



# A multi-agent reinforcement learning model for maintenance optimization of interdependent highway pavement networks

L. Yao<sup>1</sup> | Z. Leng<sup>1</sup> | J. Jiang<sup>2</sup> | F. Ni<sup>2</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

<sup>2</sup>Department of Highway and Railway Engineering, School of Transportation, Southeast University, Nanjing, Jiangsu, China

## Correspondence

Z. Leng, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, 181 Chatham Road South, Hung Hom, Kowloon, Hong Kong.  
Email: [zhen.leng@polyu.edu.hk](mailto:zhen.leng@polyu.edu.hk)

## Abstract

Pavement segments are functionally interdependent under traffic equilibrium, leading to interdependent maintenance and rehabilitation (M&R) decisions for different segments, but it has not received significant attention in the pavement management community yet. This study developed a maintenance optimization model for interdependent pavement networks based on the simultaneous network optimization (SNO) framework and a multi-agent reinforcement learning algorithm. The established model was demonstrated on a highway pavement network in the real-world, compared to a previously built two-stage bottom-up (TSBU) model. The results showed that, compared to TSBU, SNO produced a 3.0% reduction in total costs and an average pavement performance improvement of up to 17.5%. It prefers concentrated M&R schedules and tends to take more frequent preventive maintenance to reduce costly rehabilitation. The results of this research are anticipated to provide practitioners with quantitative estimates of the possible impact of ignoring segment interdependencies in M&R planning.

## 1 | INTRODUCTION

Road networks play a vital role in providing mobility for people and goods and ensuring accessibility to a variety of locations and services. However, road pavements are continuously subjected to repeated traffic loads and various climatic conditions, resulting in an inevitable deterioration in pavement performance (Guo et al., 2021; Madanat, 1993). To restore pavement functionality, periodic maintenance and rehabilitation (M&R) are required (Z. Liu et al., 2022; Yao et al., 2019), which consume large amounts of natural and financial resources. The limited resource availability necessitates the effective planning and scheduling of M&R

actions (Adeli & Karim, 1997). Maintenance planning is one of the key components of a pavement management system (PMS), which can generally be divided into two levels: the network (or system) level and the project (or segment) level (Hudson et al., 1979). Network-level maintenance decision-making is often more complex due to the large scale of road network and the heterogeneous and interdependent nature of pavement segments that comprise the network (Durango-Cohen & Sarutipand, 2007; Medury & Madanat, 2014). Segment heterogeneity has been increasingly recognized in the pavement management community in recent years, with a growing number of studies and models being developed on this topic

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Computer-Aided Civil and Infrastructure Engineering* published by Wiley Periodicals LLC on behalf of Editor.



(Durango-Cohen & Sarutipand, 2007; Hong & Prozzi, 2010; Yeo et al., 2010, 2013; Z. Zhang et al., 2017).

In contrast, interdependencies among pavement segments, especially those related to road function, are rarely considered. In general, interdependencies in pavement networks can be either economic, stochastic, or functional (Durango-Cohen & Sarutipand, 2007, 2009). Economic dependence arises when components are linked by resource constraints or the cost of executing multiple actions together are different from the sum of individual execution costs (Durango-Cohen & Sarutipand, 2007, 2009). Budget constraints are the most common example of economic interdependencies in the field of pavement management and are one of the important considerations in network-level M&R decision-making. Stochastic dependence occurs in cases where the failure probabilities or timings of components are correlated due to some common causes (e.g., environment or traffic loading; Durango-Cohen & Sarutipand, 2007, 2009). Functional dependence refers to situations where the functionality of one component depends on the functionality of another (Durango-Cohen & Sarutipand, 2007, 2009; Medury & Madanat, 2013). In the context of pavement management, functional dependence arises from the connectivity of the road network and the desire of drivers to choose routes with lowest travel costs. Previous research has shown that lane closures during work zone operations and poor pavement conditions due to inadequate maintenance will cause increased traffic delays and vehicle operation costs (VOCs; Adeli & Ghosh-Dastidar, 2004; Adeli & Jiang, 2008; Santos et al., 2017). As a result, travel costs on routes passing through these segments may increase, and drivers may reroute to avoid high travel cost segments, leading to a redistribution of traffic flow across the road network (Guan et al., 2022; Uchida & Kagaya, 2006). Therefore, the condition of one segment may affect the traffic level on another segment, which will further affect the pavement performance and corresponding M&R strategies of this segment (Durango-Cohen & Sarutipand, 2009). Conversely, the M&R decision of a segment is also related to the condition of other segments in the network. This indicates the functional dependence of pavement segments. M&R strategies derived without considering functional dependence may hinder effective decision support, as significant benefits and costs in the management process can be attributed to the interdependencies that connect the segments of a system (Durango-Cohen & Sarutipand, 2009). However, existing pavement management studies rarely consider such functional dependence among road segments and lacks quantitative justification for the rationality of the assumption of segment independence. Many studies adopt a two-stage bottom-up (TSBU) framework, which first identifies alternative M&R schedules for each segment,

and then performs budget allocation at the network-level to solve the maintenance optimization problem for real-world-scale pavement networks (Guo et al., 2020; Lee & Madanat, 2015a; Swei et al., 2019). Although the TSBU approach is generally computationally tractable, it cannot take into account the functional dependence among road segments arising from dynamic traffic distribution. While the simultaneous network optimization (SNO; Medury & Madanat, 2014) framework may address this challenge, most of the existing literature only considers the economic dependence caused by budget constraints (Cao et al., 2020; Medury & Madanat, 2014; Wang et al., 2003).

To the best of the authors' knowledge, only 12 studies so far have considered the functional dependence of road segments due to the impact of road conditions on traffic distribution across the entire road network when developing long-term pavement maintenance optimization models, as summarized in Table 1. These studies differ in network topology, influence mechanism of road conditions on traffic distribution, and traffic assignment methods, in addition to those common differences among pavement maintenance optimization models such as optimization objectives and algorithms. Seven of the 12 studies applied models to simple hypothetical networks and three were based on a popular small-to-medium-scale real-world network, that is, the Sioux Falls network (Gao et al., 2011; C. Liu et al., 2020; Yin et al., 2008). One study utilized a network consisting of long-distance road segments (23–72 miles), that is, the Illinois network, to simplify the network topology and thereby reduce the complexity of the optimization problem (Hajibabai et al., 2014). Chu and Chen (2012) tested their model on a relatively large road network (consisting of hundreds of road segments), but the possible solution space considered in the model is limited due to the use of threshold-based decision variables. In addition, previous studies have identified three ways in which road conditions influence traffic distribution as indicated in Table 1. It was found that travel costs along various routes and drivers' route choice can be affected either because pavement M&R reduces road capacity (Durango-Cohen & Sarutipand, 2007, 2009; C. Liu et al., 2020; Mao et al., 2019; Ng et al., 2009; Uchida & Kagaya, 2006) or because poor pavement surface conditions increase VOCs (Durango-Cohen & Sarutipand, 2009; Guan et al., 2022; Hajibabai et al., 2014; C. Liu et al., 2020; Mao et al., 2019; Ouyang, 2007; Uchida & Kagaya, 2006; Yin et al., 2008) or travel time (Chu & Chen, 2012; Gao, Xie et al., 2011; Guan et al., 2022). To account for these effects on network traffic distribution, different traffic assignment methods have been employed, with the user equilibrium (UE) method being the most widely used (Chu & Chen, 2012; Gao et al., 2011; Guan et al., 2022; Hajibabai et al., 2014; C. Liu et al., 2020; Yin et al., 2008). UE is a user-driven



TABLE 1 Existing studies on long-term pavement maintenance optimization that consider the functional dependence of road segments due to the impact of road conditions on the network traffic distribution.

Reference	Optimization objectives	Planning horizon	No. of segments	Network topology	Influence mechanism of road conditions on traffic distribution			Traffic assignment method	Optimization algorithm
					A1	A2	A3		
Uchida and Kagaya (2006)	Min. LCC	40	10	Hypothetical grid network with six nodes and 10 links	✓	✓		Probit-based SUE	Sensitivity analysis technique
Durango-Cohen and Sarutipand (2007)	Min. discounted cost over the planning horizon	Infinite	5	Hypothetical diamond network with four nodes and five links	✓			N/A <sup>a</sup>	Quadratic programming
Ouyang (2007)	Min. discounted LCC	Infinite	2	Hypothetical road network with two nodes and two links		✓		DUE	M-PPI
Yin et al. (2008)	Min. total cost needed to achieve the required conditions and performance	20	76	Sioux Falls network with 24 nodes and 76 links		✓		UE	Cutting plane algorithm
Durango-Cohen and Sarutipand (2009)	Min. total discounted social cost, that is, user plus agency costs (ACs)	25	3	Hypothetical road network with two nodes and three links	✓	✓		N/A <sup>a</sup>	Quadratic programming
Ng et al. (2009)	Min. maintenance cost and user delay	3	24	Hypothetical grid network with 13 nodes and 24 links	✓			CTM	Genetic algorithm
Gao et al. (2011)	Min. system-level travel time under the annual budget constraint	5	76	Sioux Falls network with 24 nodes and 76 links			✓	UE	Generalized benders decomposition
Chu and Chen (2012)	Min. system-wide traffic weighted IRI for the planning horizon	10	270	Highway network of urban Dasi Township in Taoyuan County, Taiwan, with 70 nodes and 270 links			✓	UE	Modified tabu search algorithm
Hajibabai et al. (2014)	Min. total system cost due to facility investment, transportation cost, and pavement LCC	Infinite	38	Highway network in Illinois with 15 nodes and 38 links		✓		UE	Bi-level MINLP

(Continues)



TABLE 1 (Continued)

Reference	Optimization objectives	Planning horizon	No. of segments	Network topology	Influence mechanism of road conditions on traffic distribution			Traffic assignment method	Optimization algorithm
					A1	A2	A3		
Mao et al. (2019)	Min. LCC	30	11	Hypothetical road network with nine nodes and 11 links	✓	✓		SUE	Heuristic iterative algorithm
C. Liu et al. (2020)	Min. fuel consumption in a network constrained by an annual maintenance budget	10	22	Simplified Sioux Falls network with 15 nodes and 22 links	✓	✓		UE	Modified active set algorithm
Guan et al. (2022)	Min. ACs, user costs and environmental impacts	10	24	Hypothetical grid network with nine nodes and 24 links		✓	✓	UE	Modified NSGA II

Note: A1, pavement maintenance and rehabilitation (M&R) work results in reduced roadway capacity, thereby altering vehicle travel time and drivers' route choice; A2, poor pavement surface conditions increase vehicle operation costs (VOCs), which changes vehicle travel cost and drivers' route choice; A3, poor pavement surface conditions decrease vehicle free flow speed or road capacity, thereby altering vehicle travel time and drivers' route choice;

Abbreviations: CTM, cell transmission model; DUE, deterministic user equilibrium; IRI, international roughness index; LCC, life cycle costs; M-PPI, modified parametric policy iteration; MINLP, mixed integer non-linear program; NSGA II, non-dominated sorting genetic algorithm II; SUE, stochastic user equilibrium.

<sup>a</sup>These studies did not address specific traffic assignment problem but rather used pre-determined parameters to estimate the traffic disruption and rerouting costs that are dependent on the conditions or M&R decisions of other segments.



traffic assignment method in which all drivers are selfishly choosing routes that minimize their delay (Wardrop, 1952). It falls under the category of static traffic assignment, which contrasts to dynamic traffic assignment (DTA). In Table 1, only one study has simultaneously accounted for traffic dynamics and long-term pavement maintenance planning by combining DTA and genetic algorithm (Ng et al., 2009). It utilized mesoscopic traffic simulation to estimate the impact of M&R work of the current segment on the travel time of the upstream segment. While this study is an important advance, the model it proposed is too computationally expensive. Even with a 3-year planning period, a hypothetical grid network of 24 road segments, and a 3-h traffic simulation period, it took about 2 days to get the solution (Ng et al., 2009). Meanwhile, it is usually considered unnecessary to capture detailed time-varying traffic flows across the network in long-term pavement maintenance optimization (Chu & Chen, 2012; Guan et al., 2022). In summary, existing network-level long-term pavement maintenance optimization models that take into account the dynamic distribution of traffic flows in a road network due to different conditions of road segments in the network have been demonstrated only for topologically simple road networks or optimization problems with limited solution spaces, leaving their applicability to more complex tasks unknown. In addition, they did not quantitatively assess the potential impact of ignoring segment functional dependence in maintenance planning.

To deal with the aforementioned research gaps, this study aims to develop a long-term pavement maintenance optimization model, which is applicable to real-world-scale highway networks and can take into account the functional dependence among pavement segments under traffic equilibrium and budget constraints. The effects of traffic loads on pavement degradation were modeled by previously built pavement performance models (Yao et al., 2022a). The deteriorated pavement conditions, in turn, influence drivers' route choices by including VOCs as a component of the travel cost function. The optimization objectives considered in this study include maintenance investment, that is, agency cost (AC); VOC and travel time, that is, user cost (UC); and environmental impact. In order to trade off these conflicting objectives, common approaches include using multi-objective optimization (MOO) algorithms to find Pareto-optimal solutions (Cao et al., 2020; Guan et al., 2022; Santos et al., 2017; Yu et al., 2015), transforming some of the objectives into constraints (de la Garza et al., 2011; Lee & Madanat, 2017), and converting MOO problems into single-objective optimization (SOO) problems (Gao et al., 2012; Lee & Madanat, 2014; H. Zhang et al., 2010). This study falls under the third method, which transforms the MOO problem into

a SOO problem by monetizing intangibles, but the proposed modeling framework can also be easily extended to assign unequal weights to different costs. Furthermore, sequential decision making in long-term pavement maintenance optimization makes reinforcement learning (RL) a perfect fit for this problem (Sutton & Barto, 2018). The maintenance optimization problem of a road network with interdependent segments is essentially a multi-agent system (MAS) task, that is, a set of agents interacting with each other and with the environment to achieve system-wide goals. In light of these characteristics, a multi-agent RL(MARL) algorithm called QMIX (Rashid et al., 2018) was adopted to develop an optimization model that is capable of coordinating M&R actions on different segments to achieve long-term goals of the entire road network. Based on these considerations, three main contributions are expected to be made by this research, including: (1) the functional dependence of road segments arising from dynamic traffic distribution as a result of pavement deterioration is incorporated into long-term pavement maintenance optimization; (2) MARL is applied to the pavement management community for the first time, confirming its ability to capture segment interdependencies in multi-year M&R planning; and (3) the potential impact of ignoring segment functional dependence in maintenance planning is estimated to provide justification for the rationality of the assumption of segment independence.

## 2 | MODEL FORMULATION

### 2.1 | Problem description

This research aims to solve the network-level long-term pavement M&R optimization problem to minimize the total cost for different stakeholders while taking into account the functional dependence of road segments under traffic equilibrium. There are interactions among pavement conditions, network traffic flows, and M&R decisions. While it is known to all that traffic flow affects the deterioration of pavement performance, little attention has been paid to the impact of pavement conditions on the distribution of network traffic flows. The M&R decisions also exert an influence on network traffic distribution due to the reduction in roadway capacity resulting from lane closures during M&R. Nevertheless, as the decision interval in long-term pavement maintenance optimization (typically 1 year) is often much longer than the duration of M&R activities (several hours or days), this effect is usually regarded as insignificant (Guan et al., 2022; Ng et al., 2009) and is therefore not considered in this study. In other words, it is assumed that pavement maintenance





operations will not cause vehicles to reroute but will only cause vehicles to slow down or queue up. Pavement M&R will be carried out lane by lane in order to not excessively disrupt traffic. The additional travel time cost (TTC) caused by traffic delays in the maintenance work zones when vehicles pass through the maintenance section is considered part of the total cost.

The problem in this study is essentially a bi-level problem, where the upper-level is a pavement maintenance optimization problem that identifies the optimal maintenance decisions for each segment, and the lower-level is a traffic assignment problem that assigns traffic flows to different road sections. It is supposed that there is an access-controlled highway network  $G(V, E)$ , where  $V$  and  $E$  are the set of intersections (toll stations and hubs) and traffic sections (the one-way road section between two toll stations/hubs), respectively. Due to the different pavement structures and the general length of M&R sections, traffic sections in  $E$  can be further divided into multiple subsections with two or more lanes. Since traffic assignment and pavement maintenance planning are performed for each road section and single-lane pavement segment respectively, the remainder of this paper will use the terms “section” and “segment” to distinguish road sections that contain multiple lanes and a single lane. The assigned traffic was distributed to each lane through the pre-determined lane distribution factors.

## 2.2 | Notation list

To enhance the readability of the model formulation, the notations and meanings of indices, parameters, sets, and variables utilized in the model are first defined in Table 2.

## 2.3 | Mathematical formulation

The overall optimization goal of the problem is to minimize the total discounted cost, that is, the sum of discounted ACs and extra UCs (equal to the sum of VOCs and TTCs) and environmental damage costs (EDCs) for the entire pavement network over a given planning horizon  $T$  under traffic equilibrium and annual budget constraints. The upper-level problem can be formulated as follows:

Minimize:

$$\sum_{t=1}^T \gamma^{t-1} \left\{ AC_t + \sum_{n=1}^N [\Delta VOC_{t,n} + \Delta EDC_{t,n}] + \sum_{m=1}^M \Delta TTC_{t,m} \right\} \quad (1a)$$

subject to:

$$AC_t = \sum_{n=1}^N c(x_{t,n}) l_n w + p \cdot \max \left[ \left( \sum_{n=1}^N c(x_{t,n}) l_n w - B_t \right), 0 \right] \quad \forall t = 1, \dots, T \quad (1b)$$

$$VOC_{t,n} = VOC_{t,n}(v_{t,n}, IRI_{t,n}) - VOC_{0,n}(v_{0,n}, IRI^{bl}) \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1c)$$

$$VOC_{t,n}(v_{t,n}, IRI_{t,n}) = FC_{t,n} + TW_{t,n} + VMR_{t,n} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1d)$$

$$FC_{t,n} = l_n v_{t,n} \sum_{ve \in VE} (\beta_0^{ve} + \beta_1^{ve} IRI_{t,n}) f_c(f_{ve}) p_{n,ve} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1e)$$

$$TW_{t,n} = l_n v_{t,n} \sum_{ve \in VE} (\beta_2^{ve} IRI_{t,n}^2 + \beta_3^{ve} IRI_{t,n} + \beta_4^{ve}) t_i \cdot p_{n,ve} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1f)$$

$$VMR_{t,n} = l_n v_{t,n} \sum_{ve \in VE} vmr_{ve,IRI_{t,n}} p_{n,ve} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1g)$$

$$\Delta TTC_{t,m} = TTC_{t,m}(x_{t,m}, v_{t,m}) - TTC_{0,m}(\text{“do-nothing”}, v_{0,m}) \quad \forall t = 1, \dots, T, m = 1, \dots, M \quad (1h)$$

$$TTC_{t,m}(x_{t,m}, v_{t,m}) = \varphi_t^f \left[ 1 + \alpha \left( \frac{v_{t,m}}{C_{t,m}} \right)^\beta \right] d_{t,m}^{dn} + \varphi_t^f \left[ 1 + \alpha \left( \frac{v_{t,m}}{C_{t,m}(nl_m - 1)/nl_m} \right)^\beta \right] d_{t,m}^{mr} \quad \forall t = 1, \dots, T, m = 1, \dots, M \quad (1i)$$

$$\Delta EDC_{t,n} = EDC_{t,n}^a(x_{t,n}) + [EDC_{t,n}^u(v_{t,n}, IRI_{t,n}) - EDC_{0,n}^u(v_{0,n}, IRI^{bl})] \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1j)$$

$$EDC_{t,n}^a(x_{t,n}) = EF(x_{t,n}) l_n w \cdot cp \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1k)$$



TABLE 2 Notation list.

Type	Variables	Meanings
Index	$t$	Year of decision
	$n, m$	Segment and section ID
	$ve$	Vehicle type
	$i$	Index of pavement performance indicators (PPI)
	$r, s$	Origin and destination node in the road network
	$k$	Path between origin–destination (OD) pairs
Parameter	$T$	Planning horizon
	$N, M$	Total number of pavement segments and sections
	$IRI^{bl}$	The baseline value of international roughness index (IRI), which is equal to 1 m/km in this study
	$\epsilon_i$	The maximum difference allowed between the $i$ th PPI and its initial state when do-nothing is forcibly selected
	$B_t$	Annual maintenance budget in year $t$
	$OD\ trip\ matrix_t$	OD trip matrix describing people movement in a certain area in year $t$
	$q^{r,s}(t)$	Travel demand for OD pair $r - s$ in year $t$
	$Th_i^{M\&R}$	Threshold of the $i$ th PPI when forcing the selection of an action other than do-nothing
	$Th_i^R$	Threshold of the $i$ th PPI when forcing the selection of a rehabilitation action
	$Th_i^{new}$	The initial state of the $i$ th PPI
	$\varphi$	Time value coefficient
	$\alpha, \beta$	Model constants, generally $\alpha = 0.15, \beta = 4$ for bureau of public roads function
	$t_m^f, C_{t,m}$	Free-flow travel time on section $m$ , traffic capacity of section $m$ in year $t$
	$nl_m$	Number of lanes of section $m$ in year $t$
	$d_{t,m}^{mr}$	Cumulative time for M&R in year $t$ on a lane-by-lane basis on section $m$
	$d_{t,m}^{dn}$	Cumulative time in year $t$ without M&R on all lanes on section $m$
	$\delta_{m,k}^{r,s}$	A binary variable equal to one if path $k \in K_{r,s}$ between OD pair $r - s$ uses section $m$ and zero otherwise
	$\sigma_{m,n}$	A binary variable equal to one if segment $n$ belongs to section $m$ and zero otherwise
	$p$	Penalty factor for budget overruns
	$LDF_n, l_n, w$	Lane distribution factor of segment $n$ , length of segment $n$ , lane width
	$f_{ve}, fc(f_{ve}), EF(f_{ve})$	Fuel type for vehicle type $ve$ , unit cost of fuel $f_{ve}$ , emission factor of fuel $f_{ve}$
	$ti, p_{n,ve}$	Tire cost, percentage of vehicle type $ve$ in the vehicle fleet of segment $n$
	$\beta_0^{ve}, \dots, \beta_4^{ve}$	Coefficient to calculate fuel and tire consumption for vehicle type $ve$
	$vmr_{ve,IRI_{t+n}}$	Unit repair and maintenance costs per vehicle kilometer traveled corresponding to vehicle type $ve$ and $IRI_{t+n}$
	$cp$	Carbon price
Set	$VE$	Set of vehicle types
	$I$	Set of PPI
	$A$	Set of M&R and do-nothing
	$A^{M\&R}$	Set of M&R
	$A^R$	Set of rehabilitations
	$R, S$	Set of origin and destination nodes in the road network
	$K_{r,s}$	Set of paths between OD pair $r - s$

(Continues)



TABLE 2 (Continued)

Type	Variables	Meanings
Variable	$AC_t$	AC of the entire pavement network in year $t$
	$c(x_{t,n}), EF(x_{t,n})$	Unit cost and greenhouse gas emission per square meter of M&R $x_{t,n}$
	$VOC_{t,n}$	VOC of segment $n$ in year $t$
	$EDC_{t,n}$	Environmental damage cost (EDC) of segment $n$ in year $t$
	$TTC_{t,m}$	Travel time cost (TTC) of section $m$ in year $t$
	$FC_{t,n}, TW_{t,n}, VMR_{t,n}$	Fuel consumption, tire wear, and vehicle repair and maintenance costs of segment $n$ in year $t$
	$EDC_{t,n}^a, EDC_{t,n}^u$	Agency- and user-induced EDC of segment $n$ in year $t$
	$VOC_{t,m}^{pv}$	VOC per vehicle on section $m$ in year $t$
	$PPI_{i,t,n}$	The $i$ th PPI of segment $n$ in year $t$
	$IRI_{t,m}, IRI_{t,n}, IRI_t$	IRI of section $m$ , segment $n$ , and the entire pavement network in year $t$
	$IRI_{t+,n}$	Mean of immediate post-treatment IRI and pre-treatment IRI in the following year for segment $n$ in year $t$
	$x_{t,n}$	Decision variable, that is, the selected M&R action, of segment $n$ in year $t$ , $x_{t,n} \in A$
	$v_{t,m}, v_{t,n}$	Traffic flow of section $m$ and segment $n$ in year $t$
	$w$	Integral variable
	$f_k^{r,s}(t)$	Traffic flow on path $k \in K_{r,s}$ between OD pair $r-s$ in year $t$

$$EDC_{t,n}^u(v_{t,n}, IRI_{t+,n}) \leq \epsilon_i, \text{ or } x_{t-1,n} \neq \text{"do-nothing"} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1q)$$

$$= l_n v_{t,n} \sum_{ve \in VE} (\beta_0^{ve} + \beta_1^{ve} IRI_{t+,n}) EF(f_{ve}) cp \cdot p_{n,ve} \quad \sum_{n=1}^N AC_{t,n}(x_{t,n}) \leq B_t \quad \forall t = 1, \dots, T \quad (1r)$$

$$\forall t = 1, \dots, T, n = 1, \dots, N \quad (1l)$$

$$v_{t,n} = f^{UE}(IRI_t, OD \text{ trip matrix}_t) \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1m)$$

$$PPI_{i,t+1,n} = f_i^{BNN}(PPI_{i,t,n}, v_{t,n}, \text{structure, climate, M\&R history, others}) \quad \forall t = 1, \dots, T, n = 1, \dots, N, i \in I \quad (1n)$$

The objective function is given by Equation (1a) where the AC and extra VOC, TTC, and EDC are discounted and summed to measure the pavement-induced impacts on road agencies, users, and the natural environment. Equation (1b) indicates that the AC of the entire pavement network consists of the M&R cost and the penalty term resulting from budget overruns. Equations (1c), (1h), and (1j) specify the expressions for extra VOC, TTC, and EDC, that is, the difference between the actual costs and the baseline costs calculated from the initial year traffic flow and baseline international roughness index (IRI).

$$x_{t,n} \in A^R, \text{ IF } \exists i \in I, \begin{cases} PPI_{i,t,n} > Th_i^R & \text{for RD and IRI} \\ PPI_{i,t,n} < Th_i^R & \text{for SFC and TCEI} \end{cases} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1o)$$

$$x_{t,n} \in A^{M\&R}, \text{ IF } \exists i \in I,$$

$$\begin{cases} PPI_{i,t,n} > Th_i^{M\&R} & \text{for RD and IRI} \\ PPI_{i,t,n} < Th_i^{M\&R} & \text{for SFC and TCEI} \end{cases}, \text{ and } \nexists i \in I, \text{ satisfying Equation (1o)} \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (1p)$$

$$x_{t,n} = \text{"do-nothing," IF } \forall i \in I, |PPI_{i,t,n} - Th_i^{new}|$$

Equation (1d) shows that VOC is composed of fuel cost, tire wear cost, and vehicle repair and maintenance cost, which are calculated as shown in Equations (1e)–(1g),





respectively (Chatti & Zaabar, 2012). Equation (1i) calculates the total TTC for section  $m$  in year  $t$  for both maintenance and normal periods based on the bureau of public roads (BPRs) function (Bureau of Public Roads, 1964). It is assumed that pavement maintenance will be done lane by lane to avoid excessive disruption to traffic, so during the maintenance period, the road capacity will be reduced to  $C_{t,m}(nl_m - 1)/nl_m$ . Equation (1k) denotes the calculation of agency-induced EDC, where the carbon price was set to 100 CNY/tonne (Slater et al., 2020). Equation (1l) represents the user-induced EDC calculated by multiplying fuel consumption by a greenhouse gas emission factor and a carbon price. Equation (1m) indicates that the traffic flow on an individual road segment is related to the pavement performance of the entire network and the traffic demand represented by the origin–destination (OD) trip matrix and can be derived by the UE method. Equation (1n) signifies the pavement performance model previously established based on the Bayesian neural network (BNN; Yao et al., 2022a). Maintenance effectiveness was not directly measured but rather simulated using the pavement performance model of maintained segments (Yao et al., 2022a). This was done by first assuming that the pavement performance returns to the initial value after maintenance and then predicting the performance evolutions after specific treatments based on the performance model of maintained segments. Equations (1o)–(1q) are the three constraints on action selection for enhancing algorithm efficiency and providing realistic solutions (Yao et al., 2022b). Equations (1o) and (1p) restrict the choice of M&R or rehabilitation when any of the nondecreasing pavement performance indicators (PPIs) exceeds or non-increasing PPIs falls below the corresponding thresholds  $Th_i^{M\&R}$  and  $Th_i^R$ , respectively. Equation (1q) ensures that no M&R action is carried out if all PPIs are close to their initial states or the segment has just been maintained in the previous year, that is, a segment is not allowed to be maintained for two consecutive years. Equation (1r) is the annual budget constraint.

Once M&R decisions are obtained for the entire pavement network for a given year, the traffic flow on each road section can be determined accordingly based on the UE model:

Minimize:

$$\sum_{m=1}^M \int_0^{v_{t,m}} tc(w, IRI_{t,m}) dw \quad \forall t = 1, \dots, T \quad (2a)$$

subject to:

$$tc(w, IRI_{t,m}) = \text{VOC}_{t,m}^{pv} (IRI_{t,m})$$

$$+ \varphi t_m^f \left[ 1 + \alpha \left( \frac{w}{C_{t,m}} \right)^\beta \right] \quad \forall t = 1, \dots, T, m = 1, \dots, M \quad (2b)$$

$$\text{VOC}_{t,m}^{pv} (IRI_{t,m}) = \sum_{n=1}^N \frac{\text{VOC}_{t,n} (v_{t,n}, IRI_{t,n})}{v_{t,n}} \sigma_{m,n} \text{LDF}_n \quad \forall t = 1, \dots, T, m = 1, \dots, M \quad (2c)$$

$$q_k^{r,s} (t) = \sum_{k \in K_{r,s}} f_k^{r,s} (t) \quad \forall t = 1, \dots, T, r \in R, s \in S \quad (2d)$$

$$v_{t,m} = \sum_{r \in R} \sum_{s \in S} \sum_{k \in K_{r,s}} f_k^{r,s} (t) \delta_{m,k}^{r,s} \quad \forall t = 1, \dots, T, m = 1, \dots, M \quad (2e)$$

$$v_{t,m} \geq 0 \quad \forall t = 1, \dots, T, m = 1, \dots, M \quad (2f)$$

$$f_k^{r,s} (t) \geq 0 \quad \forall t = 1, \dots, T, r \in R, s \in S, k \in K_{r,s} \quad (2g)$$

$$v_{t,n} = \sum_{m=1}^M v_{t,m} \sigma_{m,n} \text{LDF}_n \quad \forall t = 1, \dots, T, n = 1, \dots, N \quad (2h)$$

Equation (2a) is the objective function of the UE model. Equation (2b) is the travel cost function, where the first part denotes the VOC related to pavement surface conditions, and the second part is the TTC estimated based on the BPR function (Bureau of Public Roads, 1964). Equation (2c) denotes the relationship between the VOC of segment  $n$  and the average VOC per vehicle on section  $m$  in year  $t$ . Equation (2d) guarantees flow conservation on each OD pair, that is, the sum of the flow on all paths connecting an OD pair should equal the flow of that OD pair. Equation (2e) means that the total flow of a section is the sum of the flows of the paths that pass through the section. Equations (2f) and (2g) ensure that all flows on sections or paths are non-negative. Equation (2h) distributes the traffic flow on a section to each segment within the section.

### 3 | METHODOLOGY

#### 3.1 | Overview

The problem presented in the previous section can be formulated as a Markov decision process (MDP) with augmented states, incorporating maintenance history, including recent M&R actions, treatment age, pre-treatment pavement performance, and road age as part of the state variables. State augmentation ensures Markov property

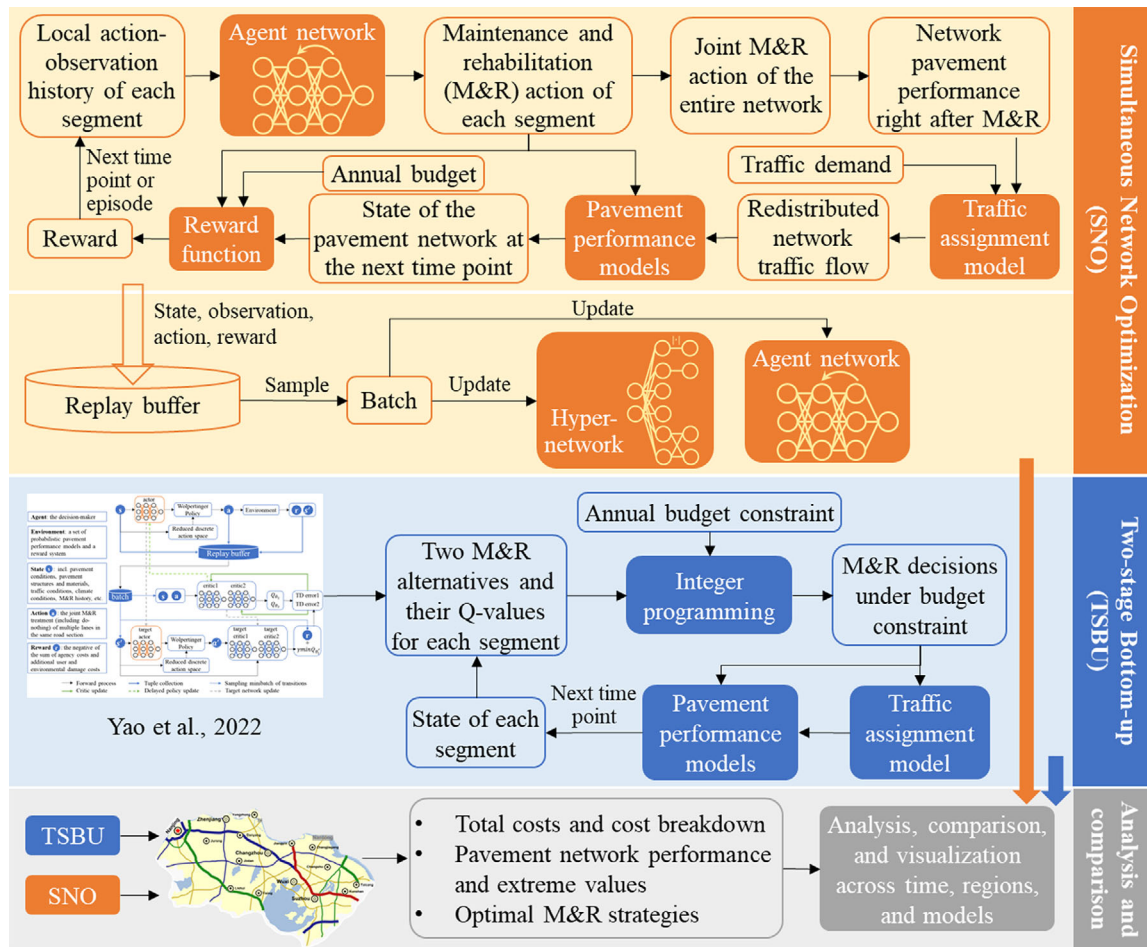


FIGURE 1 An overview of the methodology.

while accounting for the influence of previous management activities on pavement maintenance decision-making (Lee & Madanat, 2015b; Madanat, 1993). The MDP problem can be solved by dynamic programming, but the large state space and network size of the problem in this study make it susceptible to the curse of dimensionality. To address this challenge, this study resorts to the MARL algorithm called QMIX to develop an innovative maintenance optimization model for interdependent pavement networks following the SNO framework. Figure 1 shows an overview of the methodological framework. After the model was trained, the learned M&R strategy represented by the agent network that maps the local action-observation history to Q-values and that further selects the action with an  $\epsilon$ -greedy policy was saved to local storage. This strategy was denoted by SNO in the rest of this paper. In addition, another strategy, derived from the TD3-Wolpertinger model previously built on the TSBU framework (Yao et al., 2022b), referred to in this paper as TSBU, was also considered. The main drawback of TSBU lies in the lack of consideration of the interdependencies among pavement segments. Therefore, applying both strategies to an interdependent pavement network

under traffic equilibrium allows estimating the potential impact of ignoring segment interdependencies in M&R planning. The resulting costs, pavement performance, and optimal M&R schedules were then compared across time, regions, and models. The specific algorithms and methods are described in detail in the following sections.

### 3.2 | MARL and QMIX

MARL holds considerable promise for solving various MAS problems such as strategy games (Samvelyan et al., 2019), robot control (Perrusquía et al., 2021), autonomous vehicles (Schmidt et al., 2022), traffic signal control (Chen et al., 2021), and infrastructure asset management (Asghari et al., 2023). However, in the field of pavement management, none of the existing studies have so far employed MARL for network-level long-term M&R scheduling. Only a few research have recently applied single-agent RL to segment-level pavement maintenance optimization (Renard et al., 2021; Shani et al., 2021; Yao et al., 2020). Nevertheless, in an MAS setting where agents interact with each other, the environment is no longer stationary



from agents' local perspectives, which renders the Markov property in single-agent RL invalid (Hernandez-Leal et al., 2019). Therefore, this study resorts to MARL to solve the maintenance optimization problem for interdependent pavement networks.

### 3.2.1 | Decentralized partially observable MDP (Dec-POMDP)

In real-world situation, a network of highway infrastructures is typically managed by a central agency and multiple regional units under the jurisdiction of the central agency (Amin et al., 2022). This makes the decision makers behind individual pavement segments (i.e., individual agents) only have partial observability of the road network. In other words, agents do not have complete information about the state of the environment when interacting with the environment. Meanwhile, in an interdependent pavement network, each pavement segment works together to achieve system-wide goals. Although M&R decisions on different segments are constrained by the total available budget, it is still not considered a competitive task because segments acquire budgets to maximize system benefits rather than to defeat each other. The partially observable environment and the cooperative multi-agent task allow the problem in this study to be modeled as a Dec-POMDP (Rashid et al., 2018), which is defined by a tuple  $G = \langle A, S, U, P, r, O, \gamma \rangle$ . At every time step, each agent  $a \in A$  perceives its own observation  $o_t^a \in O$  and chooses an action  $u_t^a$  accordingly, forming a joint action  $u_t = \times_{a \in A} u_t^a \in U$ . This leads to the state transition from  $s_t$  to  $s_{t+1}$  according to the state transition function  $P(s_{t+1}|s_t, u_t)$ . A global reward  $r(s_t, u_t)$  is then received and the cooperative agents learn decentralized policies to maximize the return of the global reward  $R = \sum_t \gamma^t r(s_t, u_t)$ .  $\gamma \in [0, 1]$  is the discount factor.

### 3.2.2 | Centralized training and decentralized execution (CTDE)

Decentralized policy is necessary when agents cannot access the full state during the execution stage. It relies solely on each agent's local action-observation history and can mitigate the issue of joint action spaces expanding exponentially with the number of agents (Rashid et al., 2018). Decentralized policies are usually learned in a centralized manner in a simulation environment or laboratory setting. This grants agents access to global state information during the training stage that would otherwise be hidden from agents (Rashid et al., 2018). Such a paradigm

is called CTDE, which has recently attracted wide attention in the RL community.

### 3.2.3 | QMIX

The CTDE-based MARL algorithms can be generally divided into two groups: value-based methods such as value decomposition networks (Sunehag et al., 2017) and QMIX (Rashid et al., 2018) and actor-critic methods such as counterfactual multi-agent policy gradients (Foerster et al., 2018). The QMIX algorithm was employed in this study due to its outstanding performance in dealing with multi-agent tasks in the literature (Chen et al., 2021; Hu et al., 2021). QMIX mainly has three types of networks: an agent network for each agent to select an action independently according to its local action-observation history, a hypernetwork using global state as input to generate weights and biases for the mixing network, and a mixing network to mix the value function of different agents  $Q_a$  to a joint action-value function  $Q_{tot}$  in a nonlinear way. To ensure optimal individual action consistent with optimal joint action, the monotonicity constraint needs to be satisfied in the mixing network:

$$Q_{tot}(\tau, u) = g\left[\left(Q_1(\tau^1, u^1), \dots, Q_a(\tau^a, u^a) \dots\right) | W, B\right] \quad (3)$$

$$\frac{\partial Q_{tot}}{\partial Q_a} \geq 0, \forall a \in A \quad (4)$$

where  $g(\cdot)$  denotes the mixing network parameterized by  $W, B$  generated by the hypernetwork.  $\tau^a$  is the action-observation history of agent  $a$ . To enforce this monotonicity constraint, the weights of the mixing network  $W$  are restricted to be nonnegative. This is achieved by adding an absolute activation function to the output layer of the hypernetwork that generates the weights. QMIX learns by sampling a batch of transitions from the replay buffer and minimizing the sum of squares of temporal-difference error loss:

$$L(\theta) = \sum_{i=1}^b \left[ \left( y_i^{tot} - Q_{tot}(\tau_i, u_i, s_i; \theta) \right)^2 \right] \quad (5)$$

$$y_i^{tot} = r_i + \gamma \max_{u_i'} Q_{tot}(\tau_i', u_i', s_i'; \theta^-) \quad (6)$$

where  $b$  is the batch size sampled from the replay buffer,  $\theta$  is a set of parameters of the agent and hyperparameter networks,  $\theta^-$  is a set of parameters of the corresponding target networks as in Deep Q-Network (DQN) (Rashid et al., 2018), and variables with single quotes in the upper right corner indicate the variables for the next time point.



### 3.3 | QMIX model in pavement maintenance optimization

In this section, a QMIX model dedicated to maintenance optimization for interdependent pavement networks is developed by defining the elements in QMIX in the pavement maintenance optimization setting, constructing the simulation environment, and designing the reward function.

#### 3.3.1 | Definition of model elements

Assuming that there is a decision maker behind each pavement segment who is responsible for developing the M&R plan, then this decision maker is the agent, and all segments of a pavement network constitute an MAS. The environment in which agents get local observations, interact with each other, execute actions, and receive rewards was simulated using the pavement performance models (Yao et al., 2022a), traffic assignment model and reward function. Observations are information available to individual agents for making M&R decisions, which in this study refer to the input variables of the pavement performance models (Yao et al., 2022a), including variables related to pavement materials and structures, pavement performance, traffic and climate conditions, maintenance history, and so forth. States are the joint observations of all segments in the road network. Hence, the transition of observations and states is mainly achieved by using the probabilistic pavement performance models built on BNN to predict the future pavement performance after taking specific M&R actions, combined with some logical judgments and reasonable assumptions, such as increasing the road age by one for each subsequent year, assuming a constant climate, and so forth. Actions refer to the available M&R treatments plus the “do-nothing” as shown in Table 3. The reward is equivalent to the negative total cost (i.e., the sum of ACs and extra UCs and ECDs) between two consecutive time points. The calculation of each cost item refers to Equation (1), and the data source used is the same as the study by Yao et al. (2022b).

#### 3.3.2 | Model structure

The overall structure of the QMIX model as well as the meaning of the model elements as defined in the previous section are exhibited in Figure 2. At each time step of the decentralized execution, each agent observes its local action-observation history, which is then fed into the agent network to select an M&R action. The state of the pavement network is transferred to an intermediate state right

after M&R, and the network traffic flow is redistributed accordingly. Based on this, the next state can be forecasted using the pavement performance models and a global reward is received. The transition  $(o_t, s_t, u_t, r_t, o_{t+1}, s_{t+1})$  is recorded in a replay buffer. In centralized training, a batch of transitions was extracted from the buffer to calculate the temporal-difference error according to Equations (5)–(6) and update the model parameters.

### 3.4 | TSBU model

To understand the potential impact of ignoring segment interdependencies in pavement M&R planning, the previously established TSBU model was also introduced. In this TSBU model, the segment-level maintenance optimization problem that seeks to minimize long-term cumulative costs was addressed by the TD3-Wolpertinger algorithm (Yao et al., 2022b), which falls within the scope of single-agent RL. It identified two M&R treatment candidates for each segment. Then, at the network-level, the integer programming method was employed to determine the final M&R decision that minimizes the total long-term cumulative cost of the entire pavement network while satisfying the annual budget constraint. This process was repeated in each year of the planning horizon.

## 4 | NUMERICAL EXAMPLES

### 4.1 | Basic settings

The developed models and methods were demonstrated by applying them to a real-world-scale expressway pavement network south of the Yangtze River in Jiangsu Province, China. The required data, including the pavement structures and materials, pavement conditions, traffic loads and OD data, M&R histories, climate conditions, and so forth, were extracted from the PMS of Jiangsu. All expressways are divided into sections based on their different characteristics and further divided at 1-km intervals. A total of 6364 segments are obtained, each of which is a one-way, one-lane pavement segment of about 1 km. An annual budget of 2.5 billion and a 20-year planning horizon were considered. The penalty coefficient was set to five in this study. Four PPIs were incorporated into the model, including the rutting depth (RD), IRI, side-way force coefficient (SFC), and transverse cracks evaluation index (TCEI), where IRI is a component of the reward function, and the other three are inserted into the constraints. The baseline IRI was set to 1.0 m/km. The three types of thresholds related to action selection are the same as in Yao et al. (2022b).





TABLE 3 Actions in the QMIX model.

ID	M&R treatment	Unit cost (CNY/m <sup>2</sup> )	Category
1	Seal coating	15	Preventive maintenance
2	Micro-surfacing	22	
3	Hot-in-place rehabilitation	96	
4	Fine mill and fill	110	
5	Thin overlay	71	
6	Fine mill and fill and thin overlay	110	
7	Mill and fill the upper asphalt layer	134	Rehabilitation
8	Overlay with porous asphalt concrete–13	140	
9	Overlay with asphalt-rubber asphalt concrete (ARAC)–13	137	
10	Overlay with styrene-butadiene-styrene-modified AC-13	116	
11	Mill and fill the upper and middle asphalt layer	294	
12	Mill and fill the entire asphalt layer	573	

Note: AC is a dense-graded mixture, and the number “13” denotes the nominal maximum aggregate size in millimeters.

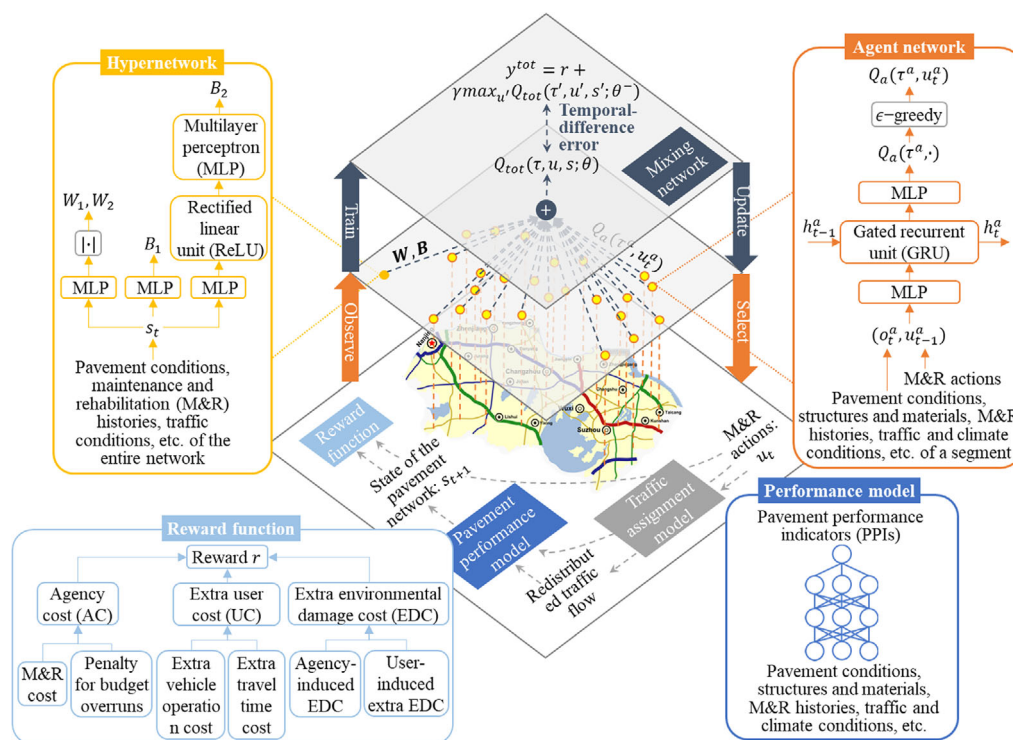


FIGURE 2 Overall structure of the QMIX model.

A new method for imposing the vertical constraint on action selections was proposed in order to ensure that the pavement elevation in the same section after M&R is the same. This was done through the following steps: (1) checking whether performing the actions selected by the agents (denoted as joint action A, such as those joint actions at the section level consisting of “do-nothing” and “overlay with Styrene-Butadiene-Styrene [SBS] modified AC-13”) would violate the vertical constraint; (2) in sections where the constraint is violated, calculating the additional over-

lay cost for segments with lower elevations rising to the highest elevation in this section (denoted as joint action B, where the action for all lanes of the above section becomes “overlay with SBS modified AC-13”); (3) adding this additional overlay cost to the AC while assuming the additional overlay thickness cannot bring any improvement to the pavement performance. The rationality of this approach lies in the fact that overlay thickness is generally considered to be a measure of rehabilitation intensity (Lee & Madanat, 2015a; Ouyang & Madanat, 2004) Segments



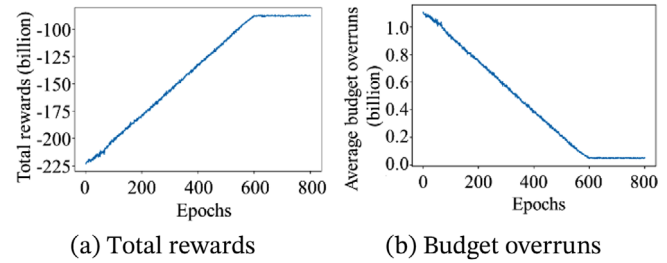
**TABLE 4** The values of the hyperparameters in the QMIX model.

Hyperparameters	Values
Optimizer	Adam
Learning rate	0.0005
Replay buffer size (episodes)	60
Batch size (episodes)	16
Initial $\epsilon$	1
Minimum $\epsilon$	0.05
$\epsilon$ anneal steps	12000
Number of episodes in an epoch	1
Total epochs	800
Update frequency of the target network parameters (epochs)	5
Number of hidden neurons in the agent network, hypernetwork, and mixing network	128, 128, and 256

rehabilitated with thicker overlays often lead to better performance. Therefore, agents selecting joint action A will produce the same cost but worse performance than joint action B, which is less cost-effective than agents directly selecting joint action B. This will motivate agents not to choose actions that violate the vertical constraint.

In addition, an approach for splitting the OD trip whose origin or destination is outside the studied pavement network was also proposed. This was achieved with the help of the Baidu Map Application Programming Interface (API) v2.0. First, the longitude and latitude of each toll station or hub were extracted. Then, for each of the cross-regional OD trips, the route with the shortest travel time was identified. Last, the node (origin or destination) outside the studied network was changed to the node on this route that is inside the network and closest to the network boundary. For the present study, this process is much easier as the OD trip across the Yangtze River must pass through a limited number of bridges.

The QMIX model was coded in Python 3.8.3. Table 4 shows the values of the hyperparameters in the developed QMIX model, which were determined by referring to the literature (Hu et al., 2021; Rashid et al., 2018) and undergoing a series of fine-tuning. In this study, a step represents a state transition with an interval of 1 year. An episode denotes a sequence of states, observations, actions, and rewards, which ends with a terminal state and has the same time span as the planning horizon. Hence, an episode consists of 20 steps in this study. Replay buffer is to store trajectories of experience that will be used for training. Batch size is the number of episodes sampled from the buffer for each training. Epsilon  $\epsilon$  refers to the probability of exploration, that is, the probability of choosing a random



**FIGURE 3** Learning curves of the QMIX model.

action, relative to exploiting the agents' current action-value estimates. The  $\epsilon$  anneal steps mean the number of steps required to reduce from the initial  $\epsilon$  to the minimum  $\epsilon$ . An epoch is an update of the network parameters, and there can be one or more episodes in an epoch, which was set to 1 after tuning in this study. The parameters of the target network were copied from the evaluation network every five epochs, and 800 epochs were sufficient for the convergence of the model in this study.

## 4.2 | Results and discussion

### 4.2.1 | Convergence performance

The evolution of the total reward that is equivalent to the negative total cost over the planning horizon and the average budget overrun calculated by Equation (7) are presented in Figure 3. It can be found that the total reward steadily increases over time and eventually converges to a stable value. The budget overrun is also gradually alleviated and finally becomes almost zero.

Average budget overrun

$$= \frac{1}{T} \sum_{t=1}^T \max \left[ \left( \sum_{n=1}^N MRC_{t,n}(x_{t,n}) - B_t \right), 0 \right] \quad (7)$$

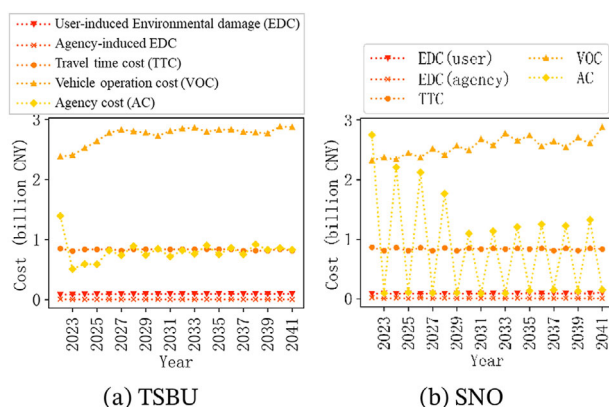
### 4.2.2 | Total cost and cost breakdown

To investigate the potential impact of ignoring segment interdependencies in M&R planning, the total cost as well as the cost breakdown were compared across the two models as shown in Table 5. It is noteworthy that for an interdependent pavement network, the total cost incurred by SNO is slightly lower than that of TSBUE, about 3.0% lower in this case. Among the many cost items, the TTC and agency-induced EDC of the two models are nearly identical. The most significant cost savings of SNO lies in the VOC and user-induced EDC, with SNO producing about 6.7% less VOC and 4.9% less user-induced EDC than

**TABLE 5** Breakdown of the total costs.

Cost items (billion CNY)	TSBU	SNO	Percentage difference (%)
AC	16.12	17.17	6.5
Vehicle operating cost	54.9	51.23	-6.7
TTC	16.63	16.66	0.2
Agency-induced EDC	0.09	0.09	0.0
User-induced EDC	1.82	1.73	-4.9
Total cost	89.57	86.89	-3.0

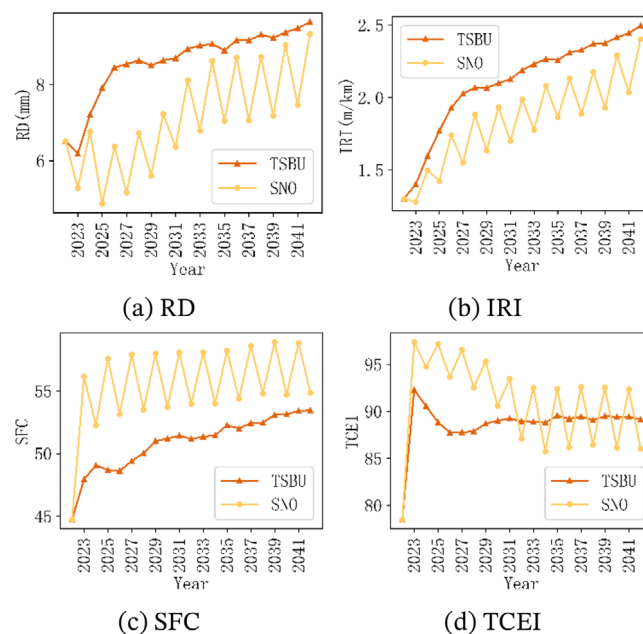
Abbreviations: SNO, simultaneous network optimization; TSBU, two-stage bottom-up.

**FIGURE 4** Costs of the two models in different years.

TSBU. These savings come at the cost of a 6.5% increase in AC, but this does not offset the VOC savings.

Figure 4 further shows the various costs of the two models in different years. It can be observed that the distribution of AC in different years varies significantly between the two models. The ACs of TSBU are relatively evenly distributed across years, except for the first few years. In contrast, an obvious periodic variation is observed from the ACs of SNO, implying that SNO tends to concentrate maintenance resources. This may be because concentrated M&R schedules allow for simultaneous improvements in the performance of the entire pavement network, thus enabling drivers to greatly optimize their routes. In addition, the VOC of SNO is clearly lower than that of TSBU and shows a slight yearly increasing trend.

Possible reasons for the insignificant difference in the total cost of the two models were also analyzed, which are twofold. First, although TSBU was obtained by assuming that the pavement segments are independent of each other, its flexibility allows the decision results to be quickly adjusted to changes in network traffic flows during the application phase. Second, the length difference between alternative routes is generally large for an access-controlled highway network. Therefore, drivers will only change routes if the alternate route has a clear advantage

**FIGURE 5** Traffic-length weighed network performance.

in pavement conditions such that VOC savings are sufficient to offset the increased TTC. In order to quantify the pavement conditions of alternative routes that may lead to vehicle rerouting, the maximum IRI value that makes the total travel cost of the alternative route lower than the shortest route is calculated for all possible IRI values on the shortest route. Route information was extracted from the Baidu Map API. VOC and TTC were calculated by the method in (Chatti & Zaabar, 2012) and based on the BPR function (Bureau of Public Roads, 1964), respectively. Table 6 shows the results, including the 25th percentile, mean, and 75th percentile of the maximum IRI values for all OD pairs that make alternative routes less costly to travel. However, not all OD pairs have such an IRI value to make the alternate route preferable since the IRI also needs to be no less than zero. Hence, the percentage of OD pairs for which no valid IRI value exists was also calculated. The common IRI values of the highway asphalt pavements in Jiangsu are generally less than 3 m/km. Therefore, it can be concluded that at general IRI values (0–3 m/km), there will be over 70% of OD pairs unlikely to be rerouted from the shortest route to the alternate route. This also partly explains why the TSBU model, which ignores segment interdependencies, does not result in a significant economic loss in this case.

#### 4.2.3 | Pavement performance

Figure 5 presents the evolution of the pavement network performance measured by the traffic-length weighed


**TABLE 6** Presence of IRI values that give alternate routes priority over shortest routes.

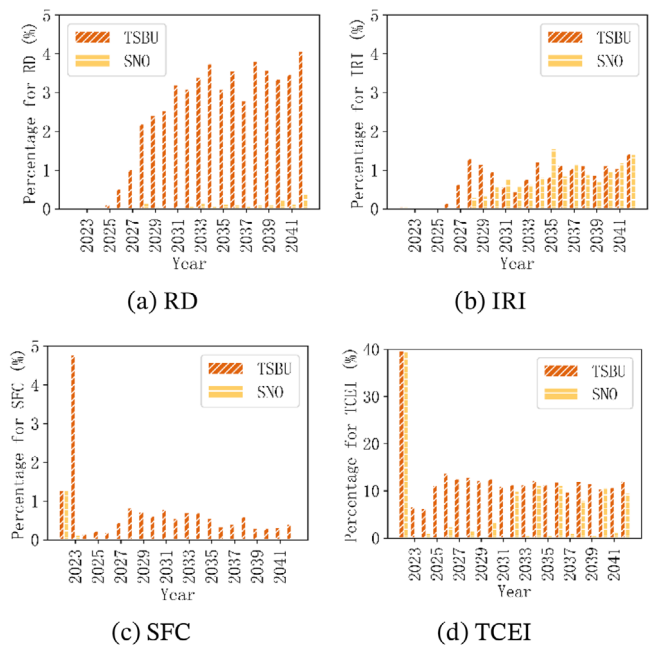
Vehicle types	Maximum IRI on alternate routes (m/km)	IRI on the shortest route (m/km)					
		1	2	3	4	5	6
Medium car	25th percentile (m/km)	0.3	0.5	0.9	1.5	3.1	4.0
	Mean (m/km)	0.5	1.0	1.6	2.5	3.6	4.5
	75th percentile (m/km)	0.8	1.5	2.4	3.5	4.6	5.5
	N/A percentage (%)	88.7	78.9	70.3	55.8	35.1	20.7
Medium truck	25th percentile (m/km)	0.3	0.6	0.8	1.9	3.5	4.2
	Mean (m/km)	0.5	1.0	1.5	2.7	3.9	4.7
	75th percentile (m/km)	0.7	1.5	2.3	3.6	4.7	5.6
	N/A percentage (%)	91.5	84.7	77.5	58.9	33.7	17.8
Articulated truck	25th percentile (m/km)	0.3	0.5	0.9	1.6	3.4	4.1
	Mean (m/km)	0.5	1.0	1.6	2.5	3.8	4.5
	75th percentile (m/km)	0.8	1.5	2.4	3.5	4.7	5.5
	N/A percentage (%)	88.7	78.8	70.3	55.4	29.8	16.9

indicator (TWI) due to the application of the two models. TWI was calculated for each of the four PPIs considered in this study, with the following equation:

$$TWI = \frac{\sum_{n=1}^N AADT_n \cdot length_n \cdot PPI_n}{\sum_{n=1}^N AADT_n \cdot length_n} \quad (8)$$

where  $AADT_n$  and  $length_n$  are the annual average daily traffic and the length of segment  $n$ , respectively. As can be seen from Figure 5, SNO produces significantly better pavement conditions than TSBU, reducing the pavement network RD by an average of 17.5% and IRI by an average of 12.4%, and improving the pavement network SFC by an average of 9.0% and TCEI by an average of 2.7%. This is natural because SNO has been developed taking into account the impact of decisions made on other segments of the network, whereas TSBU can only passively adjust M&R decisions as network traffic flows change.

Although agents are forced to select rehabilitation actions when pavement conditions are worse than the corresponding thresholds, there are still some segments where the conditions are already below the thresholds before the actions are executed. This is often due to the agents' inability to forecast the rapid deterioration in pavement conditions that may follow when the current pavement conditions are not very poor. Figure 6 shows the percentage of segments with performance worse than the rehabilitation thresholds. It clearly demonstrates that TSBU is more likely to fail to maintain pavements on time before they decline below the thresholds. This also implies that SNO is more far-sighted than TSBU.


**FIGURE 6** Percentage of pavement segments with performance worse than the rehabilitation thresholds.

The spatial distribution of segments with different pavement conditions is also illustrated in Figure 7. This was done by first extracting the geographic data from OpenStreetMap using the OSMnx library in Python and then mapping the pavement conditions to the corresponding locations using different colors. As some road sections fall outside the jurisdiction of the central agency and some have incomplete data, they are not considered in the pavement maintenance optimization model, but they are still involved in the traffic assignment model as shown



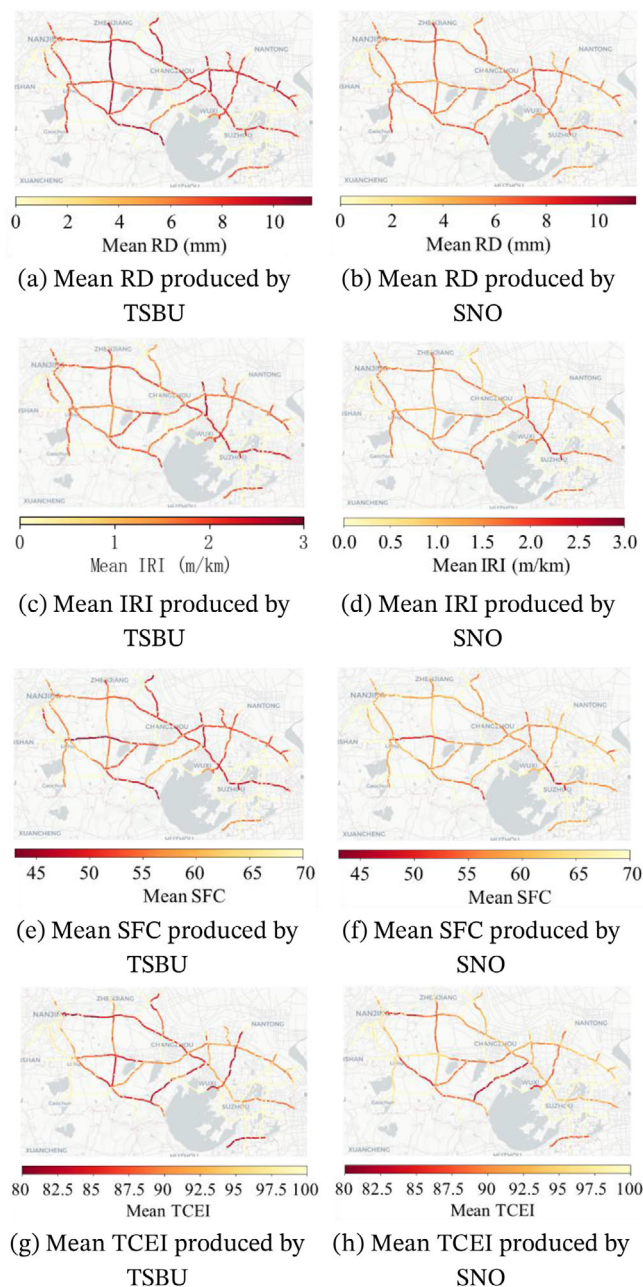


FIGURE 7 Spatial distribution of pavement conditions produced by the two models.

in the lightest color in Figure 7 to ensure the integrity of the displayed road network. It can be seen that SNO has significant advantages over TSBU in terms of various pavement performance. Moreover, the distributions of pavement segments with relatively poor conditions measured by different PPIs are significantly different. For example, segments with lower IRI values do not necessarily have better rutting, skid, and cracking resistance, which justifies the need for multi-indicator decision-making.

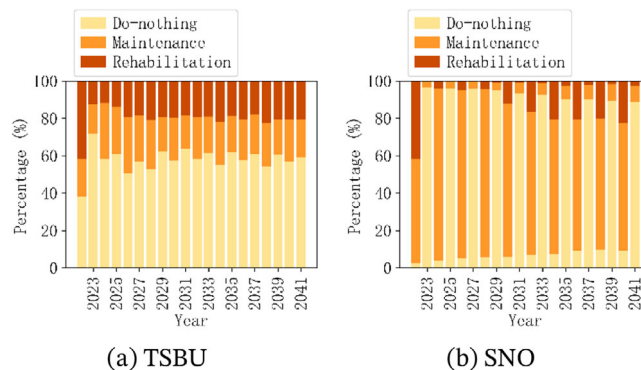


FIGURE 8 Percentage of different maintenance and rehabilitation types in each year produced by the two models.

#### 4.2.4 | Optimal maintenance strategies

To understand the difference in optimal maintenance strategies generated by the two models, the percentage of different M&R types in each year was compared as exhibited in Figure 8. The SNO model seems to prefer concentrated M&R schedules, compared to the TSBU model. Although both models constrain that a segment cannot be maintained for 2 consecutive years, the M&R schedule of the TSBU model is much more stable. Hence, there are other reasons for the concentrated M&R schedules of the SNO model. In the study by Uchida and Kagaya (2006), the repair of links was also concentrated in certain years, so they conclude that if repair work is made on a link, it is better to repair the whole link in terms of life cycle costs minimization. Hence, the reason may be that the concentrated M&R schedules could improve the performance of the entire network, thereby facilitating travel route optimization because travel costs depend on the condition of all segments along the route, not just one. In addition, the maintenance ratio of SNO is considerably higher than that of TSBU, while the rehabilitation rate is obviously lower. In other words, the SNO model tends to reduce costly rehabilitations through more frequent maintenance, a pattern also found in the study of Guan et al. (2022).

## 5 | CONCLUSION

This study follows the SNO framework and introduces the MARL algorithm QMIX for the first time to develop an innovative maintenance optimization model for interdependent highway pavement networks. The performance of the proposed model was demonstrated by applying it to a real-world-scale highway network in Jiangsu, China. The resulting M&R strategy was then compared with the one derived from the previously established TSBU model in terms of costs, performance, and optimal M&R schedules



to estimate the potential impacts of ignoring segment interdependencies in pavement maintenance planning.

When implemented on an interdependent pavement network of more than 6000 lane km, the SNO model generates a 3.0% lower total cost than TSBU, with a major cost reduction in VOC but a slightly higher AC. The SNO model produced significantly better pavement conditions, with average performance improvements of up to 17.5%. Hence, for access-controlled highway networks, ignoring the functional dependence among road segments in maintenance planning will not cause significant economic losses but may jeopardize pavement performance. Two possible reasons for the small cost differences between the two models were identified. First, the flexibility embedded in TSBU allows decision makers to flexibly adjust M&R decisions in response to changes in traffic flow and pavement conditions. Second, for access-controlled highway networks, the pavement performance advantages are often not sufficient for most OD pairs to reroute vehicles from the shortest route to the alternative route. Therefore, for road networks with many alternative routes of similar length, the developed model will be more applicable and has the potential to bring greater benefits. Otherwise, the TSBU model may be sufficient and computationally more efficient.

The SNO model is less likely to have a situation where M&R actions are not conducted on time causing pavement conditions to decline below the threshold. This indicates that the SNO model is more far-sighted than TSBU, as it was formulated taking into account the impact of M&R decisions on other segments. Moreover, the SNO model prefers concentrated M&R schedules, possibly because this could improve the performance of the entire network, thereby facilitating travel route optimization. It also tends to reduce costly rehabilitation through more frequent preventative maintenance. The value mixing method and monotonicity constraint in the QMIX algorithm also make its complexity linearly related to the number of agents (Fu et al., 2020), that is, the number of pavement segments. Thus, the developed model is expected to be applicable to large-scale road networks.

Despite the contributions of this study, there is still room to further improve the research. First, time-varying traffic demand can be considered to more realistically capture the functional dependence among road segments. Second, the impact of M&R schedule coordination on network traffic distribution and M&R costs, which are another functional and economic dependence, respectively, can also be explored in future studies. Third, the performance of the developed model can be compared with others, such as evolutionary algorithm-based models, to verify the superiority of the model. Alternatively, the model can be applied to problems that are reduced in size and for which the opti-

mal solution is easy to obtain to assess the optimality of the solution.

## ACKNOWLEDGMENTS

The authors thank the Research Institute for Sustainable Urban Development (RISUD) at the Hong Kong Polytechnic University for their support in conducting this study. Gratitude is also expressed to the engineers and researchers who contributed to the establishment and data collection of the Pavement Management System in Jiangsu province, China, which was used in this research.

## CONFLICT OF INTEREST STATEMENT

This paper is an original unpublished work and the manuscript or any variation of it has not been submitted to another publication.

## REFERENCES

- Adeli, H., & Ghosh-Dastidar, S. (2004). Mesoscopic-wavelet freeway work zone flow and congestion feature extraction model. *Journal of Transportation Engineering*, 130(1), 94–103.
- Adeli, H., & Jiang, X. (2008). *Intelligent Infrastructure: Neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures*. CRC Press.
- Adeli, H., & Karim, A. (1997). Scheduling/cost optimization and neural dynamics model for construction. *Journal of Construction Engineering and Management*, 123(4), 450–458.
- Amin, M. A., Pan, S., & Zhang, Z. (2022). Pavement maintenance and rehabilitation budget allocation considering multiple objectives and multiple stakeholders. *International Journal of Pavement Engineering*, 24(2), 2027941.
- Asghari, V., Biglari, A. J., & Hsu, S.-C. (2023). Multiagent reinforcement learning for project-level intervention planning under multiple uncertainties. *Journal of Management in Engineering*, 39(2), 04022075.
- Bureau of Public Roads. (1964). *Traffic Assignment Manual for Application with a Large, High Speed Computer*. US Department of Commerce. Urban Planning Division, Washington, DC. <https://books.google.com.hk/books?id=gkNZAAAAMAAJ&printsec=frontcover&hl=zh-CN#v=onepage&q&f=false>
- Cao, R., Leng, Z., Yu, J., & Hsu, S.-C. (2020). Multi-objective optimization for maintaining low-noise pavement network system in Hong Kong. *Transportation Research Part D: Transport and Environment*, 88, 102573.
- Chatti, K., & Zaabar, I. (2012). *Estimating the effects of pavement condition on vehicle operating costs*. Transportation Research Board.
- Chen, X., Xiong, G., Lv, Y., Chen, Y., Song, B., & Wang, F.-Y. (2021). A collaborative communication-Qmix approach for large-scale networked traffic signal control. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, IN (pp. 3450–3455).
- Chu, J. C., & Chen, Y.-J. (2012). Optimal threshold-based network-level transportation infrastructure life-cycle management with heterogeneous maintenance actions. *Transportation Research Part B: Methodological*, 46(9), 1123–1143.
- de la Garza, J. M., Akyildiz, S., Bish, D. R., & Krueger, D. A. (2011). Network-level optimization of pavement maintenance renewal strategies. *Advanced Engineering Informatics*, 25(4), 699–712.





- Durango-Cohen, P. L., & Sarutipand, P. (2007). Capturing interdependencies and heterogeneity in the management of multifacility transportation infrastructure systems. *Journal of Infrastructure Systems*, 13(2), 115–123.
- Durango-Cohen, P. L., & Sarutipand, P. (2009). Maintenance optimization for transportation systems with demand responsiveness. *Transportation Research Part C: Emerging Technologies*, 17(4), 337–348.
- Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, LA.
- Fu, Z., Zhao, Q., & Zhang, W. (2020). Reducing overestimation in value mixing for cooperative deep multi-agent reinforcement learning. *Proceedings of the International Conference on Agents and Artificial Intelligence*, Valletta, Malta.
- Gao, L., Xie, C., Zhang, Z., & Waller, S. T. (2011). Integrated maintenance and expansion planning for transportation network infrastructure. *Transportation Research Record*, 2225(1), 56–64.
- Gao, L., Xie, C., Zhang, Z., & Waller, S. T. (2012). Network-level road pavement maintenance and rehabilitation scheduling for optimal performance improvement and budget utilization. *Computer-Aided Civil and Infrastructure Engineering*, 27(4), 278–287.
- Guo, F., Gregory, J., & Kirchain, R. (2020). Incorporating cost uncertainty and path dependence into treatment selection for pavement networks. *Transportation Research Part C: Emerging Technologies*, 110, 40–55.
- Guo, F., Zhao, X., Gregory, J., & Kirchain, R. (2021). A weighted multi-output neural network model for the prediction of rigid pavement deterioration. *International Journal of Pavement Engineering*, 23(8), 2631–2643.
- Guan, J., Yang, X., You, L., Ding, L., & Cheng, X. (2022). Multi-objective optimization for sustainable road network maintenance under Traffic equilibrium: Incorporating Costs and environmental impacts. *Journal of Cleaner Production*, 334, 130103.
- Hajibabai, L., Bai, Y., & Ouyang, Y. (2014). Joint optimization of freight facility location and pavement infrastructure rehabilitation under network traffic equilibrium. *Transportation Research Part B: Methodological*, 63, 38–52.
- Hernandez-Leal, P., Kartal, B., & Taylor, M. E. (2019). A survey and critique of multiagent deep reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 33(6), 750–797.
- Hong, F., & Prozzi, J. A. (2010). Roughness model accounting for heterogeneity based on in-service pavement performance data. *Journal of transportation Engineering*, 136(3), 205–213.
- Hu, J., Jiang, S., Harding, S. A., Wu, H., & Liao, S.-W. (2021). Rethinking the implementation tricks and monotonicity constraint in cooperative multi-agent reinforcement learning. arXiv e-prints, arXiv: 2102.03479. <https://arxiv.org/abs/2102.03479>
- Hudson, W. R., Haas, R., & Pedigo, R. D. (1979). *Pavement management system development* (NCHRP Report No. 215). Transportation Research Board.
- Lee, J., & Madanat, S. (2014). Joint optimization of pavement design, resurfacing and maintenance strategies with history-dependent deterioration models. *Transportation Research Part B: Methodological*, 68, 141–153.
- Lee, J., & Madanat, S. (2015a). A joint bottom-up solution methodology for system-level pavement rehabilitation and reconstruction. *Transportation Research Part B: Methodological*, 78, 106–122.
- Lee, J., & Madanat, S. (2015b). Jointly optimal policies for pavement maintenance, resurfacing and reconstruction. *EURO Journal on Transportation and Logistics*, 4(1), 75–95.
- Lee, J., & Madanat, S. (2017). Optimal policies for greenhouse gas emission minimization under multiple agency budget constraints in pavement management. *Transportation Research Part D: Transport and Environment*, 55, 39–50.
- Liu, C., Du, Y., Wong, S., Chang, G., & Jiang, S. (2020). Eco-based pavement lifecycle maintenance scheduling optimization for equilibrated networks. *Transportation Research Part D: Transport and Environment*, 86, 102471.
- Liu, Z., Balieu, R., & Kringos, N. (2022). Integrating sustainability into pavement maintenance effectiveness evaluation: A systematic review. *Transportation Research Part D: Transport and Environment*, 104, 103187.
- Madanat, S. (1993). Optimal infrastructure management decisions under uncertainty. *Transportation Research Part C: Emerging Technologies*, 1(1), 77–88.
- Mao, X., Yuan, C., & Gan, J. (2019). Incorporating dynamic traffic distribution into pavement maintenance optimization model. *Sustainability*, 11(9), 2488.
- Medury, A., & Madanat, S. (2013). Incorporating network considerations into pavement management systems: A case for approximate dynamic programming. *Transportation Research Part C: Emerging Technologies*, 33, 134–150.
- Medury, A., & Madanat, S. (2014). Simultaneous network optimization approach for pavement management systems. *Journal of Infrastructure Systems*, 20(3), 04014010.
- Ng, M., Lin, D. Y., & Waller, S. T. (2009). Optimal long-term infrastructure maintenance planning accounting for traffic dynamics. *Computer-Aided Civil and Infrastructure Engineering*, 24(7), 459–469.
- Ouyang, Y. (2007). Pavement resurfacing planning for highway networks: Parametric policy iteration approach. *Journal of Infrastructure Systems*, 13(1), 65–71.
- Ouyang, Y., & Madanat, S. (2004). Optimal scheduling of rehabilitation activities for multiple pavement facilities: Exact and approximate solutions. *Transportation Research Part A: Policy and Practice*, 38(5), 347–365.
- Perrusquía, A., Yu, W., & Li, X. (2021). Multi-agent reinforcement learning for redundant robot control in task-space. *International Journal of Machine Learning and Cybernetics*, 12(1), 231–241.
- Rashid, T., Samvelyan, M., Schroeder, C., Farquhar, G., Foerster, J., & Whiteson, S. (2018). Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. *International Conference on Machine Learning*, PMLR, Stockholm Sweden (pp. 4295–4304).
- Renard, S., Corbett, B., & Sweil, O. (2021). Minimizing the global warming impact of pavement infrastructure through reinforcement learning. *Resources, Conservation and Recycling*, 167, 105240.
- Samvelyan, M., Rashid, T., De Witt, C. S., Farquhar, G., Nardelli, N., Rudner, T. G., Hung, C.-M., Torr, P. H., Foerster, J., & Whiteson, S. (2019). *The Starcraft Multi-Agent Challenge*. arXiv preprint arXiv:1902.04043. <https://arxiv.org/abs/1902.04043>
- Santos, J., Ferreira, A., & Flintsch, G. (2017). A multi-objective optimization-based pavement management decision-support system for enhancing pavement sustainability. *Journal of Cleaner Production*, 164, 1380–1393.



- Schmidt, L. M., Brosig, J., Plinge, A., Eskofier, B. M., & Mutschler, C. (2022). *An introduction to multi-agent reinforcement learning and review of its application to autonomous mobility*. arXiv preprint arXiv:2203.07676. <https://ieeexplore.ieee.org/abstract/document/9922205>
- Shani, P., Chau, S., & Swei, O. (2021). All roads lead to sustainability: Opportunities to reduce the life-cycle cost and global warming impact of US roadways. *Resources, Conservation and Recycling*, 173, 105701.
- Slater, H., De Boer, D., Qian, G., & Shu, W. (2020). *2020 China Carbon Pricing Survey*. China Carbon Forum. [http://www.chinacarbon.info/sdm\\_downloads/2020-china-carbon-pricing-survey](http://www.chinacarbon.info/sdm_downloads/2020-china-carbon-pricing-survey)
- Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W. M., Zambaldi, V., Jaderberg, M., Lanctot, M., Sonnerat, N., Leibo, J. Z., & Tuyls, K. (2017). *Value-decomposition networks for cooperative multi-agent learning*. arXiv preprint arXiv:1706.05296. <https://arxiv.org/abs/1706.05296>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.
- Swei, O., Gregory, J., & Kirchain, R. (2019). Embedding flexibility within pavement management: Technique to improve expected performance of roadway systems. *Journal of Infrastructure Systems*, 25(3), 05019007.
- Uchida, K., & Kagaya, S. (2006). Development of life-cycle cost evaluation model for pavements considering drivers' route choices. *Transportation Research Record*, 1985(1), 115–124.
- Wang, F., Zhang, Z., & Machemehl, R. B. (2003). Decision-making problem for managing pavement maintenance and rehabilitation projects. *Transportation Research Record*, 1853(1), 21–28.
- Wardrop, J. G. (1952). Road paper. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3), 325–362.
- Yao, L., Dong, Q., Ni, F., Jiang, J., Lu, X., & Du, Y. (2019). Effectiveness and cost-effectiveness evaluation of pavement treatments using life-cycle cost analysis. *Journal of Transportation Engineering, Part B: Pavements*, 145(2), 04019006.
- Yao, L., Dong, Q., Jiang, J., & Ni, F. (2020). Deep reinforcement learning for long-term pavement maintenance planning. *Computer-Aided Civil and Infrastructure Engineering*, 35(11), 1230–1245.
- Yao, L., Leng, Z., Jiang, J., & Ni, F. (2022a). Modelling of pavement performance evolution considering uncertainty and interpretability: A machine learning based framework. *International Journal of Pavement Engineering*, 23(14), 5211–5226.
- Yao, L., Leng, Z., Jiang, J., & Ni, F. (2022b). Large-scale maintenance and rehabilitation optimization for multi-lane highway asphalt pavement: A reinforcement learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 23(11), 22094–22105.
- Yeo, H., Yoon, Y., & Madanat, S. (2013). Algorithms for bottom-up maintenance optimisation for heterogeneous infrastructure systems. *Structure and Infrastructure Engineering*, 9(4), 317–328.
- Yeo, H., Yoon, Y., & Madanat, S. (2010). Maintenance optimization for heterogeneous infrastructure systems: Evolutionary algorithms for bottom-up methods, sustainable and resilient critical infrastructure systems. In K. Gopalakrishnan & S. Peeta (Eds.), *Sustainable and resilient critical infrastructure systems* (pp. 185–199). Springer.
- Yin, Y., Lawphongpanich, S., & Lou, Y. (2008). Estimating investment requirement for maintaining and improving highway systems. *Transportation Research Part C: Emerging Technologies*, 16(2), 199–211.
- Yu, B., Gu, X., Ni, F., & Guo, R. (2015). Multi-objective optimization for asphalt pavement maintenance plans at project level: Integrating performance, cost and environment. *Transportation Research Part D: Transport and Environment*, 41, 64–74.
- Zhang, H., Keoleian, G. A., Lepech, M. D., & Kendall, A. (2010). Life-cycle optimization of pavement overlay systems. *Journal of infrastructure systems*, 16(4), 310–322.
- Zhang, L., Fu, L., Gu, W., Ouyang, Y., & Hu, Y. (2017). A general iterative approach for the system-level joint optimization of pavement maintenance, rehabilitation, and reconstruction planning. *Transportation Research Part B: Methodological*, 105, 378–400.

**How to cite this article:** Yao, L., Leng, Z., Jiang, J., & Ni, F. (2024). A multi-agent reinforcement learning model for maintenance optimization of interdependent highway pavement networks. *Computer-Aided Civil and Infrastructure Engineering*, 39, 2951–2970. <https://doi.org/10.1111/mice.13234>