# Optimization of the Operation for the Moving Light Guide System at an Expressway Bottleneck Based on Offline Reinforcement Learning

Fumiya OKANO [a], Yasuhiro SHIOMI [b], Masahiro KINOSHITA[c]

[a] *Graduated Student, Graduate School of Science and Engineering, Ritsumeikan University, Shiga, 525-8577, Japan; E-mail: ce0054xp@ed.ritsumei.ac.jp*

[b] *Professor, Ritsumeikan University, Shiga, 525-8577, Japan; ORCid numbers: 0000-0002-7810-0971, E-mail: shiomi@fc.ritsumei.ac.jp*

[c] *Hanshin Expressway R&D Company Limited, Osaka, 530-6123, Japan, ORCid numbers: 0000-0003-3445-0942, Email: kinoshita-masahiro_gktc24002@hanshin-tech.co.jp*

## 1 Introduction

The Moving Light Guide System (MLGS) is an active traffic management (ATM) method used on expressways in Japan. This system aims to mitigate speed reductions and promote speed recovery by adjusting the lighting speed (for more details, see https://x.gd/AIDPq). Over the past decades, significant research efforts have been devoted to identifying efficient operation methods for MLGS [1]. However, existing studies have primarily relied on empirically derived operation algorithms based on naïve trial-and-error approaches, which are labor-intensive and provide no clear evidence that the derived algorithm is superior to other alternatives.

In recent years, reinforcement learning (RL) has been applied to optimize ATM strategies such as variable speed limits [2]. However, most of these studies rely solely on traffic simulations due to the challenges associated with real-world field experiments. In the case of MLGS, even simple RL-based approaches cannot be directly applied because no existing traffic simulation models account for the system's impact on driving behavior.

Nevertheless, through extensive trial and error in the past and practical operation logs, a substantial MLGS dataset has been accumulated. This dataset provides an opportunity to leverage offline RL (offline RL) [3], which enables policy optimization using pre-collected data. In this study, we employ offline RL to develop a data-driven approach for optimizing MLGS operation. Since offline RL learns optimal action selection without requiring new real-world experiments, this method allows us to optimize MLGS operations without explicitly modeling its effects on vehicle behavior.

## 2 Methodology

RL is a machine learning approach that enables agents to learn optimal actions through interaction with an environment. It can be broadly categorized into online RL and offline RL. Online RL requires trial-and-error exploration in a real-world environment, which entails high costs and risks due to the need for experimentation.

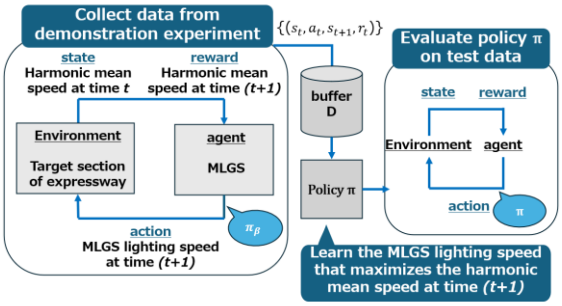


Figure 1. Schematic diagram of offline RL

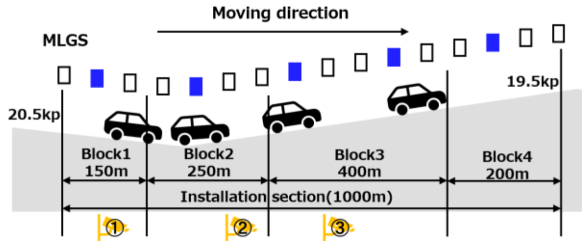


Figure 2. Definition of blocks at the target section

In contrast, offline RL learns optimal actions from pre-collected datasets and generally requires a large volume of data for effective learning. Since MLGS has accumulated extensive operational data over its long history of implementation, offline RL allows us to optimize its operation without the need for additional real-world experiments. To achieve this, we developed a program that determines the optimal lighting speed based on the traffic conditions of a given target section using offline RL. Figure 1 presents a schematic diagram of the offline RL framework used in this study. First, we collected operational data from MLGS experiments and established an environment where the system state and reward were defined. Specifically, we set the harmonic mean speed within the target section at time t as the state, while the harmonic mean speed at time t+1 served as the reward, as it directly reflects improvements in traffic throughput, particularly in bottleneck sections. The RL agent then determined the lighting speed based on the observed state. An offline RL algorithm was applied to learn a policy that maximizes the harmonic mean speed by adjusting the lighting speed using the collected dataset. For this study, we employed Discrete Conservative Q-Learning (Discrete CQL) [3] due to its strong performance in environments with complex data distributions. Additionally, Discrete Behavioral Cloning (Discrete BC) [4], an algorithm that imitates actions observed in the dataset, was used as a baseline for comparison.

## 3 Target Section and Data Overview

The target section in this study is a 1,000-meter segment of the Hanshin Expressway No. 3 Fukae Sag section (20.5 kp–19.5 kp) in Japan, where the Moving Light Guide System (MLGS) has been installed and operated. This section experiences recurrent traffic congestion due to its sag structure. It was selected as the study's target because a demonstration experiment was conducted to explore an efficient MLGS operation method [1], leading to the accumulation of experimental data logs from various operation strategies between June 2015 and July 2017. Since traffic conditions vary depending on the road alignment, the target section is divided into four blocks, as illustrated in Figure 2. Three cameras are installed to monitor traffic conditions in each block, and the lighting speed is adjusted dynamically for each block based on the monitored speed. The lighting speed is updated every minute and takes discrete values. The monitored harmonic mean speed, recorded by the cameras, is used as both the state and reward in the offline RL framework.

## 4 Results

To apply the offline RL, first the hyperparameters are set based on the results of off-policy evaluation, and then the model was separately trained for each of the four blocks. During the training process, we observed that the loss values gradually decreased and eventually converged, indicating stable learning across all blocks. Furthermore, since Discrete CQL exhibited lower loss values than Discrete BC, it suggests that Discrete CQL enables more effective learning.

Figure 3 shows the relationship between the observed harmonic mean speed at time $t$ and optimal lighting speed at $t+1$ derived by the trained policy. The results indicate that stepwise increases in lighting speed with rising harmonic mean speed were effective across all blocks, which involved modifying the lighting speed thresholds used in the current operational method. In Blocks 1 and 2, where speed decreases due to sag effects, the optimal lighting speed closely matched the vehicle speed. In Blocks 3 and 4, where speed recovers and stabilizes, setting higher lighting speeds downstream from the congestion front proved optimal. Particularly, in Block 3, a more precise adjustment of lighting speed was found to be most effective. To assess the effectiveness of the optimal policy, we calculated Q-values, which represents the expected reward for each action. It allows policy performance evaluation for various harmonic mean speeds and lighting speeds using test data. Figure 4 shows the Q-values for each block, illustrating the relationship between harmonic mean speed and lighting speed. The results demonstrated that optimal policy's actions consistently yielded higher Q-values across nearly all harmonic mean speeds, confirming the policy's effectiveness. These findings suggest that the proposed optimized operation method can improve vehicle speed and enhance congestion mitigation through MLGS.

## 5 Conclusion

This study demonstrates that stepwise increases in lighting speed in response to harmonic mean speed can enhance MLGS operation. Moreover, in congestion zones, setting the lighting speed to approximately match vehicle speed and gradually increasing it downstream proved most effective. In the future work, the multi-agent RL will be applied to enhance the collaborative optimization among the blocks.

## References

[1] Masumoto, H., Tobigaya, A., Kodama, T., Kitazawa, T., Suzuki, K., Tomoeda, Y., & Lee, R. Verification of the effectiveness of speed recovery guidance lights on the Hanshin Expressway and considerations for efficient operation methods. Journal of Traffic Engineering, 4(3) (Special Issue B), pp. B1-B9.2018. (in Japanese)

[2] Yu Han, Andreas Hegyi, Le Zhang, Zhengbing He, Edward Chung, Pan Liu. A new reinforcement learning-

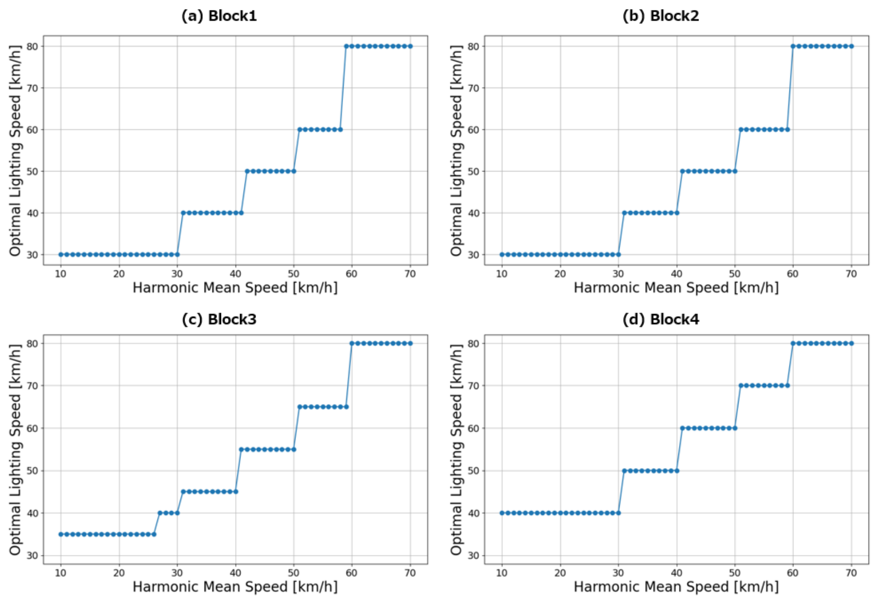


Figure 3. Relationship between harmonic mean speed and optimal lighting speed

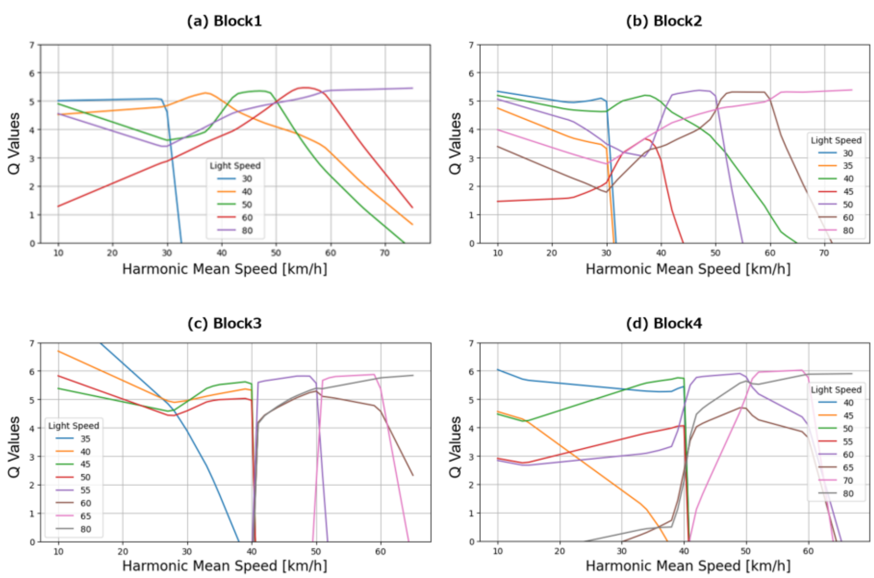


Figure 4. Q-Values for different lighting speeds across blocks


based variable speed limit control approach to improve traffic efficiency against expressway jam waves. Transportation Research Part C: Emerging Technologies, Vol. 144, 103900, 2022.

[3] Levine, S., Kumar, K., Tucker, G., Fu, J.: Conservative Q-Learning for Offline Reinforcement Learning. Advances in Neural Information Processing Systems 33 (NeurIPS 2020), pp.1179-1191, 2020.

[4] Torabi, F., Warnell, G., and Stone, P. Behavioral cloning from observation. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18). AAAI Press, 4950–4957, 2018.