Bridge Maintenance, Safety, Management, Digitalization and Sustainability –
Jensen, Frangopol & Schmidt (eds)

Deep reinforcement learning-driven life-cycle management of bridge and pavement systems

A. Bhattacharya, M. Saifullah & K.G. Papakonstantinou
Department of Civil & Environmental Engineering, The Pennsylvania State University, University Park, PA, USA

ABSTRACT: Optimal management of bridge systems and related transportation infrastructure poses multi-faceted challenges, requiring adept inspection and maintenance policies at both system and individual asset levels, to minimize life-cycle costs while considering various operational, risk, and performance constraints. This demanding type of optimization problems entails, among others, high-dimensional aspects, describing multi-component systems, long-planning horizons, diverse probabilistic and deterministic operational objectives and constraints, and inherent uncertainties associated with inspections and stochastic models. Effective coordination among individual component assets considering various inter-dependencies is also essential to enable a true system-based optimal solution. In this work, this optimization problem is formulated within the framework of Partially Observable Markov Decision Processes (POMDPs) and constrained Multi-Agent Deep Reinforcement Learning (MARL). POMDPs offer a principled mathematical approach for sequential decision-making under uncertainty, incorporating Bayesian inference to address the observation/monitoring data uncertainty, and can be suitably scaled to high-dimensional state and action spaces associated with multi-component systems, exploiting the rich representational capacities of deep learning and decentralized control settings of MARL. In this work, the recently developed DDMAC deep reinforcement learning (DRL) algorithm (Deep Decentralized Multi-Agent Actor-Critic) has been successfully deployed based on the Centralized Training and Decentralized Execution (CTDE) formulation. The efficacy and implementation aspects of the developed framework are originally studied in this work based on two existing real-world transportation networks in Virginia and Pennsylvania, USA, following all regulations imposed by the relevant agencies, as well as their overall practices, in an effort to investigate the use of the suggested framework in practical, actual settings. In both cases, DRL results significantly surpass the ones related to current state-of-practice and state-of-the-art policies, providing further support and insights toward the use of DRL-driven policies for infrastructure management.

1 INTRODUCTION

This work presents a multi-agent deep reinforcement learning framework tailored for the inspection and maintenance (I&M) planning of extensive transportation and bridge networks. This framework is methodically applied to two existing networks in Pennsylvania and Virginia, in the U.S., serving as practical, indicative, real-world case studies.

Significant computational challenges are encountered in managing multi-component infrastructure systems. These challenges include, but are not limited to, the heterogeneity of different asset classes, the extensive number of components, the presence of noisy data, limited availability of resources, and diverse performance-based constraints. The existing optimization methodologies for I&M planning range from threshold-based strategies with risk-based principles to decision tree analysis and renewal theory (Luque & Straub, 2019; Frangopol, et al., 1997). Despite their merits, many of these solutions suffer from optimality-, scalability-, and uncertainty-

DOI: 10.1201/9781003483755-400

induced complexities and are often not easily extendable to environments with constraints. Stochastic optimal control approaches based on Partially Observable Markov Decision Processes (POMDPs), e.g., in (Madanat, 1993; Papakonstantinou & Shinozuka, 2014) can support dynamic policies and noisy real-time observations in the decision-making framework, but are often hard to scale to large systems with multiple components.

To address these limitations, we combine the principles of POMDPs with the advanced capabilities of multi-agent Deep Reinforcement Learning (DRL), as explained in (Andriotis & Papakonstantinou, 2019; Andriotis & Papakonstantinou, 2021) for the Deep Centralized Multi-agent Actor-Critic (DCMAC) technique and its subsequent variant, the Deep Decentralized Multi-agent Actor-Critic (DDMAC). These techniques are designed to manage the complexities of decentralized execution and to efficiently navigate the vast parameter spaces, typical in large-scale infrastructure systems. Lately, as a further development, DDMAC is also shown under the paradigm of Centralized Training and Decentralized Execution (CTDE) (Lyu, et al., 2021) in (Saifullah, et al., 2024), with decentralization at both the action and information levels. This approach reduces the scalability involved complexities even further, by providing only local observations to the actors.

In this paper, the DDMAC-CTDE framework is uniquely applied to analyze two separate, existing, real-world transportation networks in Pennsylvania and Virginia, USA, and the resulting policies are compared against the respective current state-of-practice/state-of-the-art. Both networks under consideration consist of a set of stochastically deteriorating bridge assets necessitating maintenance, considering various deterministic and stochastic resource and condition constraints and targets, aligning closely with the actual methodologies and practices utilized by the respective transportation agencies. The Virginia network also considers multiple asset classes, i.e., pavement and bridge components, utilizing probabilistic deterioration models for the different classes of assets (Saifullah, et al., 2024), and has also been mentioned in (Saifullah, et al., 2022). In each case, results are compared with optimized Condition Based Maintenance policies (CBM) and compatible variants of current I&M policies implemented by the Pennsylvania and Virginia Department of Transportations (PennDOT and VDOT), respectively, with DDMAC-CTDE significantly outperforming both the CBM and the current state-of-practice approaches.

2 BACKGROUND

2.1 *Partially observable markov decision processes*

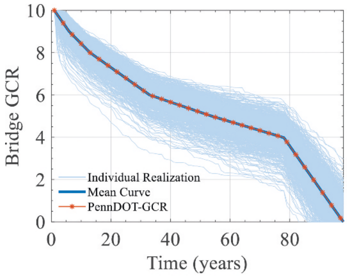
The POMDP framework is defined by the 7 elements tuple consisting of S, A, $\mathbf{P}$, Ω, $\mathbf{O}$, $\mathbf{C}$, and γ, where S, A, and Ω are sets of states, actions, and observations, respectively, $\mathbf{P}$ is the model of state transitions, $\mathbf{O}$ is the observation model, $\mathbf{C}$ is the cost function, and γ is a discount factor. In POMDPs the agent starts at a state s_t at a time step t, takes an action a_t, receives a cost c_t transitions to the next state s_{t+1} based on the transition model $p(s_{t+1}|s_t, a_t)$, and receives an observation, $o_{t+1} \in \Omega$ based on the observation model, $p(o_{t+1}|s_{t+1}, a_t)$. Due to partial observability, the agent can only form a belief $\mathbf{b}_t$ about its condition state, where $\mathbf{b}_t$ is a probability distribution over S. The goal for the agent is to choose actions at each time step that minimize its expected future discounted cumulative cost (Papakonstantinou & Shinozuka, 2014a; Papakonstantinou & Shinozuka, 2014b). Despite existing mathematical convergence guarantees, the traditional point-based POMDP solvers do not scale adequately for large systems with multiple components. However, neural network-based deep reinforcement learning methods can alleviate this curse of dimensionality.

2.2 *Deep reinforcement learning and DDMAC-CTDE*

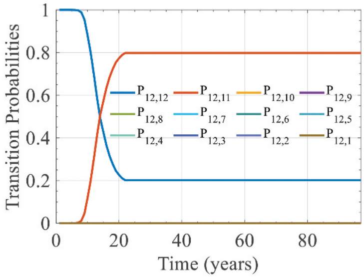
Within the context of infrastructure management, DDMAC, as developed in (Andriotis & Papakonstantinou, 2021), provides an algorithm for system-level I&M optimal planning well-suited for large multi-component systems. The framework also considers the presence of constraints through state augmentation and Lagrange multipliers. DDMAC uses a sparse parametrization of the actor-network without parameter sharing between agents (i.e., each component has its own actor network). For even larger systems, the DDMAC-CTDE formulation (Saifullah, et al., 2024) can be used, allowing for even sparser actor parametrizations. The policy, π, and its gradient are then given as (Saifullah, et al., 2024):

$$\pi(\mathbf{a}_t|\mathbf{s}_t) = \prod_{i=1}^{n} \pi_i\left(a_t^{(i)}|s_t^{(i)}\right) \tag{1}$$

$$\mathbf{g}_{\boldsymbol{\theta}^\pi} = \mathrm{E}_{\mathbf{s}_t\sim\mathbf{p},\mathbf{a}_t\sim\mu}\left[w_t\left(\sum_{i=1}^{n} \nabla_{\boldsymbol{\theta}^\pi} \log \pi_i\left(a_t^{(i)}\,|\,s_t^{(i)},\boldsymbol{\theta}^\pi\right)\right)A^\pi(\mathbf{s}_t,\mathbf{a}_t)\right] \tag{2}$$



(a)



(b)

Figure 1. (a) Fitted gamma model for GCR. (b) Transition probabilities for multi-span bridges belonging to PennDOT Family 1, with starting state 12.

3 ENVIRONMENT DESCRIPTION

3.1 *Condition states and transitions*

In the case of the Pennsylvania network, the POMDP environment is derived based on PennDOT's Bridge Asset Management System (BAMS) (PennDOT, 2022; PennDOT, 2021), focusing only on bridge decks here as the primary deteriorating constituent of bridge assets. The network considered for this study consists of 30 Family 1 multi-span bridge decks in District 8, with General Condition Rating (GCR) as the indicator of bridge deck conditions, ranging from 0 to 9.99. A probabilistic non-stationary GCR model is developed based on the deterministic deterioration model, as employed by PennDOT, as detailed in (Papakonstantinou, et al., 2024, under prep.). A non-stationary gamma process is utilized, with its mean being in time equal to the PennDOT model GCR values, and a suitable model variance. Figure 1a shows some representative simulation results for multi-span bridges belonging to PennDOT Family 1. The solid line represents the mean GCR and the red curve is the GCR model used by PennDOT.

To calculate the transition probabilities, the continuous GCR values are discretized into 15 states, with state 15 being GCR ≥9.5, state 14 being GCR $\in$ (9.5, 9.0], and so on until state 1, which is a compound state for all the GCR values < 3.0, to accommodate PennDOT's prescribed maintenance recommendations. Fifty million Monte Carlo sequences from the developed gamma process are simulated to obtain the transition probabilities for a given bridge family and span type, appropriately smoothed over a time window of 5 years. Figure 1(b) indicatively shows some computed transition probabilities for the Family 1 multi-span bridges.

In the case of the Virginia network, the Hampton Roads network is considered. It comprises 85 pavement (including interstate, primary, and secondary highways) and 11 bridge components, with 6 Critical Condition Index (CCI) states and 5 International Roughness Index (IRI) states used as condition indicators for pavements, and 7 GCR-based bridge states (Saifullah, et al., 2022). A non-stationary CCI model is devised based on VDOT's model (Katicha, et al., 2016). More details on the condition state transition probabilities can be found in (Saifullah, et al., 2024).

3.2 *Maintenance and inspections*

In the Pennsylvania network, bridge deck maintenance follows PennDOT's BAMS specifications, involving five specific maintenance actions: *Do-Nothing*, *General Preservation, Epoxy Overlay, Structural Overlay*, and *Deck Replacement*, along with their respective costs (PennDOT, 2022). All considered actions generally impact both the bridge deck condition state and deterioration rate. Detailed descriptions of these actions are provided in (Papakonstantinou, et al., 2024, under prep.). In this work, a high-fidelity inspection with 90% accuracy is assumed, performed every two years, to adhere to PennDOT's bridge assessment policies.

For the Virginia network, based on (VDOT, 2016), four different maintenance actions are considered for pavements, i.e., *Do-Nothing*, *Minor Repair*, *Major Repair*, and *Reconstruction*. Minor Repair (crack filling, moderate patching, etc.) can improve the CCI and IRI condition states but does not affect their rate of deterioration. Major Repair can improve condition states and also reduce their deterioration rate by 5 years. Reconstruction resets the pavement to an intact condition. A detailed description of maintenance actions and their costs can be found in (VDOT, 2016). Similarly, four maintenance actions are also considered for bridge decks. These actions, the related transition probabilities, action durations and costs are described in detail in (Saifullah, et al., 2024). Towards generality, inspection actions are also optimized in this case and three inspection options are considered, characterized as uninformative, low-fidelity, and high-fidelity inspection techniques, respectively. The observation probabilities for the corresponding inspections can also be seen in (Saifullah, et al., 2024).

3.3 *Costs, risks, and constraints*

For the Pennsylvania bridge network, the total cost is composed of the costs associated with maintenance actions and risk. The risk cost depends on the current condition rating along with the risk score (*RS*) associated with a bridge, as reported in BAMS. The risk score describes the importance of a bridge, which is a function of deck area, annual average daily traffic, truck traffic percent, detour length, scour rating, fracture criticality, and history of flooding (PennDOT, 2022). The risk score-based cost, $c_{RS}(t)$ at time t is quantified here as:

$$c_{RS}(t) = \sum_{i=1}^{N} \left(10 - GCR(t)_i\right) RS_i \tag{3}$$

where $GCR(t)_i$ is the GCR rating at time step t for the i^{th} bridge, and N is the total number of bridges considered. Inspections are not controlled in this case but are pre-assigned every two years. Their costs are thus not considered in the total cost to be optimized. Therefore, the total cost at time t can be estimated as:

$$c(s_t, a_t) = \underbrace{c_M(s_t, a_t)}_{\text{maintenance cost}} + \underbrace{c_{RS}(s_t)}_{\text{risk score cost}} \tag{4}$$

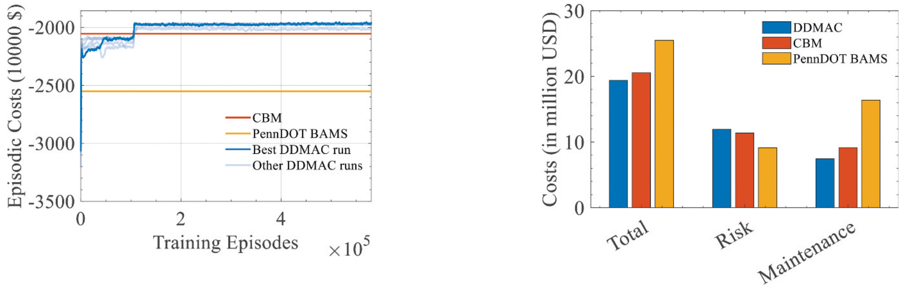
Table 1. Comparison of different solution schemes in terms of average (+/- 95% confidence bounds) total cost and performance for the Pennsylvania network.

Objective & Constraints	DDMAC-CTDE	CBM policy	BAMS policy
Total costs (million USD)	19.3779 (± 0.0211)	20.5435 (**+6.02%**) (± 0.0215)	25.5061 (**+32%**) (± 0.0116)
Poor Interstate (cap 5%)	0.5329 (± 0.0790)	0.0245 (± 0.0162)	0.00 (± 0.00)
Poor NHS (cap 7%)	3.7489 (± 0.1356)	6.4255 (± 0.1923)	0.3237 (± 0.0648)
Poor total (cap 10%)	8.3886 (± 0.2325)	4.6769 (± 0.1296)	12.3157 (± 0.1293)

Table 2. Comparison of different solution schemes in terms of average (+/- 95% confidence bounds) total cost and performance for the Virginia network.

Objective & Constraints	DDMAC-CTDE	CBM policy	VDOT policy
Total costs (billion USD)	6.55 (±0.003)	7.03 (**+7.3%**) (±0.004)	8.35 (**+27%**) (±0.008)
CCI<60 and IRI>2.2m/km for I-Hwy (%, cap 5%)	1.34 (±0.02)	1.61 (±0.02)	3.39 (±0.07)
CCI<35 of I-Hwy (%, cap 2%)	1.50 (±0.02)	0.68 (±0.01)	3.80 (±0.07)
CCI<60 for I and P-Hwy (%, cap 18%)	17.65 (±0.04)	12.67 (±0.02)	5.80 (±0.04)
IRI>2.2 m/km for I and P-Hwy (%, cap 15%)	15.40 (±0.04)	11.49 (±0.04)	15.65 (±0.07)
CCI<60 for S-Hwy (%, cap 35%)	33.00 (±0.08)	28.18 (±0.04)	37.86 (±0.13)
Bridges with GCR ≤4 (%, cap 10%)	8.33 (±0.05)	8.79 (±0.06)	15.85 (±0.15)

The Pennsylvania network in this study consists of 30 multi-span District 8 bridges belonging to Family 1, including 3 Interstate NHS and 3 Non-Interstate NHS bridges. The constraints under consideration include a hard budget limit of 5.30 million, renewed every 5 years, and 3 performance targets: (1) for Interstate bridges the average deck area in poor condition (GCR ≤ 5.0) should be less than 5% over the decision horizon; (2) for NHS bridges this target is 7%; and (3) across the entire network the average proportion of bridges in poor condition state should not exceed 10%. These constraints are integrated into the DRL framework as outlined in (Andriotis & Papakonstantinou, 2021; Saifullah, et al., 2024).



(a) (b)

Figure 2. (a) Total life cycle costs comparison of DDMAC-CTDE solution with CBM and PennDOT policy baselines. (b) Comparison of the total cost and its constituents with CBM and PennDOT policy baselines.

For the Virginia network, the total cost includes inspection, maintenance, traffic delay, terminal costs, and bridge failure risks. Budget constraints with a 5-yearly budget of 1.3 billion USD, and six performance targets are used, with the latter sought to be satisfied in an average sense over the decision horizon, based on FHWA and VDOT guidelines (Saifullah, et al., 2024).

4 RESULTS

The networks have been trained using the DDMAC-CTDE architecture, with their existing conditions as the initial condition states, an episodic length of 20 years, and considering a discount factor of $\gamma = 0.97$. 5 maintenance actions per component are considered for the Pennsylvania network, totaling ~10^{21} available actions for the entire network, and 10 inspection and maintenance actions per component are considered for the Virginia network, with 10^{96} total available actions for the system. The 5-yearly budgets used for both the Pennsylvania and the Virginia networks follow the guidelines by the respective agencies.

To compare the DDMAC-CTDE solutions, two types of baselines are evaluated, i.e., (i) a condition-based maintenance (CBM) policy, and (ii) policies following PennDOT/BAMS and VDOT practices. The CBM policy is heuristically optimized, to identify the relevant thresholds, based on the conditions of each component type. The details of CBM, BAMS, and VDOT policies can be found in (Papakonstantinou, et al., 2024, under prep.; Saifullah, et al., 2024).

Figure 2(a) demonstrates DDMAC-CTDE performance over CBM and BAMS during its training on the Pennsylvania network, with weights initialization in this case based on an unlimited budget earlier trained case. Figure 2(b) showcases relevant histograms of total, maintenance, and risk costs from 5,000 Monte Carlo policy realizations. DDMAC-CTDE policy is 6.02% and 31.64% less expensive in total cost, in relation to the CBM and BAMS policies, respectively, and it spends 120% less than BAMS and 23% less than CBM on maintenance. Table 1 presents further details on the average performance of the Pennsylvania network over 5,000 Monte Carlo simulations, showcasing the total cost and the three constraints related to poor bridge conditions percentages, together with their recommended targets. Table 2 similarly compares six performance constraints and their targets for various pavement types and bridges in the Virginia network, where I-, P-, and S-Hwy represent interstate, primary, and secondary highway components, respectively. Both Tables provide the estimated values along with their 95% confidence bounds.

To better understand the converged DDMAC-CTDE policies, Figures 3 and 4 illustrate instances of the Pennsylvania and Virginia networks, respectively. In Figure 3, the maintenance policies for five randomly selected bridges in PA are shown over time, together with the

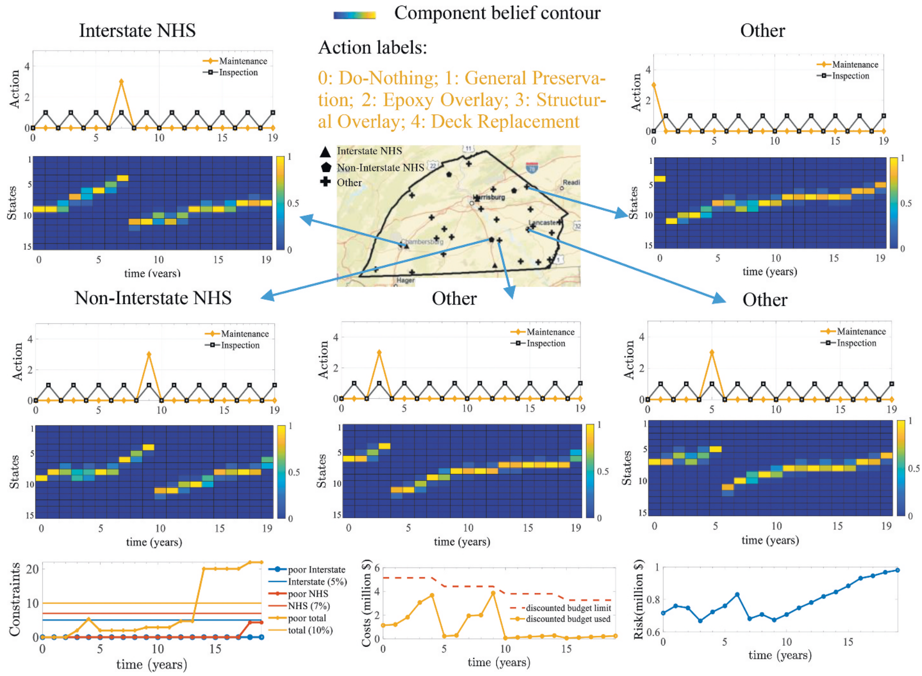


Figure 3. Indicative realization of the learned DDMAC-CTDE policy for the PA transportation network. Due to safety considerations by PennDOT, the true identities of the individual bridges are withheld. The visualization here serves solely as an illustrative depiction of the network under consideration and does not correspond to the actual bridges analyzed in this study.

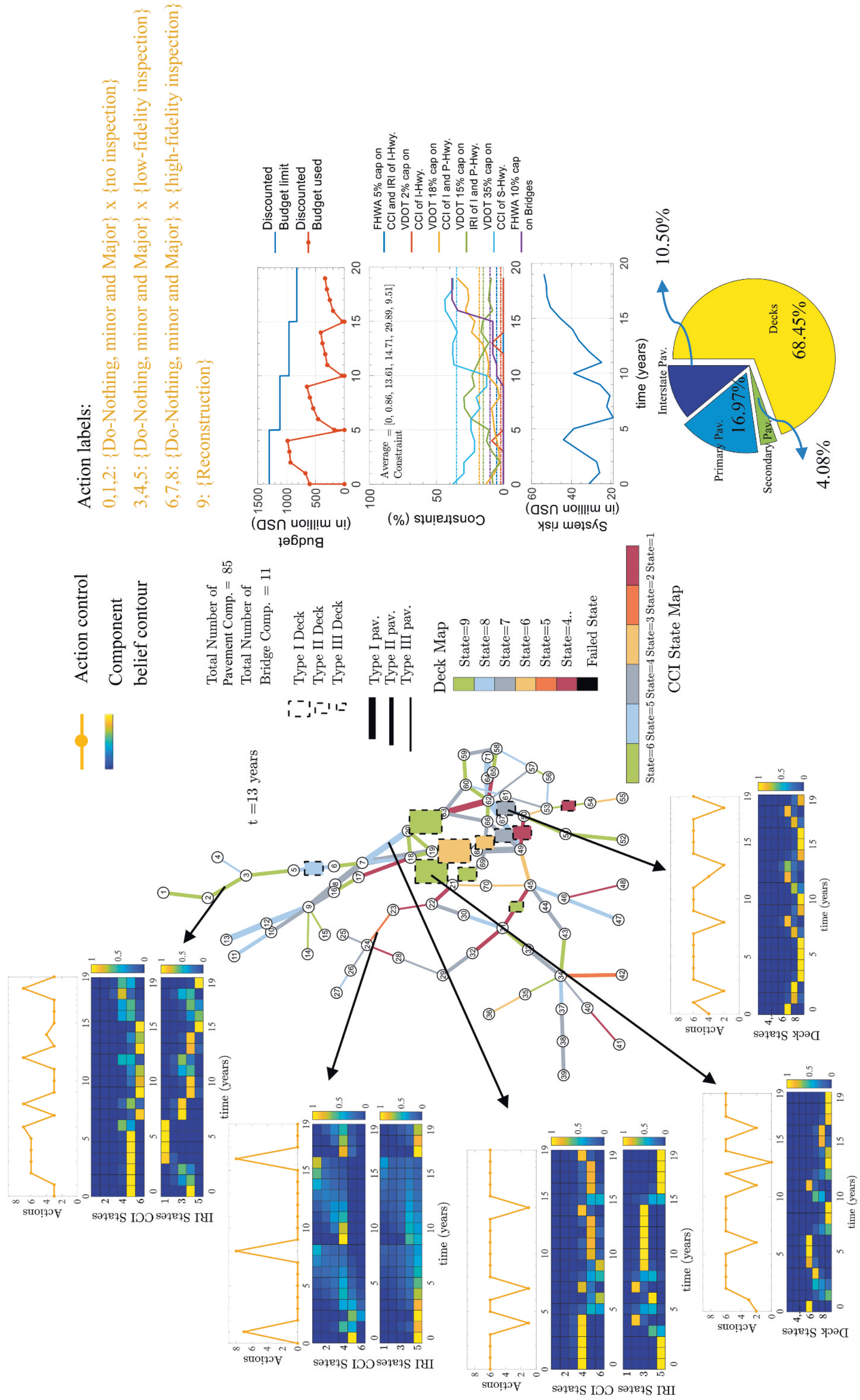


Figure 4. Illustrative realization of the learned DDMAC-CTDE policy for the Hampton Roads network.

action effects on the belief states. Figure 3 also depicts variations in considered performance constraints, their targets, the total system risk cost, and the utilized budget, with the total available budget discounted for each five-year cycle. The figure indicates that performance constraints are well-respected, on average over the decision period, and the budget remains within limits throughout the entire horizon.

Similarly, Figure 4 represents a policy instance for the Virginia network, where there are 10 available control actions. The time variations of budget utilization, system risk, and six performance constraints are also shown. The budget is always under its limit in this case as well, and discounted for each five-year cycle. A pie chart representing the I&M cost distributions for different asset types is additionally shown, with most funds spent on bridges in this case, having also a failure risk associated with them. Among the pavements, primary highway pavements utilize the most funds, as they represent the most components in this network.

5 CONCLUSIONS

Deep Decentralized Multi-Agent Actor Critic with Centralized Training and Decentralized Execution (DDMAC-CTDE) solutions are evaluated in this work, for managing bridge and bridge-pavement systems, related to two real-world networks in Pennsylvania and Virginia, USA. Comparisons with the corresponding agency policies, i.e., PennDOT (BAMS) and VDOT, respectively, and with optimized Condition-Based Maintenance (CBM) policies are performed as well. The DDMAC-CTDE solution is shown to surpass the State DOT-based baselines by 32% (for BAMS) and 27% (for VDOT) in terms of the total overall cost. CBM policies for the two networks are also surpassed by 6.02% and 7.3%, respectively, while generally meeting all the specified performance targets. Overall, this work and results indicate the applicability and effectiveness of our DRL framework for asset management of real-world networks, characterized in their full complexity, and can set solid foundations for further developments and implementations in the field, related to bridge networks and beyond.

ACKNOWLEDGEMENTS

We acknowledge the support by the U.S. National Science Foundation under CAREER Grant No. 1751941, by the Pennsylvania Department of Transportation, and by the Center for Integrated Asset Management for Multimodal Transportation Infrastructure Systems, 2018 U.S. DOT Region 3 University Center. We would further like to acknowledge the support by PennDOT's Mr. Jeff Davis and Mr. Justin Bruner, through extensive discussions, data sharing, and BAMS access.

REFERENCES

Andriotis, C. P. & Papakonstantinou, K. G., 2019. Managing engineering systems with large state and action spaces through deep reinforcement learning. *Reliability Engineering & System Safety*, Volume 191, p. 106483.

Andriotis, C. P. & Papakonstantinou, K. G., 2021. Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. *Reliability Engineering & System Safety*, Volume 212, p. 107551.

Papakonstantinou, K.G., Bhattacharya, A., Saifullah, M., Guler, I., Lu, M., Sengupta, A., Gayah, V., 2024, in preparation. *Quantifying the impact of data unavailability, inaccuracies and uncertainty on deterioration modeling and infrastructure asset management policies*. CIAMTIS US DOT Region 3 University Transportation Center, Final Report.

Frangopol, D., Lin, K. & Estes, A., 1997. Life-cycle cost design of deteriorating structures. *Journal of Structural Engineering*, Volume 123(10), p. 1390–401.

Katicha, S. W. et al., 2016. VDOT: *Development of enhanced pavement deterioration curves*.

Luque, J. & Straub, D., 2019. Risk-based optimal inspection strategies for structural systems using dynamic Bayesian networks. *Structural Safety*, Volume 76, pp. 68–80.

Lyu, X., Xiao, Y., Daley, B. & Amato, C., 2021. Contrasting centralized and decentralized critics in multi-agent reinforcement learning. *arXiv preprint arXiv:2102.04402.*

Madanat, S., 1993. Optimal infrastructure management decisions under uncertainty. *Transportation Research Part C: Emerging Technologies*, Volume 1(1), pp. 77–88.

Papakonstantinou, K. G. & Shinozuka, M., 2014a. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory. *Reliability Engineering & System Safety*, Volume 130, pp. 202–213.

Papakonstantinou, K. G. & Shinozuka, M., 2014b. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation. *Reliability Engineering & System Safety*, Volume 130, pp. 214–224.

Papakonstantinou, K. G. & Shinozuka, M., 2014. Optimum inspection and maintenance policies for corroded structures using partially observable Markov decision processes and stochastic, physically based models. *Probabilistic Engineering Mechanics*, Volume 37, pp. 93–108.

PennDOT, 2021. *Bridge Care*. [Online] Available at: https://bams.penndot.gov/ [Accessed 4th January 2024].

PennDOT, 2022. *Transportation Asset Management Plan*, Harrisburg, PA: Pennsylvania Department of Transportation.

Saifullah, M., Andriotis, C. P., Papakonstantinou, K. G. & Stoffels, S. M., 2022. *Deep reinforcement learning-based life-cycle management of deteriorating transportation systems*. Proceedings of the 11th International Conference on Bridge Maintenance, Safety and Management (IABMAS), CRC press, p. 293–301.

Saifullah, M., Papakonstantinou, K. G., Andriotis, C. P. & Stoffels, S. M., 2024. Multi-agent deep reinforcement learning with centralized training and decentralized execution for transportation infrastructure management. *arXiv preprint arXiv:2401.12455.*

VDOT, 2016. *Supporting document for the development and enhancement of the pavement maintenance decision matrices used in the needs-based analysis*, Richmond: Virginia Department of Transportation.