicles with gas driven and diesel driven types. In this work, we only simulated one type of vehicle that releases the same amount of gas to the air and consumes the same amount of fuel for all the vehicles.

IV. Experimental Evaluation

In this section, we experimented the impact of DRL-based TSCs on fuel consumption and CO_2 emission statistics using SUMO [20] microscopic vehicular traffic simulator with Tensorflow Python API for controlling multi-agent A2C agents. Both synthetic and real networks are trained on the same agent parameters with 2000 experience replay memory size, discount factor $\gamma = 0.95$, as well as, 0.00001 and 0.000005 learning rates for actor and critic networks, respectively.

All our experiments compare the performance of DRL TSCs with three baselines. One of the baselines is standard fixed time TSC where green light times are allocated to each direction with pre-defined duration. We also compared our method with two adaptive control methods: queue-based vehicle-actuated TSC [21], and max-pressure-based TSC [22]. Maximum phase duration for the vehicle-actuated controller and the max-pressure controller and DRL controller is set to be 40 seconds.

A. Results from the Synthetic Network

In this section, we perform experiments on a multi-intersection environment with A2C DRL-based TSCs using 4 connected intersections (see Fig. 2). One traffic intersection has only 3 incoming roads while the other three intersections have 4 incoming roads. The roads connecting the different intersections are 1000 meters long, while the roads on the edges are 500meters long. 1 hour traffic flows on the synthetic traffic network constitutes one episode. The traffic is generated one vehicle per second by selecting the origin and destinations randomly. We trained our DRL agent on synthetic network for 20 episodes.

Due to space limitations we do not include comparison results with other DRL methods here but our previous experiments show that multi-agent A2C model achieves the best performance among other DRL models. In this paper, we only showed the impact of multi-agent A2C (MA2C) model on

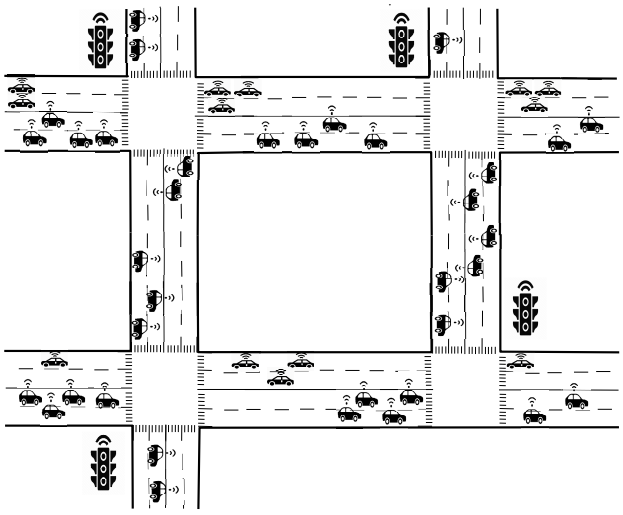


Fig. 2. Traffic scenario for multi-agent multi-intersection TSCs.

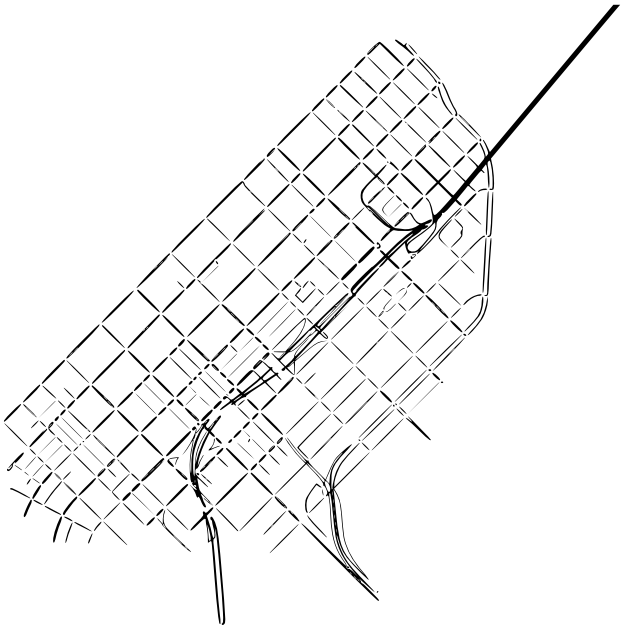


Fig. 3. San Francisco downtown traffic network

fuel consumption and CO_2 emission rate in addition to total network waiting time. Fig. 4 shows the air pollution statistics and total vehicle waiting time with established baselines fixed-time, actuated and max-pressure TSCs throughout the simulation for the synthetic network. Fig. 4(a) exhibits the learning curve of multi-agent A2C agents in terms of total travel waiting time. Fig. 4(b) and Fig. 4(c) show the total fuel consumption rate and total CO_2 emission rate. Results in Fig. 4 shows that DRL based TSC can achieve the minimal fuel consumption and CO2 emission rate, along with the total waiting time. In general, as the total vehicles travel time decreases, fuel consumption and CO_2 emission rate also decrease proportionally.

B. Results from the Real Network

In addition to simulating the synthetic road network, we evaluated the DRL-based traffic controllers and state-of-the-art conventional TSC controllers using a real dataset on San Francisco downtown road network, which follows a grid structure. The traffic from the bay bridge is also a part of the traffic flow in San Francisco downtown, where the bridge is merged with the main downtown traffic network. Fig. 3 shows the downtown San Francisco traffic network with 115 signalized intersections in total. Since it is not practical to control all the signalized intersections, we trained and tested only 10 neighboring intersections in the lower central downtown area. In addition, we tested our DRL agent with 4 neighboring intersections on the real road network, closer to the synthetic network. The results of such a 4 intersection controller have similar results with 10 intersections. Hence, in this paper we only present the 10 intersection controller model results below. We trained our DRL-based TSC controller with a 24 hours replicated traffic route file where similar timely traffic patterns are preserved. Cumulative CO_2 emission and fuel consumption rates are collected at around the signalized intersections. We presented real network test results in separate tables for three scenarios: 24-hour all-day traffic, 8am-11am morning traffic, and 5pm-8pm evening traffic.

First, the all-day simulation results for the San Francisco traffic network is shown in Table I. Although multi-agent A2C achieves the highest performance in the synthetic network, it performs slightly worse than Max-pressured-based TSC in the San Francisco network. Among the four TSC models we evaluated, DRL-based TSC controller achieves the second best performance in terms of total vehicle waiting time, total fuel consumption and total CO_2 emission.

TABLE I
Comparison of different TSC controllers using 24 hours traffic flow on San Francisco downtown network

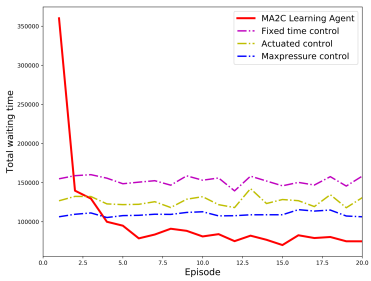
TSC	Waiting time (sec)	Fuel (liter)	CO_2 (gram)
Max-pressure	**658656**	**1658.5**	**128443.5**
MA2C	783140	1835.2	146028.4
Actuated	845829	1925.4	159295.3
Fixed-time	1453968	2543.5	254762.2

Next, we studied the San Francisco network with 3 hours traffic flow for two groups of time periods: 8am-11am and 5pm-8pm. The purpose of this analysis is to identify how learning agents behave in different time periods of the day. SUMO runs traffic flow with a given route file. Since we have only one all-day dataset, we need to train the network with replicated traffic flow route files before testing learning agent with the actual traffic conditions. We randomly sampled traffic routes and replaced some of the routes with sampled routes for creating a replicated route file. This way, we preserved the same traffic behaviors for the given time period. However, we observe that training with the 3-hour dataset with one replicated route file does not provide sufficient learning for the DRL agent. Therefore, we generated 10 different route files with the same traffic behaviours and trained the DRL agent 10 episodes. Then we tested real traffic routes with DRL agent. Tables II and III summarize our results.

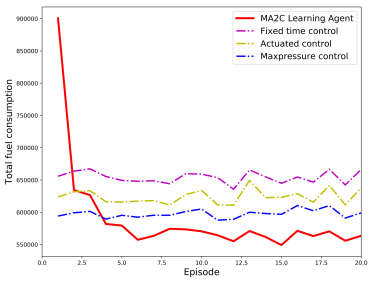
TABLE II
Comparison of different TSC controllers using 3 hours traffic flow on San Francisco downtown network between 8am and 11am

TSC	Waiting time (sec)	Fuel (liter)	CO_2 (gram)
Max-pressure	**94762**	**285.6**	**19594.0**
MA2C	110485	310.1	21789.2
Actuated	141265	341.7	27024.0
Fixed-time	218525	421.1	38889.0

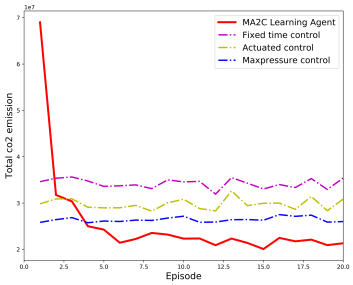
We begin with presenting the morning simulation results in Table II. Similar to the all-day results in Table I, the max-pressure TSC performs best in lowering traffic congestion, fuel consumption and CO2 emissions than other controllers, with MA2C comes second.



(a) Total waiting time for 1 hour traffic flow.



(b) Total fuel consumption for 1 hour traffic flow.



(c) Total CO_2 emission rate for 1 hour traffic flow.

Fig. 4. Waiting time, fuel consumption and CO2 emission rate results compared with standard fixed time and actuated controller models including max-pressure control

TABLE III
Comparison of different TSC controllers using 3 hours traffic flow on San Francisco downtown network between 5pm and 8pm

TSC	Waiting time (sec)	Fuel (liter)	CO_2 (gram)
Max-pressure	**40537**	**146.7**	**8941.3**
MA2C	61584	170.6	11972.0
Actuated	69748	181.2	13495.7
Fixed-time	117731	227.5	20824.9

Next we present the results for the evening period shown in Table III. Compared with the morning period, the evening period has lower congestion, fuel consumption, and CO_2 emissions, largely due a difference in traffic demand between the two peak commuting periods. Among all the control methods, the max pressure controller still performs the best, with MA2C being the second best. But the performance of MA2C is closer to that of the actuated controller in the evening period than in the morning period.

V. Conclusion

This paper investigated the effectiveness of learning based TSCs in reducing fuel and emissions, as compared with other state-of-the-art conventional TSCs, on both a synthetic and a real road network. The main findings are (i) there is a high correlation between the CO2 emission and fuel consumption rates and the total waiting time, (ii) learning based TSC controllers are not universally more effective than other types of controllers in our application context. While the multi-agent A2C controller achieves the best performance on the synthetic network, it was outperformed by the max pressure traffic controller on the San Francisco downtown network in all three testing scenarios. Nevertheless, the DRL controller still performs the second best in these cases. Several factors influence the ability of DRL controllers to learn and generalize, one of which is the reward function. Our current study used a simple reward function based on vehicle waiting time only. We will explore other forms of reward functions including the emission in our future work to see if the performance of the DRL controller can be further improved.

References

[1] U. E. P. Agency, "Fast facts on transportation greenhouse gas emissions."

[2] A. Haydari and Y. Yilmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[3] S. Hausberger, D. Engler, M. Ivanisin, and M. Rexeis, "Update of the emission functions for heavy duty vehicles in the handbook emission factors for road traffic," *Federal Environment Agency, Vienna/Austria 2003*, 2003.

[4] W. Genders and S. Razavi, "Asynchronous n-step q-learning adaptive traffic signal control," *Journal of Intelligent Transportation Systems*, vol. 23, no. 4, pp. 319–331, 2019.

[5] T. Tan, F. Bao, Y. Deng, A. Jin, Q. Dai, and J. Wang, "Cooperative deep reinforcement learning for large-scale traffic grid signal control," *IEEE transactions on cybernetics*, 2019.

[6] T. Nishi, K. Otaki, K. Hayakawa, and T. Yoshimura, "Traffic signal control based on reinforcement learning with graph convolutional neural nets," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 877–883.

[7] N. Casas, "Deep deterministic policy gradient for urban traffic light control," *arXiv preprint arXiv:1703.09035*, 2017.

[8] G. Zheng, X. Zang, N. Xu, H. Wei, Z. Yu, V. Gayah, K. Xu, and Z. Li, "Diagnosing reinforcement learning for traffic signal control," *arXiv preprint arXiv:1905.04716*, 2019.

[9] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[10] A. Y. Bigazzi and M. Rouleau, "Can traffic management strategies improve urban air quality? a review of the evidence," *Journal of Transport & Health*, vol. 7, pp. 111–124, 2017.

[11] J. Kwak, B. Park, and J. Lee, "Evaluating the impacts of urban corridor traffic signal optimization on vehicle emissions and fuel consumption," *Transportation Planning and Technology*, vol. 35, no. 2, pp. 145–160, 2012.

[12] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *Journal of transportation engineering*, vol. 128, no. 2, pp. 182–190, 2002.

[13] B. De Coensel, A. Can, B. Degraeuwe, I. De Vlieger, and D. Botteldooren, "Effects of traffic signal coordination on noise and air pollutant emissions," *Environmental Modelling & Software*, vol. 35, pp. 74–83, 2012.

[14] M. Gastaldi, C. Meneguzzer, R. Rossi, L. Della Lucia, and G. Gecchele, "Evaluation of air pollution impacts of a signal control to roundabout conversion using microsimulation," *Transportation research procedia*, vol. 3, pp. 1031–1040, 2014.

[15] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[17] T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *CoRR*, vol. abs/1509.02971, 2016.

[18] HBEFA, "Handbuch für emissionsfaktoren des strassenverkehrs (hbefa) (handbook of emission factors for road traffic)," *Umweltbundesamt Berlin, Bundesamt für Umwelt, Wald und Landschaft Bern*, pp. Infras AG, Bern, August 2019.

[19] J. Young Park, R. B. Noland, and J. W. Polak, "Microscopic model of air pollutant concentrations: Comparison of simulated results with measured and macroscopic estimates," *Transportation research record*, vol. 1750, no. 1, pp. 64–73, 2001.

[20] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. WieBner, "Microscopic traffic simulation using sumo," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2575–2582.

[21] H. Wei, G. Zheng, V. V. Gayah, and Z. Li, "A survey on traffic signal control methods," *CoRR*, vol. abs/1904.08117, 2019. [Online]. Available: http://arxiv.org/abs/1904.08117

[22] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177–195, 2013.