The following publication L. Yao, Z. Leng, J. Jiang and F. Ni, "Large-Scale Maintenance and Rehabilitation Optimization for Multi-Lane Highway Asphalt Pavement: A Reinforcement Learning Approach," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 11, pp. 22094-22105, Nov. 2022 is available at https://dx.doi.org/10.1109/TITS.2022.3161689.

Large-Scale Maintenance and Rehabilitation Optimization for Multi-Lane Highway Asphalt Pavement: A Reinforcement Learning Approach

Linyi Yao, Zhen Leng, Jiwang Jiang, Fujian Ni

Abstract—**Pavement maintenance and rehabilitation (M&R) optimization is of great importance to the sustainable development of roadway infrastructure. Various models have been developed for supporting M&R decision-making. However, there is still a lack of research that can provide lane-specific M&R strategies for large-scale pavement networks which consider uncertainty in optimization to achieve management flexibility. This study proposes an innovative M&R optimization approach for multi-lane highway pavement based on a reinforcement learning (RL) method. Life cycle assessment (LCA) and life cycle cost analysis (LCCA) were integrated to assess the environmental and economic impact of M&R decisions, respectively. The uncertainty of pavement deterioration was considered by constructing an RL simulation environment that contains several probabilistic pavement performance models. The proposed method was applied to a large-scale real-world highway network as demonstration, and compared with the state-of-the-practice hierarchical threshold-based approach (HT). The results show that the RL model saved about 26.59% of the cost in comparison to the HT approach, which was equal to 18147.27 million CNY. It could keep the long-term pavement performance within an acceptable range in a cost-effective manner. The RL model tends to select less rehabilitations and more preventive maintenance than the HT model. It was also found that incorporating uncertainty into optimization allows the model to balance the expected return and the negative (risk) and positive (opportunity) uncertainty of the solution. The outcomes of this study are expected to improve the current pavement management practice and demonstrate the potential of RL in pavement M&R optimization.**

Index Terms—**Pavement maintenance optimization; reinforcement learning; lane-specific solution; large-scale pavement network; managerial flexibility**

I. Introduction

Road pavement inevitably deteriorates over time under the repeated traffic loading and joint impact of climatic factors. The deteriorated pavement would incur high maintenance and rehabilitation (M&R) cost, extra fuel consumption and greenhouse gas (GHG) emissions, and increased vehicle operating cost, etc., which significantly affect the sustainable development of roadway infrastructure. Improving the cost-effectiveness of M&R activities while maintaining the pavement in good conditions has long been the concern of highway agencies. Recent years, reducing the environmental impact through effective management of pavement infrastructure is also an important consideration in pavement maintenance planning [1]-[3]. However, the limited resource, expanding road network, rapid aging of pavement, and mutually conflicting goals impose great challenges to decision-makers.

The methodological frameworks for determining the optimal pavement M&R strategy for the whole road network can be broadly classified into top-down, two-stage bottom-up (TSBU), and simultaneous network optimization (SNO) [4]. The top-down approach divides the pavement network into groups and applies the same M&R strategy to segments within each group [5]. Though computationally efficient, this method ignores the segment-specific features, and thereby cannot provide segment-specific decisions. By contrast, the bottom-up approach derives segment-specific M&R strategies by decomposing the problem into two stages. In the first stage (i.e., segment-level), the segment-optimal M&R activities are selected for each segment, which are then considered in the second stage (i.e., system-level or network-level) to determine the system-optimal M&R strategy that optimizes the system-wide objectives while meeting the resource constraints [6]. However, the main drawback of TSBU is that it does not account for future resource limitations [4]. Another alternative is the SNO approach which selects the optimal segment-level strategy and solves the network-level resource allocation problem simultaneously. Although the SNO method takes into account the interdependency of distinct segments (e.g., the shared budget), it is computationally expensive. This study falls within the TSBU framework for two reasons: 1) the future available budgets are usually highly uncertain, depending on the revenues from multiple sources, and 2) the TSBU approach is more computationally tractable than SNO, thereby allowing more detailed formulation of the problem. Specifically, this study aims to solve the segment-level optimization problem in the TSBU framework as there remains room for improvement in this area.

Table I presents a summary of the representative studies in pavement management that have attempted to optimize the segment-level M&R strategy while providing segment-specific solutions. These studies formulate the segment-level optimization problem with various objectives, performance indicators, M&R options, deterioration models, planning horizon lengths, etc. The solution methods generally fall into two categories: mathematical programming and heuristic approaches. Mathematical programming could guarantee the

This work was supported in part by the Research Institute for Sustainable Urban Development (RISUD) at the Hong Kong Polytechnic University
(Corresponding author: Zhen Leng)

Linyi Yao, Zhen Leng, and Jiwang Jiang are with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnique University, Kowloon, Hong Kong (e-mail: linyi.yao@connect.polyu.hk, zhen.leng@polyu.edu.hk, jiwang.jiang@polyu.edu.hk).

Fujian Ni is with the School of Transportation, Southeast University, Nanjing 211189, China (e-mail: nifujian@gmail.com).

TABLE I
SUMMARY OF EXISTING STUDIES IN SEGMENT-LEVEL PAVEMENT MAINTENANCE OPTIMIZATION WITH SEGMENT-SPECIFIC SOLUTIONS

Studies	Optimization objectives				Pavement performance indicators	M&R options	Horizon	Optimization algorithm	Action forms				Uncertainty
	Agency cost	User cost	Environmental impact	Pavement condition					A	B	C	D	
Ouyang and Madanat [7]	√	√			Roughness ^{c}	Overlay thickness ^{c}	100	Calculus of variations		√			
Gao and Zhang [8]	√				IRI ^{c}	4	20	Robust optimization				√	√
Zhang *et al.* [9]	√	√	√		DI ^{d}	3	40	DP	√				√
Irfan *et al.* [10]	√	√		√	IRI ^{c}	5	25~32	MINLP				√	√
Yeo *et al.* [6]	√	√			PSR ^{d}	3	40	DP				√	
Yu *et al.* [11]	√	√	√		PSI ^{d}	3	40	DP	√				√
Yu *et al.* [2]	√	√	√	√	PSI ^{c}	5	40	GA	√				
Bai et al. [12]	√	√			IRI ^{c}	1	20	DP	√				
Lee and Madanat [13]	√	√			IRI ^{d}	6	infinite	DP		√			
Lee and Madanat [14]	√	√			IRI ^{c}	M&R intensity ^{c}	infinite	MINLP		√			
Hadiwardoyo *et al.* [15]	√			√	SDI ^{c}	5	5	GA				√	
Santos *et al.* [1]	√	√	√		CCI ^{c}, IRI ^{c}	7	50	GA	√				
Zhang *et al.* [16]	√	√			IRI $^{d, c}$	3	infinite	DP, greedy heuristic				√	
Santos *et al.* [17]	√				CCI ^{c}	6	30/50	GA	√				
Heidari *et al.* [18]	√	√	√		PSI ^{d}	8	35	DP	√				√
Guo *et al.* [19]	√	√			IRI ^{c}	8	20	Backtrack-search				√	√
Yao *et al.* [20]	√			√	RD ^{c}, IRI ^{c}, SFC ^{c}, and PDCI ^{c}	38	15	Deep Q-learning			√		
Renard *et al.* [21]			√		IRI ^{d}	7	50	Q-learning	√				√
Shani *et al.* [22]	√	√	√		IRI ^{d}	7	20/30	Q-learning	√				√
Han *et al.* [23]	√			√	PCI ^{c}, RDI ^{c}, RQI ^{c}, and SRI ^{c}	14	30	PPO				√	

Note: ^{c} = continuous; ^{d} = discrete; A= applying the same M&R treatment to the entire section; B= testing only on one-lane highway or one lane of a multi-lane highway; C= lane-specific M&R treatment; D= not mentioned; MINLP= mixed-integer nonlinear program; DI= distress index; IRI= international roughness index; SDI= Surface Distress Index; CCI= critical condition index; PSR= pavement serviceability rating; PSI= present serviceability index; PCI= pavement surface condition index; RDI= pavement rutting depth index; RQI= pavement riding quality index; SRI= pavement skidding resistance index; RD= rutting depth; SFC= side-way force coefficient; PDCI= pavement distress condition index; PPO= proximal policy optimization;

global optimality of the solution. The most commonly used mathematical programming approach in segment-level optimization is dynamic programming (DP) [6] [13][18], as it is essentially a sequential decision problem. However, traditional DP suffers heavily from the so-called "curse of dimensionality", with limited applicability to problems that have high-dimensional state space or action space. Common solutions include reducing the state variables to only contain the current pavement conditions, discretizing the continuous state variables, limiting the number of actions, etc. However, these simplifications also cause other problems. For example, several studies have demonstrated the necessities of developing history-dependent deterioration model and incorporating the history variables (e.g., pavement age) into the state space [14][16][24]. It was also reported that the state discretization would lead to significant information loss which is inversely proportional to the number of discrete states [25]. In addition, most studies only considered limited M&R options, and usually described these options in general terms, e.g., maintenance, preservation, rehabilitation, reconstruction, etc., so as to reduce the solution space [16][24]. This simplification does not recognize the characteristics of different M&R activities, and thus may introduce considerable bias. On the other hand, heuristic methods are sometimes deemed to be more promising in solving problems with large solution space due to their better computational efficiency. Genetic algorithm (GA) is one of the heuristic methods most commonly used in pavement management [1][26]. However, GA could only find near optimal solutions, and is prone to converge to the local optimum in some cases. Therefore, improved GA techniques and other new heuristic algorithms have also been introduced to pavement maintenance decision-making [27].

Recently, some researchers leveraged reinforcement learning (RL) to solve the segment-level maintenance optimization problem [20]-[23]. RL is a subfield of machine learning (ML) where the intelligent agent learns an optimal policy through interaction with the environment to maximize long-term cumulative rewards [28]. Compared with other optimization algorithms, RL learns an optimal policy that maps the states to actions or learns a value function that maps the states to the expected returns of a specific state-action pair. It takes advantage of individual behavioral interactions which enables more efficient search [28]. RL has also been proven by many studies to provide flexibility for decision-making [21][29]. Managerial flexibility, which is defined as the available flexibilities for decision-makers to alter their policies as new information arrives, has gained increasing attention in pavement community in recent years [18][30]. A typical example is the ability of decision-makers to quickly adjust their M&R strategies when future pavement conditions are different from expectations. Managerial flexibility in pavement management is often achieved by combining uncertainty consideration with multiple M&R and construction alternatives [18]. RL can not only simultaneously address these two needs, but its mapping structure also makes it easier to adjust the strategy in future years. By inputting new state variables into the policy network or value function, the updated strategy could be obtained without the need to reconstruct the optimization algorithm [21].

Despite the contributions made by previous researchers, there are still limitations. Firstly, very few studies have dealt with the lane-specific M&R decisions. Most of the existing studies either apply the same M&R treatment to the entire road section or test the algorithm only on one-lane highway or one lane of a multi-lane highway. It should be noted that due to the uneven distribution of traffic over multiple lanes, the pavement conditions of different lanes in the same road section may vary significantly. Hence, applying the same M&R treatment to the entire road section may result in a waste of resources for some lanes, and insufficient maintenance for the others. Besides, the actions of adjacent lanes are interdependent because of the vertical restriction [31]. For instance, overlay without milling the existing pavement would add the pavement elevation, so it must be implemented on the entire road section to maintain the same elevation. To address this problem, this study will use the joint action of multiple lanes in the same road section for modelling. Though such method exponentially increases the computational complexity, a subtype of RL algorithm was introduced to tackle the challenge. Secondly, previous studies that resort to RL to solve the segment-level optimization problem only learn from a small number of segments or a single highway. This will lead to an inadequate exploration of the actual state space (i.e., the information required to make an informed choice in the real world), and the learnt policy cannot be directly applied to other segments in the road network.

This study developed a decision support tool that aims to improve the sustainability of roadway infrastructure while providing lane-specific M&R strategies for large scale pavement networks and increasing the flexibility of pavement management. Life cycle assessment (LCA) and life cycle cost analysis (LCCA) methods were integrated to assess the sustainability performance of the M&R decision. Twin Delayed Deep Deterministic policy gradient algorithm (TD3) [32] was combined with the Wolpertinger Policy [33] to achieve lane-specific decision-making with vertical restriction. The learning was performed on a large number of real-world road segments, and a complex state space was built, allowing the policy model to be applied to a large-scale pavement network. Managerial flexibility was embedded in the model to proactively adapt to the uncertain future pavement deterioration. While previous studies have proven the efficacy of utilizing RL to enhance flexibility [21], this study further expanded upon this by using the TD3-Wolpertinger algorithm to replace the tabular Q-learning which is a bit cumbersome especially in high-dimensional state space.

II. Problem Formulation

The problem in this study is a segment-level system-wide pavement maintenance optimization problem. It is supposed that there is a road network consisting of N_s one-way pavement segments with 2 to 4 lanes (in one direction). The segment set is denoted by $S, S = \{1,2,\ldots,N_s\}$. The optimization objective is to minimize the sum of discounted agency costs (AC), user costs (UC) and environmental damage costs (EDC) for each segment $n \in S$ over a given planning horizon T, as shown in Eq.1a. AC_{tnl} represents the agency cost of a certain lane l in segment n in year t, which is a function of the selected M&R activity (including do-nothing) x_{tnl} and consists of material, machine, and labor costs. UC_{tnl} signifies the user costs, including the costs incurred during the usage phase as well as the work zone (WZ) traffic management phase. Environmental impacts are assessed by converting the GHG emissions associated with material extraction and production, construction, transportation, and pavement-vehicle interaction (PVI) into monetary values, i.e., environmental damage cost (EDC) [11]. Three constraints are imposed to avoid ineffective exploration and improve algorithm efficiency, as shown in Eq.1b~1d. Constraint 1b specifies that a maintenance or rehabilitation treatment must be selected when any of the pavement performance indicators (PPI) falls below (for non-increasing PPI, e.g., side-way force coefficient (SFC)) or rise above (for non-decreasing PPI, e.g., international roughness index (IRI)) the threshold. Constraint 1c enforces that a rehabilitation treatment must be implemented when the threshold of any indicators is reached. Constraint 1d guarantees that no M&R activity is performed if all PPIs are very close to brand new conditions.

$\forall n \in S$, minimize:

$$\sum_{t=1}^{T}\sum_{l=1}^{L_n}[\gamma^{t-1}(AC_{tnl} + UC_{tnl} + EDC_{tnl})] \tag{1a}$$

subject to:

$$\exists i \in I, PPI_{itnl} < or > PPI_{iM\&R}: x_{tnl} \in A_{M\&R} \tag{1b}$$

$$\exists i \in I, PPI_{itnl} < or > PPI_{iR}: x_{tnl} \in A_R \tag{1c}$$

$$\forall i \in I, |PPI_{itnl} - PPI_{inew}| \le \epsilon_i: x_{tnl} = \text{do-nothing} \tag{1d}$$

where L_n= number of lanes in segment n; γ= discount factor; i= index of PPI; I= set of PPI index; PPI_{itnl}= the i^{th} PPI of lane l in segment n in year t; $PPI_{iM\&R}$= the threshold of the i^{th} PPI that a maintenance or rehabilitation treatment must be selected; PPI_{iR} = the threshold of the i^{th} PPI that a rehabilitation treatment must be selected; $A_{M\&R}$= set of M&R treatments; A_R= set of rehabilitation treatments; PPI_{inew}= the brand new condition of the i^{th} PPI; ϵ_i= the distance between the value of the i^{th} PPI and its brand new condition for mandatory do-nothing.

III. Methodology

Fig. 1 presents an overview of the present study with a general presentation of the used methods. The pavement maintenance optimization problem was solved through both our innovative TD3-Wolpertinger algorithm and the state-of-the-practice hierarchical threshold-based approach (HT). A set of probabilistic pavement performance models developed by a recently published paper [34] were used for predicting the future pavement deterioration. The integrated LCA-LCCA method was adopted to quantify the cost of applying the selected M&R actions. To demonstrate the validity and performance of the proposed method, we compared the performance of the TD3-Wolpertinger algorithm with the HT approach in terms of the cost savings, pavement performance evolutions and treatment type distributions. Monte Carlo simulation (MCs) was performed to obtain the cumulative reward distributions of the selected actions and to provide some insight into the trade-off between the expected return and the uncertainty of the actions. The following sections further detail the techniques employed in this study.

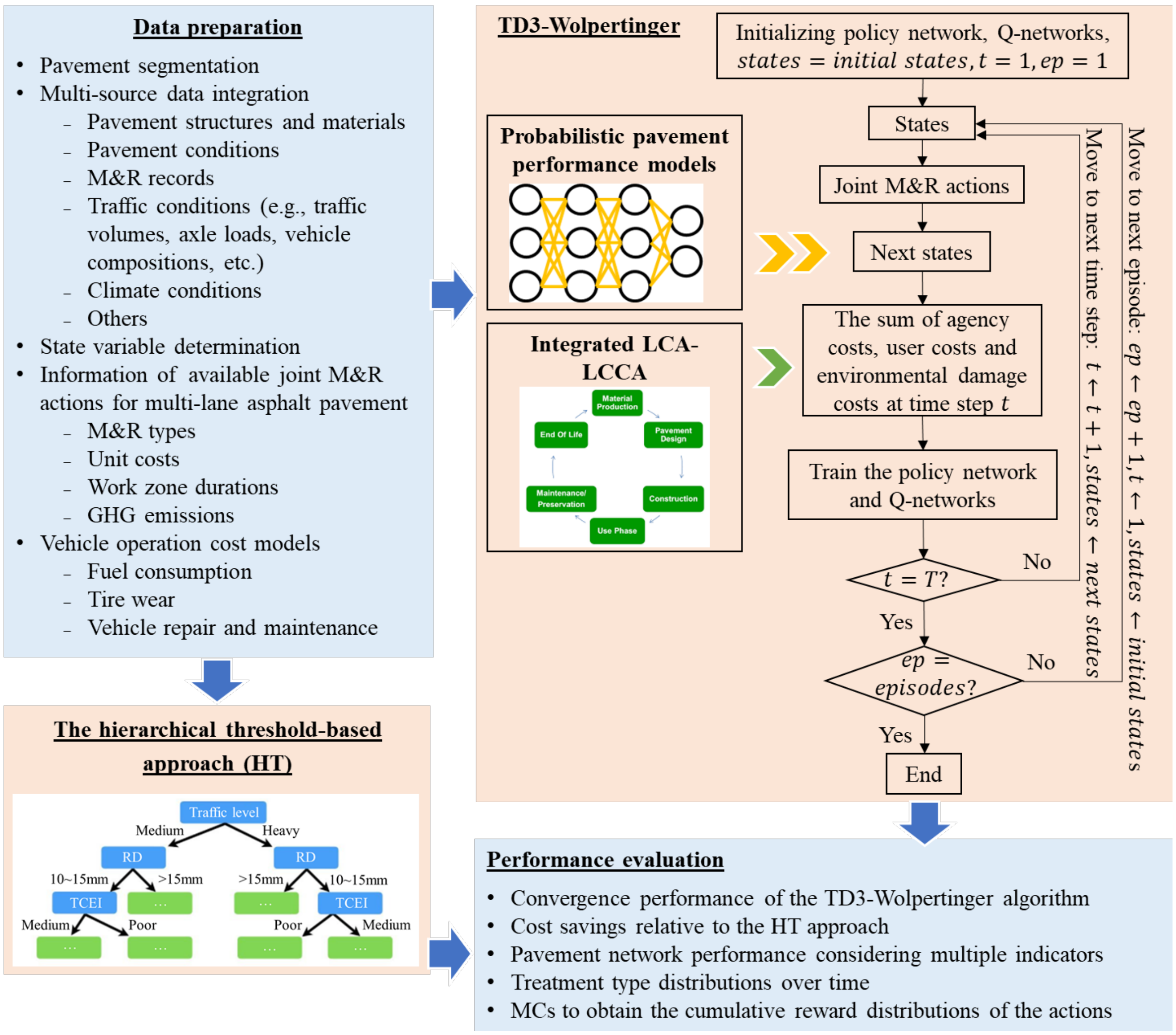


Fig. 1. An overview of the present study

A. Integrated LCA-LCCA Method

We employed the integrated LCA-LCCA method to estimate the cost and environmental impact induced in various pavement life cycle phases, with LCA accounting for the environmental aspects and LCCA dealing with the economic aspects of the pavements. The integration was achieved through the monetary method, i.e., monetarizing the GHG emissions. The functional unit is a one-way asphalt pavement segment of 1 km in length with 2 to 4 lanes (in one direction). The planning horizon was set to be 20 years, from 2021 to 2041. The carbon price was obtained from the 2020 China Carbon Pricing Survey [35], taking the expected average value at the midpoint (i.e., 2031) of the analysis period for calculation. The cost and environmental burden of applying the selected M&R action was assessed on a yearly basis. As this study aims to solve the pavement M&R optimization problem, the material production and construction stages for constructing the new pavement were not included. The pavement life cycle only encompasses three phases: M&R, usage, and WZ traffic management phases. We used the process-specific data for compiling the life cycle inventory (LCI) [36]-[39]. The costs of maintaining and rehabilitating the pavement were calculated using the local reference prices. The LCA consideration during the usage phase mainly results from the pavement-vehicle interaction. PVI describes the excess fuel consumption of vehicles driving on the uneven pavement. We employed the calibrated HDM-4 fuel consumption model [40] to determine the effect of pavement roughness on vehicle fuel consumption and tailpipe emissions. Regarding the economic dimension in usage phase, the models developed in the same study [40] were utilized to estimate the vehicle operation costs associated with fuel consumption, tire wear, and vehicle maintenance and repair. Moreover, work zone operation causes additional user costs due to the lane closure and reduced vehicle speed. In this study, the RealCost software [41] was applied to account for the impact of work zone on traffic congestion.

B. TD3 and Wolpertinger Algorithm

1) Reinforcement learning

Recent years have witnessed the significant success of reinforcement learning in solving various sequential decision-making problems [42]-[44]. RL learns the decision policy by enabling the agent to interact with the simulation environment and receive reward signals. At each time step t, with a given state s, the agent selects an action $a \in A$ according to its policy $\pi: S \rightarrow A$. By applying the action a to the simulation environment, a reward signal r is received and the next state of the environment s' is observed. These steps are repeated until a terminal state is reached (e.g., $t = T$). The return of an episode is the discounted sum of the rewards $R_t = \sum_{i=t}^{T} \gamma^{i-t} r(s_i, a_i)$, and the goal of RL is to find the optimal policy that maximize the expected return over all episodes. There are generally two types of RL algorithms: value-based and policy-based. Additional information about these two types of RL algorithms are introduced in Section 1.1 of the supplementary materials. Both types have been applied to addressing the pavement maintenance optimization problem [20]-[23].

2) TD3 Algorithm

In this study, we resorted to the TD3 algorithm to solve the pavement maintenance optimization problem. TD3 builds upon the actor-critic framework. Actor-critic methods combine the advantages of policy-based and value-based methods, aimed to learn both the value function (i.e., the critic) and the policy (i.e., the actor). In actor-critic methods, the actor decides which action should be taken and the critic informs the actor how good the action is and how it should adjust. TD3 algorithm was developed to tackle the overestimation bias with the value function by making a couple of modifications on Deep the Deterministic Policy Gradient algorithm (DDPG) [32]. More information about TD3 algorithm can be found in Section 1.2 of the supplementary materials.

3) Wolpertinger policy

Wolpertinger is a policy architecture which allows the agent to efficiently learn from large discrete action spaces. A schematic diagram is shown in Fig. 2 to prompt understanding. The method embeds the discrete actions in a continuous space based on priori information. The policy network outputs continuous actions within this space, but they may not be valid actions (i.e., the actions that are not within the discrete action space and thus cannot be performed in real world). Thus, an approximate nearest neighbor method is used to find a set of closest discrete actions and selects the action with the maximum Q-value as the final action. In this study, we combined the Wolpertinger policy with the TD3 algorithm instead of the DDPG algorithm that was used in the original study [33] due to the effort made by TD3 in mitigating the overestimation of value function [32].

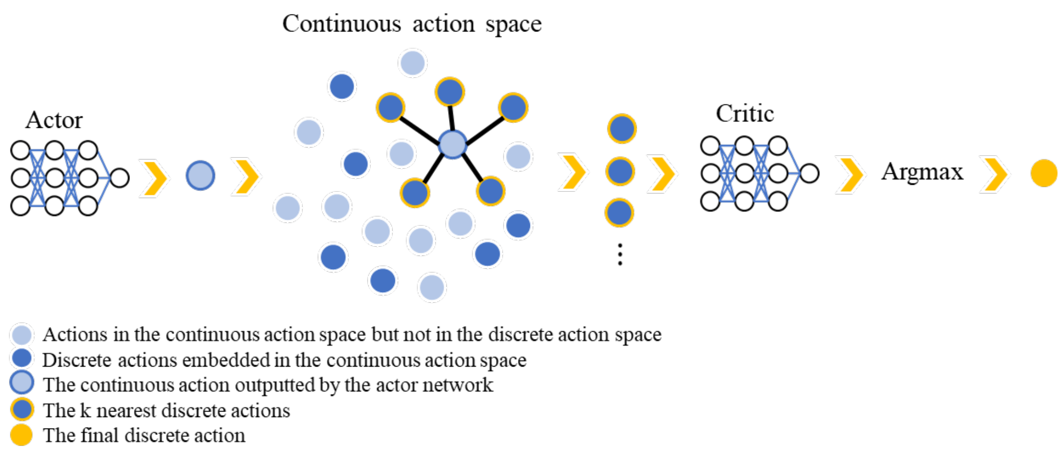


Fig. 2. A schematic diagram of the Wolpertinger policy

C. RL Components Definition in Pavement Maintenance Optimization

There are generally five key components in RL, i.e., agent, environment, state, action, and reward. The following are more detailed descriptions of their definitions in our problem.

1) Agent and environment

In the context of highway pavement maintenance optimization, the agent is the decision-maker who is in charge of making M&R plans. Decision-makers implement M&R activities on pavement after observing the pavement conditions and considering all other factors that may affect the M&R decision. Therefore, the environment encompasses the pavement segment itself, as well as the external environment in which it is located. In this study, we built our own RL simulation environment which contains a set of probabilistic pavement performance models and a reward system.

2) State

The state represents the minimum amount of information required by the agent to make M&R decisions in the simulation environment. In segment-level optimization problem, state variables need to cover all or at least the most important factors involved in the pavement performance models, while excluding variables that remain constant between road segments and over the planning horizon, or that can be fully deduced from other variables. Based on these principles, the state variables considered in the RL model were presented in Table S1 of the supplementary materials.

3) Action

Action space in pavement maintenance optimization problem typically refers to a set of available M&R treatments. In this study, the joint M&R treatment of multiple lanes in a certain segment was regarded as the action. Let x_{tnl} be the M&R treatment for lane l in segment n in year t, then the joint M&R treatment for segment n could be expressed as $x_{tn1}\& \ldots \& x_{tnL_n}$. To impose the vertical restriction, joint M&R treatments that violate the restriction were excluded when constructing the action space. Table II illustrates the various M&R treatments for a typical asphalt pavement in China with three asphalt layers, which are divided into preventive maintenance and rehabilitation treatments. Mill & fill and overlay differ in whether the existing pavement is milled to maintain the same pavement elevation before and after maintenance. The treatment thickness for fine mill & fill and thin overlay is generally within 2.5 cm for a typical road segment with upper, middle and lower asphalt layer thicknesses of 4 cm, 6 cm and 8 cm, respectively. PAC, ARAC and AC denote the porous asphalt concrete, asphalt-rubber concrete, and a dense-graded mixture, respectively, whereas the number "13" signifies the nominal maximum aggregate size in millimeters.

For a single lane, there are a total of 13 M&R options including the 12 treatments listed in Table II plus the "do-nothing" (ID=0). Hence, the number of joint M&R treatments for a one-way 4-lane segment could be up to $13^4 = 28561$ before the vertical restriction is considered. Even though imposing a vertical restriction would reduce the number of available actions to 4193, this is still a large action space compared to previous studies that did not provide lane-specific M&R decisions. Moreover, the discrete actions need to be embedded in a continuous action space. This was done by first ranking the actions according to their unit cost and then embedding them in a continuous space from -1 to 1. Fig. 3 is a schematic diagram for a one-way 4-lane segment. The different colored circles have the same meanings as those in Fig. 2.

TABLE II
THE AVAILABLE M&R TREATMENTS

ID	M&R treatment	Category
1	Seal coating	Preventive maintenance
2	Micro-surfacing	
3	Hot-in-place rehabilitation	
4	Fine mill & fill	
5	Thin overlay	
6	Fine mill & fill and thin overlay	
7	Mill & fill the upper asphalt layer	Rehabilitation
8	Overlay with PAC-13	
9	Overlay with ARAC-13	
10	Overlay with SBS modified AC-13	
11	Mill & fill the upper and middle asphalt layer	
12	Mill & fill the entire asphalt layer	

4) Reward

The reward in this study is defined as the negative of the sum of agency costs and additional user and environmental damage costs at each time step, as shown in Eq. 2(a)~(b). The additional user costs and EDC represent the additional costs and carbon emissions associated with the change in IRI from its baseline condition (IRI = 1 m/km) as well as the different traffic conditions between work zone operation and normal operation.

$$reward_{nt} = -cost_{nt} \tag{2a}$$

$$cost_{nt} = \sum_{l=1}^{L_n} (AC_{tnl} + \Delta UC_{tnl} + \Delta EDC_{tnl}), \forall n \in S, t \in T \tag{2b}$$

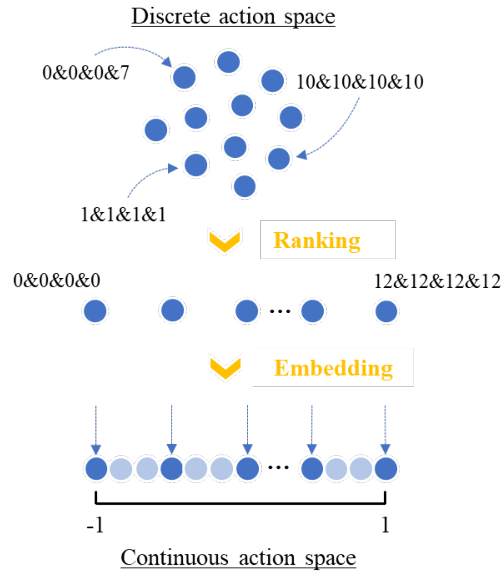


Fig. 3. Action embedding

D. Framework of TD3-Wolpertinger Algorithm

The pseudo-code of the TD3-Wolpertinger algorithm developed in this study is presented in Algorithm 1 of the supplementary materials. The detailed flowchart of the algorithm and the definitions of the RL components are illustrated in Fig. 4. A particle method [45][46] was used to propagate the uncertainty in pavement performance prediction. To do so, a set of particles with the same initial states were created at the beginning of each episode. Dynamic models are sampled from the previously developed probabilistic models [34]. These particles are then passed through different dynamic models to yield different outputs (i.e., next states). The expected reward was obtained by averaging over multiple particle trajectories. Moreover, the constraints in Eq.1 are imposed by removing the actions that violate the constrains from the original action space. A similar approach was employed and verified by Yuan *et al.* [47].

E. The Hierarchical Threshold-based Approach

The same problem was also solved through the hierarchical threshold-based approach which is widely used in practice. The HT approach makes decisions by constructing a decision-tree and going through a sequential list of questions. A specific M&R treatment was triggered once a sequence of thresholds was met. In this study, the HT model used in Jiangsu province, China [48] was adopted to serve as a benchmark method for our TD3-Wolpertinger model.

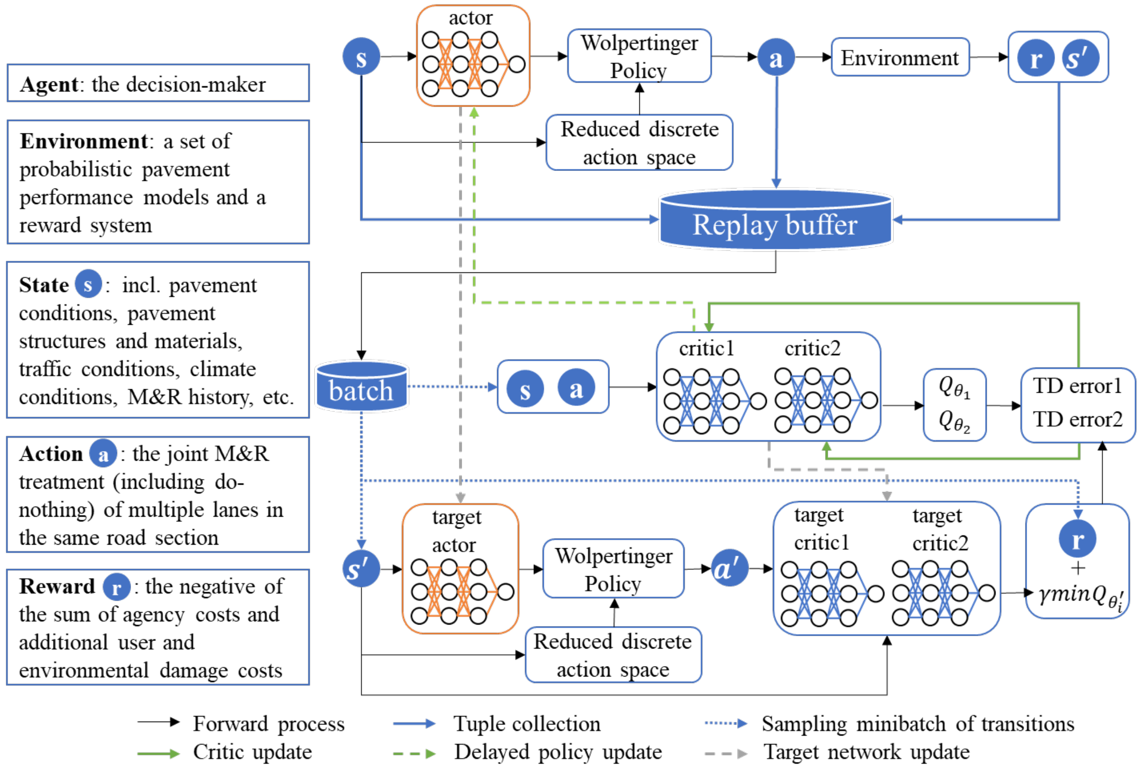


Fig. 4. Detailed flowchart of TD3-Wolpertinger algorithm

IV. Experiments and Results

To demonstrate the ability of our TD3-Wolpertinger algorithm in solving pavement M&R optimization problem, it was applied to the highway network in Jiangsu and compared with the HT approach. Road sections with complete data required for developing the RL model were extracted from the pavement management system (PMS). They were then segmented into shorter sections according to the directions, locations, structure sections, and traffic sections they belong to, and were further divided at 1-km intervals. Hence, a segment in this study generally refers to a 1-km one-way highway section containing 2 to 4 lanes (in one direction). After segmentation, a total of 7109 segments were obtained including 3967 2-lane segments, 2570 3-lane segments, and 572 4-lane segments. We developed three TD3-Wolpertinger models for the three groups of segments, denoted as RL-2, RL-3, and RL-4. The large number of pavement segments allows the model to cover a wide range of state space and enables it to be flexibly applied to large-scale road networks. Additionally, thresholds for mandatory do-nothing, maintenance or rehabilitation, and rehabilitation only, were set to the values when the evaluation index (e.g., RDI, RQI, SRI, and Transverse Cracks Evaluation Index (TCEI)) corresponding to each inspection index (e.g., RD, IRI, and SFC) reaches 95, 80 and 75, respectively.

F. Implementation Details

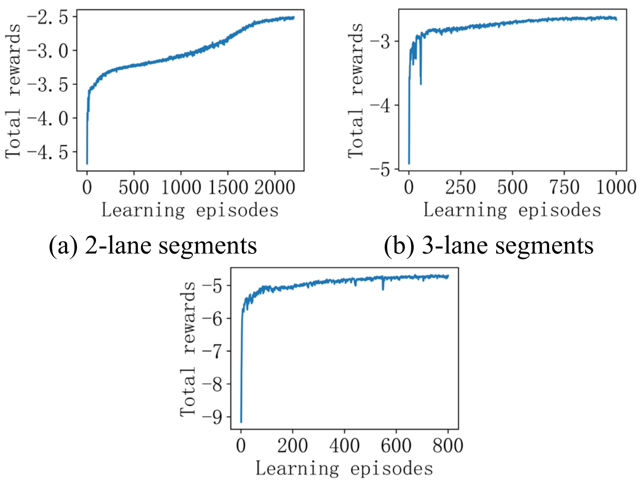
The TD3-Wolpertinger algorithm was coded in Python 3.8.3 using the PyTorch framework. The numerical experiment was performed on a desktop computer with Intel Core i7 2.90 GHz CPU and 64 GB RAM. Table III shows the adopted parameters in our TD3-Wolpertinger model, which were determined after a process of trial and error. Other parameters, such as the Gaussian exploration noise, target network update rate, clipped noise added to target policy, range to clip target policy noise, and frequency of delayed policy updates, followed the default settings in the original paper [32].

TABLE III
PARAMETERS IN TD3-WOLPERTINGER

Parameters		RL-2	RL-3	RL-4
Number of hidden layers		2	2	2
Number of hidden neurons	Actor	256, 128	256, 128	256, 128
	Critic	256, 256	256, 256	256, 256
Learning rate		0.0003	0.0003	0.0008
Learning episodes		1500	1000	800
Batch size		256	256	256
Memory size		30000	30000	20000
Discount factor		0.99	0.99	0.99
Number of nearest neighbors		10	200	500
Number of discrete actions		77	547	4193

G. Results and Discussion

Fig. 5 illustrates the learning curves of the three TD3-Wolpertinger models. The x-axis is the learning episodes. One episode represents a sequence of states, actions, and rewards, which ends with the terminal states. The y-axis denotes the total rewards (i.e., the negative of the total costs) over a planning horizon averaged by the number of segments and lanes, whose unit is million CNY/lane-km/20 years. It can be observed that the total rewards gradually increase over time and finally converge to a stable value. This implies that the M&R policy can no longer be improved, and so the optimal policies have been found.



(a) 2-lane segments (b) 3-lane segments

(c) 4-lane segments

Fig. 5. Learning curve of the three TD3-Wolpertinger models

To demonstrate the superiority of the innovative TD3-Wolpertinger model over the traditional HT approach, the M&R strategies for the entire pavement network were determined using both methods. Fig. 6 presents the distributions of the cost savings of the TD3-Wolpertinger model relative to the HT approach. The yellow diamond symbol and black horizontal line within each box signify the mean and median values of the cost savings. The top and bottom edges of the boxes are the 75th and 25th percentiles. The two whiskers specify the maximum and minimum values. Fig. 6 suggests that using the TD3-Wolpertinger model to replace the traditional HT approach could produce significant cost savings, and the average savings per lane-km in 20 years is about 1.72, 0.76, and 0.84 million CNY for 2-lane, 3-lane, and 4-lane segments, respectively. When considering the entire pavement network, such cost savings could reach 18147.27 million CNY, about 26.59% of the network cost of the HT approach.

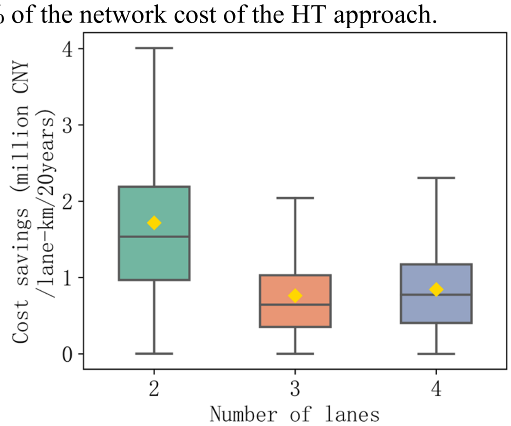


Fig. 6. Distribution of cost savings using TD3-Wolpertinger model compared to HT approach

Fig. 7 shows the breakdowns of the network cost savings for the three case studies. The reduced environmental impact was expressed in both monetary values (i.e., EDC) and original forms (i.e., GHG emission). In the three cases, the number of 2-lane segments was the largest, resulting in the greatest cost savings, totaling about 12058.47 million CNY. Among them, 61.03% came from the reduced agency costs, 37.88% was derived from the lower user costs due to the smoother pavement surface, and 1.09% was generated by the reduction of GHG emissions throughout the planning horizon. For the 3-lane segments, a total of 4548.50 million CNY cost savings were achieved, including 94.85% agency cost savings, 4.42% user cost savings and 0.73% EDC savings. The 4-lane segments have the minimal cost savings (1540.30 million CNY) because of its relatively small number of segments compared to the other two cases. The three cost items account for 40.16%, 58.63%, and 1.21% of the total cost savings, respectively. Additionally, despite the relatively small percentage of EDC savings, the savings in terms of GHG emissions of the three cases could reach 1321.14, 331.15, and 186.84 kilotonnes carbon dioxide equivalent (CO_2e), respectively.

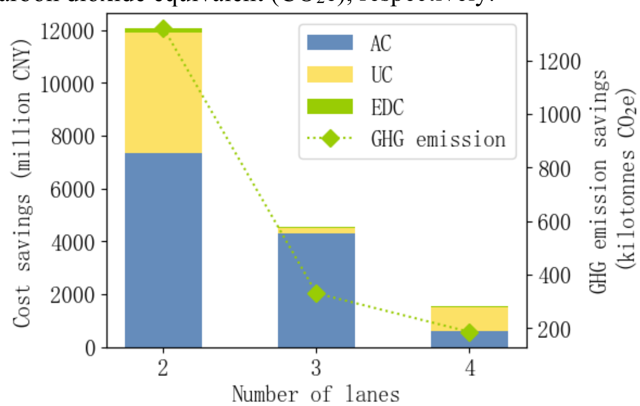


Fig. 7. Breakdown of the network cost savings

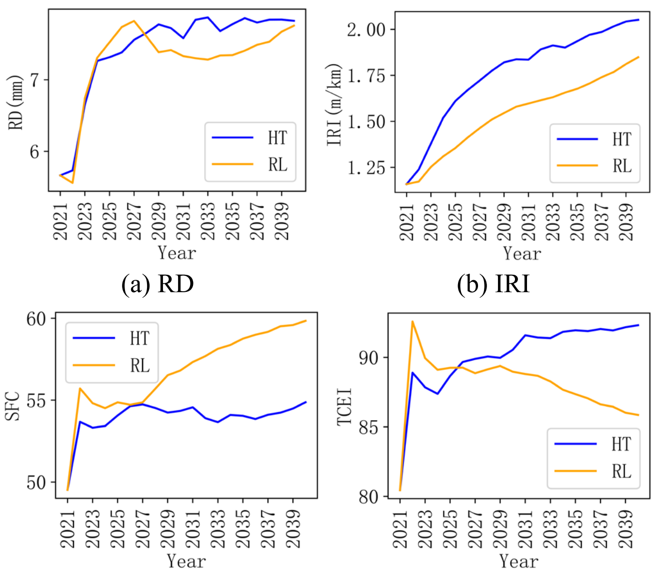
Based on the above analysis, it can be concluded that the developed TD3-Wolpertinger model dramatically improves the current M&R practice, and is beneficial to various stakeholders, including the highway agencies, road users and people who are affected by climate change due to the additional carbon emissions arising from the deteriorated pavement and the implementation of M&R treatments.

The pavement network performances across the planning horizon using the M&R strategies developed by both techniques were compared, as shown in Fig. 8. The network performance was evaluated using the traffic-length weighed indicator (TWI) [19], as shown in Eq.3.

$$TWI = \frac{\sum_{n=1}^{N_s} AADT_n \cdot length_n \cdot PPI_n}{\sum_{n=1}^{N_s} AADT_n \cdot length_n} \tag{3}$$

where $AADT_n$ is the annual average daily traffic of segment n, and $length_n$ denotes the length of segment n. Since this study seeks to build a multi-indicator decision process, four pavement performance indicators were embraced. IRI was directly incorporated into the reward function. A higher IRI would considerably increase the user cost, so the agent tends to reduce the IRI values to avoid reward losses. This was confirmed by Fig. 8(b) (also Fig. S1(b) and (f) of the supplementary materials) which show that the RL model produces better or at least similar roughness conditions compared to the HT model. As for the other pavement performance indicators (i.e., RD, IRI, and SFC), they were not directly targeted for optimization, but were subject to some constraints to avoid particularly poor conditions. Thus, the agent will not always pursue to optimize these indicators. Instead, it trades off between performing M&R to improve pavement conditions and deferring M&R to reduce M&R expenses. As demonstrated in Fig. 8(d) (and Fig.

S1(a)~(e) and (h) of the supplementary materials), the agent attempts to maintain the pavement condition within an acceptable range rather than keeping it in an optimal state at a high expense. There are also situations where the RL model yields better long-term performance as well as lower costs, such as those in Fig. 8(a)~(c) (and Fig. S1(f) and (g) of the supplementary materials).

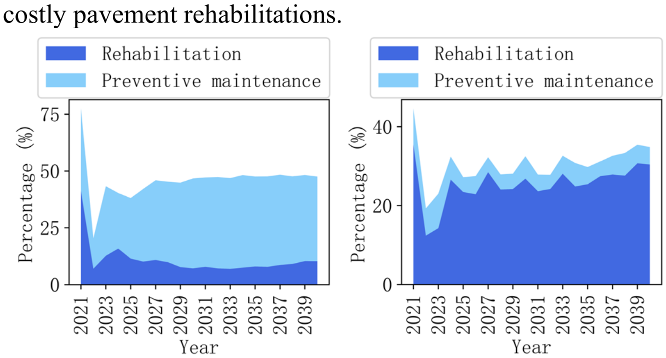


(a) RD (b) IRI

(c) SFC (d) TCEI

Fig. 8. Traffic-length weighed network performance for 2-lane segments

Furthermore, in order to shed some light on how the M&R strategies generated by the two approaches were different, the treatment type distributions over time for both models were plotted in Fig. 9 and Fig. S2 of the supplementary materials. The RL model has higher overall treatment ratios, but compared with the HT model, it tends to select less rehabilitations and more preventive maintenance throughout the analysis period. This reveals the benefits of preventive maintenance in reducing costly pavement rehabilitations.



(a) TD3-Wolpertinger (b) HT approach

Fig. 9. Ratios of treated segments by different treatment types for 2-lane segments

In addition, the cumulative reward of an M&R action is actually a random variable because of the uncertainties embedded in the pavement deterioration process. As pointed by Guo *et al.* [19], introducing uncertainty to the optimization forces the algorithm to make trade-off between the expected value of the solution and its uncertainty. To demonstrate this, we first obtained the optimal and suboptimal actions for each segment by comparing the Q-values of the candidate actions (the action with the largest Q-value was considered to be the optimal one). MCs was then adopted to randomly sample 1000 future deterioration trajectories for each segment starting from the selected current action and following the learned policy thereafter. Finally, the cumulative probability curves for the network costs over the entire analysis period were drawn in Fig. 10. From left to right in Fig. 10 are the network costs for the two-, three- and four-lane segments, respectively. The solid and dashed lines in each subplot represent the network cost distributions when each road segment takes its respective optimal and suboptimal actions, respectively.

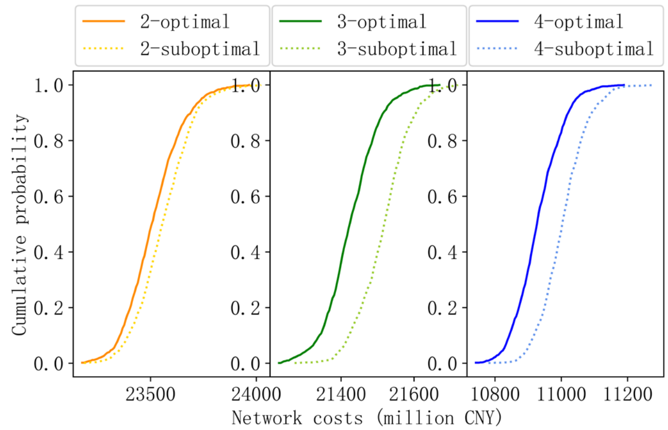
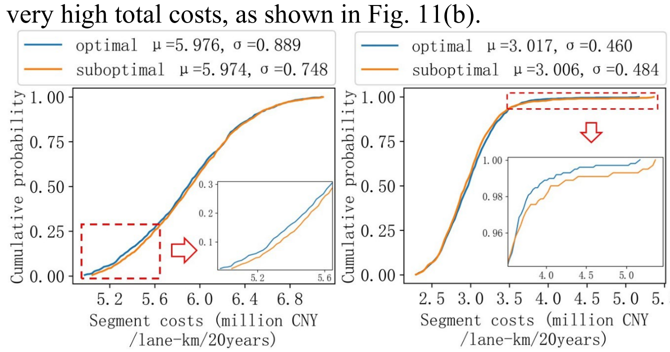


Fig. 10. Distribution of the network costs

We also investigated the segment-specific characteristics. In most cases, the optimal action of a certain segment has a lower mean cost than the suboptimal action. However, there are also exceptions where the trade-off was reflected. Fig. 11 shows the cost distributions of two representative segments when the optimal and suboptimal actions were selected respectively. For the first segment, the optimal action has larger mean value and standard deviation(std) than the suboptimal action. In this case, the larger deviation of the cost has positive effect as it provides better opportunity to get lower costs, as shown in Fig. 11(a). For the second segment, the mean cost of the optimal action is also larger but its std value is smaller. Thus, the larger deviation here exerts negative effect because it increases the risk of having a very high total costs, as shown in Fig. 11(b).



(a) Segment 1 (b) Segment 2

Fig. 11. Cost distributions of two representative segments if taking the optimal and suboptimal actions

Hence, it can be concluded that incorporating uncertainty into the optimization allows the agent to make trade-off between the expected return, the risk, and the opportunity of the solution. Meanwhile, the positive and negative deviations need to be differentiated. The uncertainty or variance not only measures the risk of the solution, but also implies the opportunity.

V. Conclusions

In this study, we proposed an innovative M&R optimization approach for multi-lane asphalt pavement in large-scale highway network based on a variant of RL algorithm, namely TD3. The Wolpertinger policy was combined with the TD3 algorithm to deal with the large discrete action space arising from the consideration of joint M&R actions for multi-lane road segments in this study. The integrated LCA-LCCA method was employed to assess the environmental and economic impact of the selected M&R actions. The uncertainty in pavement deterioration was introduced to the optimization process by integrating a set of probabilistic pavement performance models into the simulation environment of RL. The TD3-Wolpertinger method was demonstrated on a large-scale real-world highway pavement network and compared with the state-of-the-practice HT approach. Three case studies including all the one-way two-lane, three-lane, and four-lane segments in the road network were conducted.

The results show that the developed TD3-Wolpertinger models have good convergence performance. Compared with the traditional HT approach, the TD3-Wolpertinger models could offer significant cost savings of 18147.27 million CNY, approximately 26.59% of the network cost of the HT approach. Agency cost, user costs and environmental damage costs all made contributions to the total cost savings, which reveals that the proposed RL-based method is beneficial to various stakeholders. The TD3-Wolpertinger models always yield better or at least similar pavement roughness performance compared to the HT approach since roughness was directly incorporated into the reward function. For other PPIs, the first priority of the algorithm was not to keep them in the best conditions, but to ensure the long-term pavement performance within an acceptable range in a cost-effective manner. Moreover, the TD3-Wolpertinger models tend to select less rehabilitations and more preventive maintenance in contrast to the HT model, which leads to considerable cost savings.

In addition, incorporating the uncertainty of future pavement deterioration into the M&R optimization allows the model to trade off between the expected return and the uncertainty of the solution. It is worth noting that negative uncertainty measures the risk of the solution, while positive uncertainty actually describes the opportunity of obtaining better results. Thus, these two effects of uncertainty need to be distinguished in M&R decision-making. The proposed method also enhanced the flexibility of pavement management and made the policy model more adaptable to future variations due to the mapping structure of the policy model, the introduction of uncertainty in optimization, the large number of M&R alternatives and the large-scale pavement network.

Despite the contributions of this study, there is still much work to be done to further extend this research. One important issue is to solve the network-level resource allocation problem based on the outcomes of this study, thus enabling a complete bottom-up decision-making process. In addition, the current study considered the environmental impact of M&R strategies in decision-making by simply monetarizing the GHG emissions. Future research could improve this by applying multi-objective optimizations or incorporating other impact categories. Finally, other uncertain factors besides the uncertainty in pavement deterioration, such as the uncertain unit costs, emission factors, traffic volumes, etc., also need to be considered in future work.

Acknowledgements

This study was conducted under the support of the Research Institute for Sustainable Urban Development (RISUD) at the Hong Kong Polytechnic University. In addition, the data used in this research were collected from the Pavement Management System in Jiangsu province, China. The engineers and researchers who built the system and collected the data are also acknowledged for their contribution.

Reference

[1] J. Santos, A. Ferreira, and G. Flintsch, “A multi-objective optimization-based pavement management decision-support system for enhancing pavement sustainability,” Journal of cleaner production, vol. 164, pp. 1380–1393, 2017, doi: 10.1016/j.jclepro.2017.07.027.

[2] B. Yu, X. Gu, F. Ni, and R. Guo, “Multi-objective optimization for asphalt pavement maintenance plans at project level: Integrating performance, cost and environment,” Transportation research. Part D, Transport and environment, vol. 41, pp. 64–74, 2015, doi: 10.1016/j.trd.2015.09.016.

[3] L. Yao, Z. Leng, J. Lan, R. Chen, and J. Jiang, “Environmental and economic assessment of collective recycling waste plastic and reclaimed asphalt pavement into pavement construction: A case study in Hong Kong,” Journal of cleaner production, vol. 336, 2022, doi: 10.1016/j.jclepro.2022.130405.

[4] A. Medury and S. Madanat, “Simultaneous Network Optimization Approach for Pavement Management Systems,” Journal of infrastructure systems, vol. 20, no. 3, p. 4014010, 2014, doi: 10.1061/(ASCE)IS.1943-555X.0000149.

[5] K. Golabi, R. B. Kulkarni, and G. B. Way, “A Statewide Pavement Management System,” Interfaces (Providence), vol. 12, no. 6, pp. 5–21, 1982, doi: 10.1287/inte.12.6.5.

[6] H. Yeo, Y. Yoon, and S. Madanat, “Algorithms for bottom-up maintenance optimisation for heterogeneous infrastructure systems,” Structure and infrastructure engineering, vol. 9, no. 4, pp. 317–328, 2013, doi: 10.1080/15732479.2012.657649.

[7] Y. Ouyang and S. Madanat, “An analytical solution for the finite-horizon pavement resurfacing planning problem,” Transportation research. Part B: methodological, vol. 40, no. 9, pp. 767–778, 2006, doi: 10.1016/j.trb.2005.11.001.

[8] L. Gao and Z. Zhang, “Robust optimization for managing pavement maintenance and rehabilitation,” Transportation research record, vol. 2084, no. 2084, pp. 55–61, 2008, doi: 10.3141/2084-07.

[9] H. Zhang, G. A. Keoleian, M. D. Lepech, and A. Kendall, “Life-Cycle Optimization of Pavement Overlay Systems,”

Journal of infrastructure systems, vol. 16, no. 4, pp. 310–322, 2010, doi: 10.1061/(ASCE)IS.1943-555X.0000042.
[10] M. Irfan, M. B. Khurshid, Q. Bai, S. Labi, and T. L. Morin, "Establishing optimal project-level strategies for pavement maintenance and rehabilitation - A framework and case study," Engineering optimization, vol. 44, no. 5, pp. 565–589, 2012, doi: 10.1080/0305215X.2011.588226.
[11] B. Yu, Q. Lu, and J. Xu, "An improved pavement maintenance optimization methodology: Integrating LCA and LCCA," Transportation research. Part A, Policy and practice, vol. 55, pp. 1–11, 2013, doi: 10.1016/j.tra.2013.07.004.
[12] Y. Bai, O. E. Gungor, J. A. Hernandez-Urrea, Y. Ouyang, and I. L. Al-Qadi, "Optimal pavement design and rehabilitation planning using a mechanistic-empirical approach," EURO Journal on Transportation and Logistics, vol. 4, no. 1, pp. 57–73, 2015, doi: 10.1007/s13676-014-0072-2.
[13] J. Lee and S. Madanat, "Jointly optimal policies for pavement maintenance, resurfacing and reconstruction," EURO Journal on Transportation and Logistics, vol. 4, no. 1, pp. 75–95, 2015, doi: 10.1007/s13676-014-0065-1.
[14] J. Lee and S. Madanat, "A joint bottom-up solution methodology for system-level pavement rehabilitation and reconstruction," Transportation research. Part B: methodological, vol. 78, pp. 106–122, 2015, doi: 10.1016/j.trb.2015.05.001.
[15] Hamdi, S. P. Hadiwardoyo, A. G. Correia, and P. Pereira, "Pavement Maintenance Optimization Strategies for National Road Network in Indonesia Applying Genetic Algorithm," Procedia engineering, vol. 210, pp. 253–260, 2017, doi: 10.1016/j.proeng.2017.11.074.
[16] L. Zhang, L. Fu, W. Gu, Y. Ouyang, and Y. Hu, "A general iterative approach for the system-level joint optimization of pavement maintenance, rehabilitation, and reconstruction planning," Transportation research. Part B: methodological, vol. 105, pp. 378–400, 2017, doi: 10.1016/j.trb.2017.09.014.
[17] J. Santos, A. Ferreira, and G. Flintsch, "An adaptive hybrid genetic algorithm for pavement management," The international journal of pavement engineering, vol. 20, no. 3, pp. 266–286, 2019, doi: 10.1080/10298436.2017.1293260.
[18] M. R. Heidari, G. Heravi, and A. N. Esmaeeli, "Integrating life-cycle assessment and life-cycle cost analysis to select sustainable pavement: A probabilistic model using managerial flexibilities," Journal of cleaner production, vol. 254, p. 120046, 2020, doi: 10.1016/j.jclepro.2020.120046.
[19] F. Guo, J. Gregory, and R. Kirchain, "Incorporating cost uncertainty and path dependence into treatment selection for pavement networks," Transportation research. Part C, Emerging technologies, vol. 110, pp. 40–55, 2020, doi: 10.1016/j.trc.2019.11.015.
[20] L. Yao, Q. Dong, J. Jiang, and F. Ni, "Deep reinforcement learning for long-term pavement maintenance planning," Computer-aided civil and infrastructure engineering, vol. 35, no. 11, pp. 1230–1245, 2020, doi: 10.1111/mice.12558.
[21] S. Renard, B. Corbett, and O. Swei, "Minimizing the global warming impact of pavement infrastructure through reinforcement learning," Resources, conservation and recycling, vol. 167, p. 105240, 2021, doi: 10.1016/j.resconrec.2020.105240.
[22] P. Shani, S. Chau, and O. Swei, "All roads lead to sustainability: Opportunities to reduce the life-cycle cost and global warming impact of U.S. roadways," Resources, conservation and recycling, vol. 173, p. 105701, 2021, doi: 10.1016/j.resconrec.2021.105701.
[23] C. Han, T. Ma, and S. Chen, "Asphalt pavement maintenance plans intelligent decision model based on reinforcement learning algorithm," Construction & building materials, vol. 299, 2021, doi: 10.1016/j.conbuildmat.2021.124278.
[24] J. Lee and S. Madanat, "Joint optimization of pavement design, resurfacing and maintenance strategies with history-dependent deterioration models," Transportation research. Part B: methodological, vol. 68, pp. 141–153, 2014, doi: 10.1016/j.trb.2014.06.008.
[25] K. D. Kuhn, "Network-Level Infrastructure Management Using Approximate Dynamic Programming," Journal of infrastructure systems, vol. 16, no. 2, pp. 103–111, 2010, doi: 10.1061/(ASCE)IS.1943-555X.0000019.
[26] W. T. Chan, T. F. Fwa, and C. Y. Tan, "Road-Maintenance Planning Using Genetic Algorithms. I: Formulation," Journal of transportation engineering, vol. 120, no. 5, pp. 693–709, 1994, doi: 10.1061/(ASCE)0733-947X(1994)120:5(693).
[27] W. Chen and M. Zheng, "Multi-objective optimization for pavement maintenance and rehabilitation decision-making: A critical review and future directions," Automation in construction, vol. 130, 2021, doi: 10.1016/j.autcon.2021.103840.
[28] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction. MIT Press, 2018.
[29] A.-M. Karimi-Majd, M. Mahootchi, and A. Zakery, "A reinforcement learning methodology for a human resource planning problem considering knowledge-based promotion," Simulation modelling practice and theory, vol. 79, pp. 87–99, 2017, doi: 10.1016/j.simpat.2015.07.004.
[30] O. Swei, J. Gregory, and R. Kirchain, "Embedding Flexibility within Pavement Management: Technique to Improve Expected Performance of Roadway Systems," Journal of infrastructure systems, vol. 25, no. 3, p. 5019007, 2019, doi: 10.1061/(ASCE)IS.1943-555X.0000496.
[31] V. Donev and M. Hoffmann, "Optimisation of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types," The international journal of pavement engineering, vol. 21, no. 5, pp. 583–607, 2020, doi: 10.1080/10298436.2018.1502433.
[32] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," 2018.
[33] G. Dulac-Arnold *et al.*, "Deep Reinforcement Learning in Large Discrete Action Spaces," 2015.

[34] L. Yao, Z.Leng, J. Jiang, and F. Ni, " Modelling of Pavement Performance Evolution Considering Uncertainty and Interpretability: A Machine Learning Based Framework," International journal of pavement engineering, 2021, doi:10.1080/10298436.2021.2001814.

[35] H. Slater, D. De Boer, G. Qian, and W. Shu, "2020 China Carbon Pricing Survey," China Carbon Forum, Beijing, China, Dec. 2020.

[36] Eurobitume, "Life Cycle Inventory: Bitumen," 2011. [Online].Available:https://www.eurobitume.eu/fileadmin/pdf-downloads/LCI%20Report-Website-2ndEdition-20120726.pdf

[37] H. Stripple, "Life Cycle Assessment of Road: A Pilot Study for Inventory Analysis, " 2nd rev. ed. Swedish Environmental Research Institute (IVL) Report, 2001. [Online].Available:https://www.ivl.se/download/18.34244ba71728fcb3f3f57f/1591704221839/B1210E.pdf

[38] Ministry of Transport of the People's Republic of China, "Highway engineering budget quota (JTG/T 3832-2018)," 2018.[Online].Available:https://xxgk.mot.gov.cn/2020/jigou/glj/202103/t20210331_3547339.html

[39] Ministry of Transport of the People's Republic of China, "Highway Engineering Machinery Shifts Quota (JTG/T 3833-2018)," 2018.[Online].Available:https://xxgk.mot.gov.cn/2020/jigou/glj/202103/t20210331_3547339.html

[40] I. Zaabar and K. Chatti, "Estimating vehicle operating costs caused by pavement surface conditions," Transportation Research Record, vol. 2455, no. 2455, pp. 63–76, 2014, doi: 10.3141/2455-08.

[41] Federal Highway Administration (FHWA), "Life-Cycle Cost Analysis Software," Federal Highway Administration, Washington, DC, 2011.

[42] D. Silver *et al.*, "Mastering the game of Go without human knowledge," Nature (London), vol. 550, no. 7676, pp. 354–359, 2017, doi: 10.1038/nature24270.

[43] C. . Andriotis and K. . Papakonstantinou, "Managing engineering systems with large state and action spaces through deep reinforcement learning," Reliability engineering & system safety, vol. 191, p. 106483, 2019, doi: 10.1016/j.ress.2019.04.036.

[44] B. R. Kiran *et al.*, "Deep Reinforcement Learning for Autonomous Driving: A Survey," IEEE transactions on intelligent transportation systems, pp. 1–18, 2021, doi: 10.1109/TITS.2021.3054625.

[45] Y. Gal, R. McAllister, and C. E. Rasmussen, "Improving PILCO with Bayesian neural network dynamics models," In Data-Efficient Machine Learning workshop, ICML (Vol. 4, No. 34, p. 25). April, 2016.

[46] S. Depeweg, J. Hernández-Lobato, F. Doshi-Velez, and S. Udluft, "Learning and policy search in stochastic dynamical systems with Bayesian neural networks," 2017, doi: 10.17863/CAM.55902.

[47] Y. Yuan, L. Lei, T. X. Vu, S. Chatzinotas, S. Sun, and B. Ottersten, "Energy Minimization in UAV-Aided Networks: Actor-Critic Learning for Constrained Scheduling Optimization," IEEE transactions on vehicular technology, vol. 70, no. 5, pp. 5028–5042, 2021, doi: 10.1109/TVT.2021.3075860.

[48] Q. Wang, "Research of Long-Term Pavement Maintenance of Highway based on Life Cycle Analysis in Jiangsu Province, " Master's thesis, Southeast University, Nanjing, China, 2019. [Online]. Available: https://d.wanfangdata.com.cn/thesis/ChJUaGVzaXNOZXdTMjAyMTA1MTkSCUQwMTY1OTczNRoINGJxcnMzaGs%3D

Linyi Yao received the B.S. and M.S. degrees from Southeast University, Nanjing, China, in 2016 and 2019, respectively, and is currently a Ph.D. candidate in The Hong Kong Polytechnic University. Her research interests mainly include pavement management based on data analysis and machine learning, pavement sustainability assessment, data quality assessment and improvement, and risk and uncertainty analysis.

Zhen Leng received the PhD degree from University of Illinois, Urbana-Champaign, Urbana, IL, USA, in 2011. He is an Associate Professor of Civil and Environmental Engineering at The Hong Kong Polytechnic University. He serves as the Director of Road Research Laboratory. His research interests mainly include sustainable infrastructure materials and technologies, eco-efficiency analysis and life cycle assessment, and nondestructive evaluation of transportation infrastructure.

Jiwang Jiang received his PhD degree from Southeast University, Nanjing, Jiangsu, China, in 2019, and jointly trained in the University of Wisconsin-Madison, WI, USA from 2017 to 2018. He is currently a Postdoctoral Research Fellow in the Hong Kong Polytechnic University. His research interests mainly include sustainable pavement maintenance materials, multiscale characterization and modeling of asphalt pavement, and intelligent pavement management.

Fujian Ni received the B.S., M.S., and Ph.D. degrees from Southeast University, Nanjing, China. From 1997 to 1999, he was supported by the China Ministry to held a post-doctoral position with Nagaoka University of Technology, Japan. He is currently a Professor with Southeast University. He has served as the Primary Investigator for over 100 research projects, including the National Natural Science Foundation of China and the Natural Science Foundation of Jiangsu Province and has published more than 100 journal articles. His research interests include construction materials, pavement management systems, material research and mechanics analysis of steel deck pavement, and cold in-place recycling.