Adaptive Parameterized Control for Coordinated traffic Management Using Reinforcement Learning

Sun, D.; Jamshidnejad, A.; De Schutter, B.H.K.

DOI
10.1016/j.ifacol.2023.10.198

Publication date
2023

Document Version
Final published version

Published in
IFAC-PapersOnLine

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 5463–5468

Adaptive Parameterized Control for Coordinated Traffic Management Using Reinforcement Learning ⋆

Dingshan Sun * **Anahita Jamshidnejad** ** **Bart De Schutter** *

* *Delft Center for Systems and Control, Delft University of Technology, The Netherlands (e-mail: d.sun-1@tudelft.nl, b.deschutter@tudelft.nl).*
** *Department of Control and Operations, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands (e-mail: a.jamshidnejad@tudelft.nl)*

Abstract: Traffic control is essential to reduce congestion in both urban and freeway traffic networks. These control measures include ramp metering and variable speed limits for freeways, and traffic signal control for urban traffic. However, current traffic control methods are either too simple to respond to complex traffic environment, or too sophisticated for real-life implementation. In this paper, we propose an adaptive parameterized control method for traffic management by using reinforcement learning algorithms. This method takes advantage of the simple structure of parameterized state-feedback controllers for traffic; meanwhile, a reinforcement learning agent is employed to adjust the parameters of the controllers on-line to react to the varying environment. Therefore, the proposed method requires limited real-time computational efforts, and is adaptive to external disturbances. Furthermore, the reinforcement learning agent can coordinate multiple local traffic controllers when adjusting their parameters. The method is validated by a numerical case study on a freeway network. Results show that the proposed method outperforms conventional controllers when the system is exposed to a changing environment.

Keywords: Parameterized control, adaptive control, reinforcement learning, coordinated control, traffic network system.

1. INTRODUCTION

1.1 Background

The ever-increasing traffic demand is bringing more and more burden to existing traffic systems, both urban and freeway networks. When the demand exceeds the road capacity, traffic congestion can occur easily, which results in negative impacts on environmental, economic, and societal aspects. Individuals also suffer from severe traffic congestion, since they could be exposed to harmful emissions and their daily commute becomes inefficient. To relieve heavy traffic jams, many traffic control measures have been proposed and achieved significant success in field applications (Papageorgiou et al., 2003). For example, traffic signal control for intersections in urban networks is proved to be effective in reducing the total time spent (TTS) of all the vehicles. For freeway networks, ramp metering and variable speed limits are the most widely used control measures (Papageorgiou and Kotsialos, 2002).

1.2 Parameterized Traffic Control

Many studies have been carried out to improve the performance of traffic control measures (see e.g. (Kotsialos and Papageorgiou, 2004)). Traditional control methods, such as fixed-time traffic signal control, calculate the signal cycle settings off-line to minimize the total delay time for a certain pattern of traffic demand. These strategies are easy to implement and need low maintenance costs. However, they rely on historical data and cannot adapt to changing conditions of the network (e.g., varying demands). Traffic responsive strategies were proposed to address this issue. A well-known state-feedback controller is ALINEA (Papageorgiou et al., 1991), which was designed to regulate the downstream density of the freeway on-ramps using ramp metering. At every control time step, the current downstream density is measured and fed to the controller. Then the ramp metering rate is adjusted to reduce the error between the desired density and actual density. A similar strategy can be deployed for variable speed limit control and urban traffic signal control. Zegeye et al. (2012) proposed parameterized variable speed limits controllers that have a simple structure to minimize the differences in speed and density between the segments. van Kooten et al. (2017) utilized a similar idea for the green time split in urban traffic signal control. The green times are distributed to different phases according to several

⋆ This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101018826 - ERC Advanced Grant CLariNet). This work is also sponsored by China Scholarship Council, Grant No. 201806230254.

10.1016/j.ifacol.2023.10.198

indices, including the queue length, the waiting time of the first vehicle, and the total number of vehicles on the corresponding lanes.

The methods discussed above are state-feedback and can react to the changing states of traffic networks efficiently. However, they all need fine-tuned parameters to work properly, especially when considering uncertainties and unknown disturbances. Moreover, the controllers of a traffic network in general work independently and do not coordinate with each other.

1.3 Reinforcement Learning for Traffic Control

Reinforcement learning (RL) is one the most important control methods for traffic management. RL is basically a training method that interacts with the environment to find the best action according to the feedback rewards or penalties. From traditional Q-learning (Abdulhai et al., 2003; Li et al., 2017) to recently emerging deep RL algorithms (Chu et al., 2019; Wang et al., 2022), RL has achieved significant success in both urban and freeway networks control, due to its adaptive nature to deal with uncertainties. Moreover, it is reported by Wang et al. (2022) that a single deep RL agent can handle a large scale traffic network by coordinating multiple local controllers. Nonetheless, RL methods still struggle with low sample efficiency (i.e., a prolonged training process is needed before implementation), as well as the lack of performance and safety guarantees.

1.4 Proposed RL-based Adaptive Parameterized Control

To overcome the shortcomings of current traffic control methods, this paper for the first time proposes a RL-based adaptive parameterized traffic controller. Specifically, RL is trained to tune the parameters of the controllers, such that they can adapt to the changing environment, and multiple local controllers can be coordinated by a central RL agent. The proposed method inherits the simple structure of the state-feedback controllers; so it is cheap to implement. Compared with standalone RL-based control, the proposed method is more robust since it is used to tune the parameters, instead of directly determining the control inputs.

The remainder of this paper is organized as follows. Section 2 introduces related work about adaptive parameterized traffic controllers. The newly proposed RL-based adaptive parameterized traffic control method is presented in Section 3. The method is evaluated with a freeway network considering ramp metering control in Section 4. Conclusions and future work in Section 5 end this paper.

2. RELATED WORK

2.1 Parameterized MPC for Traffic Control

Model predictive control (MPC) (Camacho and Alba, 2013) is one major research direction in the field of traffic management. MPC is basically a model-based optimal control method that optimize an objective function every control time step based on the prediction of future states. Gartner et al. (1976) first applied optimal control method in traffic signal control, and De Schutter and De Moor (1998) formally suggested to use MPC in this field. Similar work was done by Kotsialos et al. (2002b). After that, MPC had been applied in freeway network by Hegyi et al. (2005), in which ramp metering and variable speed limit control are coordinated. But the real-time implementation of MPC is not always feasible, due to the significant computational efforts required to solve the non-linear and non-convex optimization problems caused by the complex mathematical model of traffic networks. Parameterized MPC (Zegeye et al., 2012) is one of the approaches that address the computational issue of MPC.

Parameterized MPC has been applied in both urban and freeway networks to combine MPC and parameterized traffic controllers. Zegeye et al. (2012) used MPC to optimize the parameters of ramp metering controllers and variable speed limits controllers simultaneously at every control time step. van de Weg et al. (2018) further extended this method by considering adjusting the parameters of ALINEA for ramp metering control, as well as the switching times of the controller. Jeschke et al. (2023) applied parameterized MPC in an urban traffic network. Similar to van Kooten et al. (2017), they used a parameterized state-feedback function to distribute the green time length for each phase, based on the queue length and number of vehicles on the corresponding lanes. The parameters in the function are then taken as the optimization variables for MPC. Parameterized MPC can significantly reduce the computational complexity compared to conventional MPC, since the optimization variables are the parameters which are independent of the prediction horizon. However, parameterized MPC still suffers from the model mismatch between the mathematical model and the real traffic system. In addition, external disturbances (e.g., weather change or incidents) also limit the performance of MPC methods.

2.2 Adaptive Parameterized Traffic Control

Only few studies can be found that focus on improving the adaptivity of the parameterized traffic controllers. Ghods et al. (2009) proposed a genetic-based fuzzy approach for adaptive freeway ramp metering and variable speed limit control. A genetic algorithm is running in real time to adjust the parameters of the baseline fuzzy controller every 5 min, trying to minimize the TTS of the network for a prediction horizon. However, a model is still needed in the genetic algorithm block for prediction, and it is not explained how the genetic block tunes the parameters of the fuzzy controller. In addition, external disturbances and uncertainties are not considered in the paper. Chen et al. (2019) used big data to analyze the weather patterns and the congestion evolution patterns that influence the ramp metering control of freeway networks. The weather is clustered into two conditions: normal weather and heavy weather, each of which corresponds to a set of parameters for ALINEA. The congestion evolution has three patterns, and it also determines another parameter to be adjusted. All the analysis are done off-line, and the parameters are prepared in advance. During real implementation, the controller just needs to switch among the preset parameters according to the traffic condition. However, this method needs a large amount of historical data for off-line analysis before it can be applied, which limits its adaptivity to an unknown scenario. In addition, coordinated control is not considered in this method.

An on-line model-free adaptive traffic controller was reported by Smaragdis et al. (2004). Based on the original ALINEA controller, Smaragdis et al. (2004) proposed an adaptive ALINEA (AD-ALINEA) method, in which the real-time measurements of traffic states are used to tune the set-point occupancy of downstream segment, in order to maximize the traffic flow. This method was shown to track the varying critical occupancy levels effectively. However, this method can work only when the states are around the critical point, which is difficult to satisfy in practice. Besides, this is a local ramp metering strategy that does not consider coordinated control. To our best knowledge, no work that uses RL in adaptive parameterized control has been studied yet. As mentioned in Section 1.4, RL-based adaptive parameterized control merges the advantages of current controllers, and it is introduced in the next section.

3. THE PROPOSED RL-BASED ADAPTIVE PARAMETERIZED CONTROL METHOD

In this section, the framework of the proposed algorithm is presented for an intuitive illustration. Then more details are given for the definition of the RL component.

3.1 Framework of the Method

Assume a traffic network is described as a discrete system

$$\boldsymbol{x}(k+1) = F(\boldsymbol{x}(k), \boldsymbol{u}(k), \boldsymbol{d}(k)), \tag{1}$$

where $\boldsymbol{x}$ denotes the state vector of the traffic system, which can include the number of vehicles on the roads, the queue lengths, or the density of traffic flow; F denotes the unknown traffic dynamics, $\boldsymbol{u}$ is the control input vector to the system (e.g., ramp metering rates, variable speed limits, or traffic signal settings), and $\boldsymbol{d}$ is the external disturbances that influence the traffic condition (e.g., weather change or incidents); k is the time step, which is usually taken from 10 s to 1 min according to the traffic system.

For parameterized traffic control, the control input can be written as

$$\boldsymbol{u}(k) = \boldsymbol{f}(\boldsymbol{x}(k), \boldsymbol{\theta}), \tag{2}$$

in which $\boldsymbol{f}$ is the vector-valued function that maps the state to the control input, and $\boldsymbol{\theta}$ contains the parameters of the state-feedback controllers. Considering there are multiple local controllers in a traffic network, $\boldsymbol{u}(k)$ is a combination of multiple control inputs $\boldsymbol{u} = [u_1, \ldots, u_N]^\top$, where N is the number of local controllers and u_i is the corresponding control input for $i = 1, \ldots, N$.

In the traditional parameterized controller, the parameters $\boldsymbol{\theta}$ are fixed. The control input can vary with the state evolution, but it can then not react to a changing environment. For example, when the weather turns from sunny to rainy, which would significantly influence the driver behavior, the controllers that perform well under sunny weather may not suitable for a rainy traffic condition. Therefore, it is necessary to adjust the parameters of the state-feedback controllers during the real-time implementation.

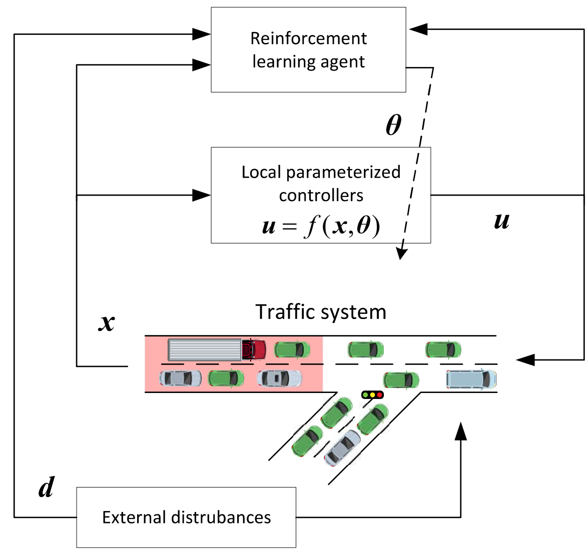

Fig. 1. The diagram of the proposed RL-based adaptive parameterized controller

In our method, a reinforcement learning agent is used to adjust the parameters $\boldsymbol{\theta}$. As shown in Figure 1, the parameterized controllers work at the lower level and with a fast time step (e.g., every 10 s), while the RL agent works at a higher level and with a slower time step (e.g., every 30 min). All the relevant states of the network are given as the input to the RL agent, as well as all the external disturbances. Note that the external disturbances refer to the factors that can be measured, such as the humidity of the surface, a lane close due to an accident, or the light intensity. At every operation time step of the RL agent, all the parameters of the local controllers would be updated in a coordinated way, according to the output of the RL agent. In order to make RL react to the changing environment properly, a training process is necessary and the states, actions, and rewards of RL should be well-defined. In next subsection, several suggestions are given about how to define the RL agent.

3.2 Definition of the RL Agent

All the components connected to the RL agent can be regarded as the environment, and the RL agent interacts with the environment to learn how to perform the best control input by trail and error. The environment can be described as a Markov decision process (MDP), which can be represented by a five-tuple $\langle S, A, P, R, \gamma \rangle$. The state space S, action space A, and reward function R are defined in this section. Moreover, $P = S \times A \times S$ denotes the transmission probability among the states, which is unknown and included implicitly in the environment (e.g., the traffic dynamics F). In addition, $\gamma \in [0, 1)$ is the discount factor on future rewards and user-defined during implementation.

State space All the measurable states related to the local controllers could be given to the RL agent. In addition, the disturbance measurements $\boldsymbol{d}$ and current control input $\boldsymbol{u}(k)$ are also included in the state space.

Action space The dimension of the action space is equal to $\dim(\boldsymbol{\theta})$, and the actions are used to modify $\boldsymbol{\theta}$ values of all the local controllers. Since the actions do not determine the control input $\boldsymbol{u}$ directly, and modify the parameters instead, the performance of the proposed method is in general more stable even during the exploration process.

Reward function The reward function represents the goal of learning. In this method, we consider coordinated control, so the global performance of the traffic network is included in the reward function. A popular choice is the TTS of all the vehicles on the network during a given period of time (i.e., the control sampling time of RL).

Remark 1. Note that the definitions can be adjusted according to the application and the control aim. When choosing the specific RL algorithm, it may be necessary to consider the scale of the target traffic network, and the dimension of the state and action space. For example, the state and action space can either be continuous or discrete. In a simple case, all the states and actions can be discretized as a set of values, each of which represents a traffic condition, such as the degree of congestion or the weather indicator. In this case, the simple Q-learning algorithm (Watkins and Dayan, 1992) may be enough to learn the mapping from action to state. If the states and actions are continuous, then it is recommended to use an actor-critic RL algorithm (Grondman et al., 2012).

The proposed method can be applied to any parameterized controller, including the ones mentioned in Section 1 and 2. Next, a case on RL-based adaptive ALINEA for coordinated ramp metering control is presented.

4. A CASE STUDY: RL-BASED ADAPTIVE ALINEA

In this section, our method is illustrated by a case of ramp metering control for a freeway network using ALINEA. First, the basic equations of ALINEA are introduced. Then the target freeway network and the simulation settings are presented, and the state and action spaces of RL are defined. Finally, the training results and evaluation results of our method and the comparison methods (no-control, standalone RL, ALINEA) are analyzed.

4.1 RL-based Adaptive ALINEA

In original ALINEA method (Papageorgiou et al., 1991), traffic occupancy is used as the state for control. Here we use the traffic density instead, because this state reflects the same traffic condition and is easier to measure in the traffic simulator. The basic equation of ALINEA reads

$$r(k) = r(k-1) + K_{\mathrm{R}}(\bar{\rho} - \rho_{\mathrm{d}}(k-1)), \tag{3}$$

where $k = 1, 2, \ldots$ is the discrete time step; $r(k)$ is the ramp metering rate to be implemented during the next time step k; K_{R} is a positive gain parameter; $\bar{\rho}$ is the desired downstream density, which is usually taken as the critical density ρ_{cr}; $\rho_{\mathrm{d}}(k-1)$ is the measured downstream density at last time step $k-1$. In the traditional ALINEA method, the parameters K_{R} and $\bar{\rho}$ are fixed. Therefore, the actions of the RL agent consist of the values of these two parameters, which is denoted as K_{rl} and $\bar{\rho}_{\mathrm{rl}}$ to replace the original K_{R} and $\bar{\rho}$.

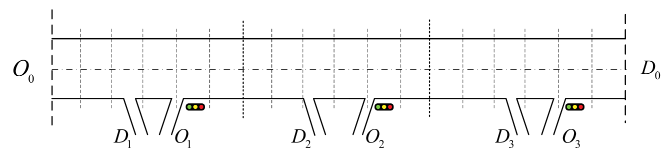

Fig. 2. The freeway network for the case study

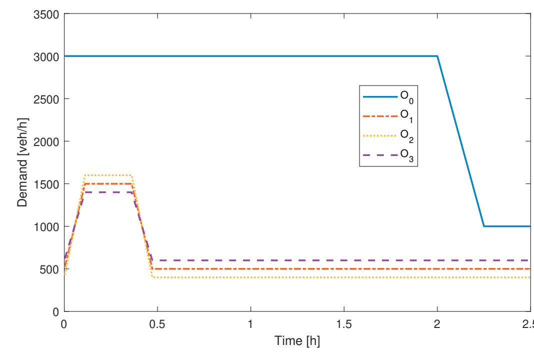

Fig. 3. The rush hour demands for the case study

4.2 Network Settings and RL Definitions

In this case study, we consider a freeway network from Liu et al. (2022), as shown in Figure 2. This freeway network is divided into 18 segments, each of which has a length of 1 km. There are 3 on-ramps (O_1, O_2, O_3) that are under control $(N = 3)$, 3 off-ramps (D_1, D_2, D_3) that are unrestricted, one mainstream origin (O_0), and one unrestricted mainstream destination (D_0). The mainstream stretch has two lanes, while the on-ramps have one lane. For the sake of simplicity, queue constraints are not considered. METANET is used to formulate the freeway network. We refer the reader to Kotsialos et al. (2002a) for more details on METANET. The parameters of METANET for this freeway network are taken from Liu et al. (2022), as shown in Table. 1.

Table 1. METANET parameters for the freeway network

C_{main}	C_{onramp}	τ	κ
2000 veh/h/lane	2000 veh/h/lane	18 s	40 veh/h/lane
η	a_m	σ	v_{free}
60 km^2/h	1.867	0.0122	102 km/h
ρ_{cr}	ρ_{max}	α	T
37.5 veh/km/lane	180 veh/km/lane	0.1	10 s

The typical rush hour traffic demands are considered for a 2.5 hours simulation time interval, as shown in Figure 3. The turning rates of all the three off-ramps are fixed as 5% of the mainstream flow. To reproduce the stochastic phenomena of the traffic network, a random noise e_{v} with Gaussian distribution is added to the velocity update equation of the three on-ramp downstream segments. When the density of the downstream segment is below 30 veh/km/lane, the noise is $e_{\mathrm{v}} \sim \mathcal{N}(-1, 1)$; when the density is between 30-40 veh/km/lane, $e_{\mathrm{v}} \sim \mathcal{N}(-3, 2)$; when the density is above 40 veh/km/lane, $e_{\mathrm{v}} \sim \mathcal{N}(-5, 3)$. This means that the drivers tend to decrease their speed when the traffic is crowded. In addition, the weather condition

Table 2. METANET parameters under different weather conditions

Parameter	Good	Bad	Extreme
ρ_{cr} (veh/km/lane)	37.5	29.5	21.5
v_{free} (km/h)	102	91.8	81.6
τ (s)	18	19.8	21.6

is also considered explicitly, which is divided into three levels: good weather, bad weather, and extreme weather. Each weather condition corresponds to a set of parameters that influence the traffic system, as shown in Table 2.

In this case study, we choose continuous state space and discrete action space. Therefore, deep Q-network (DQN) (Mnih et al., 2013) is a suitable RL algorithm for this case. The state space is defined as $s = [W, d_{O_0}, d_{O_i}, \rho_{O_i}, q_{O_0}, q_{O_i}, r_{O_i}, i = 1, 2, 3]^\top$, where $W \in \{\text{good}, \text{bad}, \text{extreme}\}$ is a discrete indicator of the weather condition, d_{O_1} is the demand of mainstream and d_{O_i} is the demand of on-ramp O_i, ρ_{O_i} is the downstream density of on-ramp O_i, q_{O_i} is the queue length of on-ramp O_i, and r_{O_i} is the previous ramp metering rate of on-ramp O_i. Therefore, the dimension of the state space is 15. For simplicity and a better learning process, the action space is reduced as much as possible. The three on-ramp metering installations share the same parameters K_{rl} and $\bar{\rho}_{\text{rl}}$, and $K_{\text{rl}} \in \{0.01, 0.001\}$ (high gain and low gain) while $\bar{\rho}_{\text{rl}} \in \{23.5, 31.5, 39.5, 47.5\}$. These parameter values are chosen based on experience. The dimension of action space is then 8, and the RL update the parameters every 15 min.

We consider three weather scenarios in total, in which the weather is fixed as bad for the first simulation hour. After one hour, the weather can change to good or extreme weather or keep as bad weather. So the weather scenarios are bad-good weather (scenario 1), bad-bad weather (scenario 2), and bad-extreme weather (scenario 3). The proposed RL-based adaptive ALINEA method (or for short: RL-based ALINEA) is trained for these three scenarios, as well as a RL-based ramp metering control method (RL-based RM). For the latter method, the RL algorithm and parameters are the RL-based ALINEA, except for the action space. The RL-based RM directly output the ramp metering rates. For simplicity, the actions for all the three ramp metering installations are the same, and the ramp metering rates are discretized into 11 values distributed equidistantly between 0 and 1. The control sampling time is 1 min.

4.3 Simulation Results and Discussion

To show the learning process better, the episode rewards of RL-based ALINEA and RL-based RM for three weather scenarios are presented separately, as shown in Figures 4 to 6. Note that the episode reward is the negative value of the TTS for each run of simulation. Obviously, the RL-based ALINEA method has a better sample efficiency than RL-based ramp metering (i.e., the RL-based ALINEA method converges faster). In addition, RL-based ALINEA has a higher initial episode reward, which means the proposed method can even provide control performance during exploration. The episode rewards variation of RL-based ALINEA during training is also smaller than that of RL-based RM.

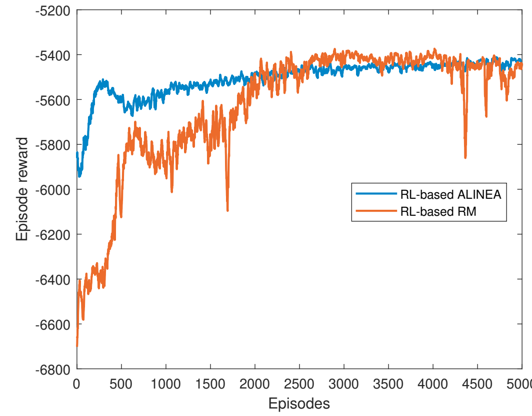

Fig. 4. Learning process for weather scenario 1

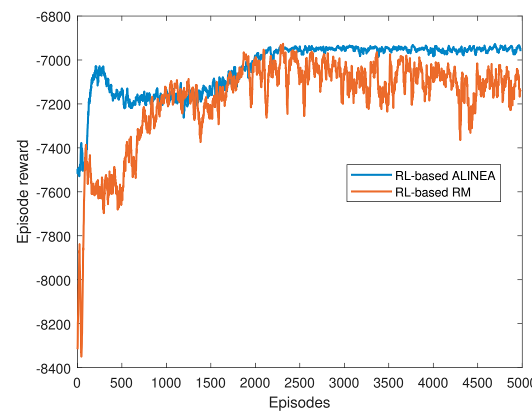

Fig. 5. Learning process for weather scenario 2

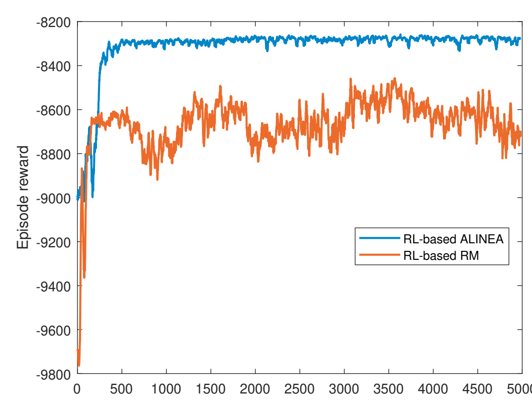

Fig. 6. Learning process for weather scenario 3

The TTS simulation results are presented in Table 3, including the no-control case, ALINEA (with $\bar{\rho} = 37.5$ veh/km/lane), trained RL-based RM, and the trained RL-based ALINEA. All the methods are evaluated for the three weather scenarios, and the TTS values are averaged over 10 independent runs for each scenario. The results show that RL-based ALINEA can achieve the best performance in terms of TTS among all the control methods.

Table 3. TTS simulation results of all the comparison control methods

	Scenario 1	Scenario 2	Scenario 3
No control	6793.9	8496.4	9848.9
ALINEA	5505.7	7120.2	8797.2
RL-based RM	5408.1	7005.4	8495.2
RL-based ALINEA	5379.1	6987.6	8267.3

5. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a reinforcement learning (RL)-based adaptive parameterized control method for coordinated traffic management. The proposed method is cheap to implement without requiring significant online computational efforts. Compared with traditional parameterized control, our method can adapt to various unknown traffic conditions. In contrast to standalone RL-based control, our method has a better sample efficiency and can learn faster. Results of the case study show the effectiveness of the method.

For future work, an extensive case study can be conducted by considering more complex environments and scenarios. In addition, multi-agent RL can be employed to handle a larger network by coordinating several local RL agents.

REFERENCES

Abdulhai, B., Pringle, R., and Karakoulas, G.J. (2003). Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering*, 129(3).

Camacho, E.F. and Alba, C.B. (2013). *Model Predictive Control*. Springer science & business media.

Chen, J., Lin, W., Yang, Z., Li, J., and Cheng, P. (2019). Adaptive ramp metering control for urban freeway using large-scale data. *IEEE Transactions on Vehicular Technology*, 68(10), 9507–9518.

Chu, T., Wang, J., Codecà, L., and Li, Z. (2019). Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems*, 21(3), 1086–1095.

De Schutter, B. and De Moor, B. (1998). Optimal traffic light control for a single intersection. *European Journal of Control*, 4(3), 260–276.

Gartner, N.H., Little, J.D., and Gabbay, H. (1976). Simultaneous optimization of offsets, splits, and cycle time. *Transportation Research Record*, 596, 6–15.

Ghods, A.H., Kian, A.R., and Tabibi, M. (2009). Adaptive freeway ramp metering and variable speed limit control: a genetic-fuzzy approach. *IEEE Intelligent Transportation Systems Magazine*, 1(1), 27–36.

Grondman, I., Busoniu, L., Lopes, G.A., and Babuska, R. (2012). A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 1291–1307.

Hegyi, A., De Schutter, B., and Hellendoorn, H. (2005). Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies*, 13(3), 185–209.

Jeschke, J., Sun, D., Jamshidnejad, A., and De Schutter, B. (2023). Grammatical-evolution-based parameterized model predictive control for urban traffic networks. *Control Engineering Practice*, 132, 105431.

Kotsialos, A. and Papageorgiou, M. (2004). Motorway network traffic control systems. *European Journal of Operational Research*, 152(2), 321–333.

Kotsialos, A., Papageorgiou, M., Diakaki, C., Pavlis, Y., and Middelham, F. (2002a). Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET. *IEEE Transactions on Intelligent Transportation Systems*, 3(4), 282–292.

Kotsialos, A., Papageorgiou, M., Mangeas, M., and Haj-Salem, H. (2002b). Coordinated and integrated control of motorway networks via non-linear optimal control. *Transportation Research Part C: Emerging Technologies*, 10(1), 65–84.

Li, Z., Liu, P., Xu, C., Duan, H., and Wang, W. (2017). Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks. *IEEE Transactions on Intelligent Transportation Systems*, 18(11), 3204–3217.

Liu, S., Sadowska, A., and De Schutter, B. (2022). A scenario-based distributed model predictive control approach for freeway networks. *Transportation Research Part C: Emerging Technologies*, 136, 103261.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., and Wang, Y. (2003). Review of road traffic control strategies. *Proceedings of the IEEE*, 91(12), 2043–2067.

Papageorgiou, M., Hadj-Salem, H., Blosseville, J.M., et al. (1991). Alinea: A local feedback control law for on-ramp metering. *Transportation Research Record*, 1320(1), 58–67.

Papageorgiou, M. and Kotsialos, A. (2002). Freeway ramp metering: An overview. *IEEE Transactions on Intelligent Transportation Systems*, 3(4), 271–281.

Smaragdis, E., Papageorgiou, M., and Kosmatopoulos, E. (2004). A flow-maximizing adaptive local ramp metering strategy. *Transportation Research Part B: Methodological*, 38(3), 251–270.

van de Weg, G.S., Hegyi, A., Hoogendoorn, S.P., and De Schutter, B. (2018). Efficient freeway mpc by parameterization of alinea and a speed-limited area. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 16–29.

van Kooten, R., Imhof, P., Brummelhuis, K., van Pampus, M., Jamshidnejad, A., and De Schutter, B. (2017). ART-UTC: An adaptive real-time urban traffic control strategy. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 1547–1552.

Wang, C., Xu, Y., Zhang, J., and Ran, B. (2022). Integrated traffic control for freeway recurrent bottleneck based on deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*.

Watkins, C.J. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3), 279–292.

Zegeye, S.K., De Schutter, B., Hellendoorn, J., Breunesse, E.A., and Hegyi, A. (2012). A predictive traffic controller for sustainable mobility using parameterized control policies. *IEEE Transactions on Intelligent Transportation Systems*, 13(3), 1420–1429.