



Inference and dynamic decision-making for deteriorating systems with probabilistic dependencies through Bayesian networks and deep reinforcement learning

P.G. Morato ^{a,*}, C.P. Andriotis ^b, K.G. Papakonstantinou ^c, P. Rigo ^a

^a ANAST, Department of ArGENCo, University of Liege, 4000, Liege, Belgium

^b Faculty of Architecture and the Built Environment, Delft University of Technology, 2628 BL Delft, The Netherlands

^c Department of Civil & Environmental Engineering, The Pennsylvania State University, University Park, PA 16802, USA



ARTICLE INFO

Keywords:

Infrastructure management
Decision analysis
Deep reinforcement learning
Partially observable Markov decision processes
System reliability analysis
Dynamic Bayesian networks

ABSTRACT

In the context of modern engineering, environmental, and societal concerns, there is an increasing demand for methods able to identify rational management strategies for civil engineering systems, minimizing structural failure risks while optimally planning inspection and maintenance (I&M) processes. Most available methods simplify the I&M decision problem to the component level, often assuming statistical, structural, or cost independence among components, due to the computational complexity associated with global optimization methodologies under joint system-level state descriptions. In this paper, we propose an efficient algorithmic framework for inference and decision-making under uncertainty for engineering systems exposed to deteriorating environments, providing optimal management strategies directly at the system level. In our approach, the decision problem is formulated as a factored partially observable Markov decision process, whose dynamics are encoded in Bayesian network conditional structures. The methodology can handle environments under equal or general, unequal deterioration correlations among components, through Gaussian hierarchical structures and dynamic Bayesian networks, decoupling the originally joint system state space to component networks conditional on shared random variables. In terms of policy optimization, we adopt a deep decentralized multi-agent actor-critic (DDMAC) reinforcement learning approach, in which the policies are approximated by actor neural networks guided by a critic network. By including deterioration dependence in the simulated environment, and by formulating the cost model at the system level, DDMAC policies intrinsically consider the underlying system-effects. This is demonstrated through numerical experiments conducted for both a 9-out-of-10 system and a steel frame under fatigue deterioration. Results demonstrate that DDMAC policies offer substantial benefits when compared to state-of-the-art heuristic approaches. The inherent consideration of system-effects by DDMAC strategies is also interpreted based on the learned policies.

1. Introduction

Managing engineering systems, by controlling the risks of adverse events and optimally allocating inspection and repair resources, is crucial for securing societal progress, improving the quality of life at the community level and maximizing economic returns from an individual perspective [1]. Increasing societal consciousness on sustainability, along with the expanding wealth of data from our structural systems and infrastructure, require, and enable, more efficient management policies [2]. Engineering systems are exposed to deterioration mechanisms during the course of their service life, that negatively influence their expected reliability and operational performance. The prediction

of deterioration evolution and system reliability estimates from engineering models is associated with significant uncertainties. At a certain economic expense, information can be collected from inspection and/or monitoring systems and can be used to update reliability metrics in time, hence supporting more rational and optimal maintenance decisions. Inspection, maintenance, and repair interventions should be thus timely allocated throughout the lifetime of engineering systems based on reliability, cost and other related metrics. Research efforts devoted to the development of risk-based inspection and maintenance planning methods have increased considerably during the last decade [2]. Such policies need to support decision-making, for both newly designed

* Corresponding author.

E-mail address: pgmorato@uliege.be (P.G. Morato).

systems and existing ones, through life-cycle plans that integrally account for interventions (e.g., repairs, retrofits, etc.) and data collection methods (e.g., inspections, structural health monitoring, etc.).

Most available inspection and maintenance (I&M) planning methods assume independence among the constitutive components, primarily driven by the practical need to tame the involved computational complexities associated with solving such a system-level decision-making optimization problem under uncertainty [3,4]. At the component level, existing risk-based I&M methods can be classified according to their capabilities of modeling physically-based deterioration processes, e.g., fatigue or corrosion deterioration [5,6], and depending on their policy optimization approaches, namely, static decision rules, adaptive decision rules prescribed by heuristics, or adaptive decision rules defined as a function of the dynamically updated history of actions and observations. Some methods focus on the optimization of predefined static decision rules, planning inspections at equidistant intervals or when a prescribed failure probability threshold is surpassed, and prescribing maintenance interventions if a certain damage indicator is observed, e.g., crack detection [3,7,8]. While these approaches can provide reasonable and effective policies in some specific scenarios, the optimality of the policies depends greatly on the designer's experience when defining the heuristic combinations for the policy search, since they cannot consider all policies within the vast available policy space, which could in turn produce better results than the originally considered predefined heuristics [9,10]. In other existing methods, while inspection planning decision rules are defined a priori, the maintenance policy is adaptive, properly updating the involved thresholds based on new information [11]. In these cases, action planning is formulated based on optimization techniques which need to be repeated for all desired updates. While still operating on a limited policy subspace, such sophisticated methods correctly identify the need to go beyond static thresholds and accordingly provide solution approaches.

Methods based on Markov Decision Processes (MDPs) and Partially Observable MDPs (POMDPs) aim, on the other hand, at addressing the problem in a global optimization sense, outside the limitations of threshold-based or time-based formulations. Early works on the application of Markov decision processes for managing deteriorating engineering cases include [12–15]. Founded on the principled mathematical properties offered by dynamic programming, either under full or partial observability, additional formulations have been proposed, e.g., in [16–18]. In the same class of applications, a recent POMDP-based approach proposed in [9] demonstrated that POMDP policies outperform heuristic-based policies, as exemplified in physically-based numerical examples featuring fatigue deterioration. POMDP policies are defined as a function of the belief about the condition states, i.e., the probability distribution over states, which is a sufficient statistic of the prior history of actions and observations, recursively encoding it through forward Bayesian updates.

As mentioned before, many of the existing I&M methods formulate the decision-making problem at the component level. However, disregarding the essential interrelations among the system constituents, although allowing for a substantial simplification of the decision-making problem, may result in sub-optimal and even non-conservative policies in some cases. The need for I&M methods capable of determining policies at the system level has long been identified by the risk research community. Early works approaching the problem at the system level include [19,20]. In [21], the fatigue details were classified according to the fatigue design factor, establishing a simplified approach for identifying system policies. More recently, [22] proposed a static I&M planning optimization relying on dynamic Bayesian networks to efficiently model deterioration, cost and reliability dependence among the structural elements. In this method, the policy is computed by optimizing static heuristic decision rules, with decision variables including, among others, equidistant inspections, number of inspected components, component prioritization, and repair thresholds based

on observations. As with all static policy optimization methods, explained before, the policies are constrained to the set of predefined heuristic rules, out of the immense space of possible policies, which is substantially enlarged now in structural system settings.

Addressing the important complexities of managing large engineering systems, a deep reinforcement learning (DRL) method has been introduced in [23], motivated by the success of deep reinforcement learning algorithms in complex game environments, e.g., in [24–26]. In particular, a multi-agent actor-critic DRL scheme is developed in [23], relying on (PO)MDPs for simulating the deteriorating environment, and demonstrating the capabilities of deep reinforcement learning approaches for identifying optimal policies in high-dimensional state, action, and observation spaces. Thereafter, a modified version of this method has also been applied for solving system I&M decision-making problem under constraints, e.g., imposed risk thresholds or budget limitations [27]. In general, DRL approaches offer computational benefits in high-dimensional state spaces, mitigating the need for exhaustive state exploration by leveraging a function parametrization over the state space [28]. Further verifying the effectiveness of DRL methods for minimizing long-term costs and system failure risks, DRL approaches have been recently applied for the maintenance optimization of multi-component engineering systems [29,30], from gas plant and pipeline systems [31,32] to rail-road infrastructure [33], and bridge assets [34,35].

In order to identify optimal management strategies for engineering systems directly at the system level, we originally formalize here an efficient modeling framework for inference and decision-making under uncertainty for general cases with probabilistic dependencies. Inference and reliability estimation for systems managed through maintenance actions and inspections is hard when there exist statistical correlations among components. Addressing this, we provide an approach in this work that can also be consistently combined with a stochastic control framework through deep reinforcement learning. In terms of inference, our proposed methodology builds on top of adept Dynamic Bayesian Network (DBN) formulations [4,36], modeling environments described by deterioration dependencies among components through Gaussian hierarchical structures, with the objective of decoupling the joint system space to independent component networks conditional on common influencing random variables. This decomposition results in a linear computational complexity with the number of components that otherwise increases exponentially in the joint system space. Furthermore, in this paper, the Gaussian hierarchical model is originally expanded and enhanced to enable the treatment of general, unequal deterioration correlation scenarios and dependence alterations after a maintenance action is taken for a subset of the components.

While our developed generalized inference framework is applicable regardless of the decision-making method used for the generation of management policies, our proposed approach efficiently models the decision-making problem as a factored POMDP, whose dynamics are encoded as Bayesian networks that can then be seamlessly integrated with sophisticated solution methods. In this regard, we adopt a deep decentralized multi-agent actor-critic (DDMAC) scheme, in which the system policies are approximated by actor neural networks, at a component level, guided via system level value function estimates approximated by a critic network [23]. As DDMAC adjusts the weights of the actor networks according to noisy rewards collected at the system level, DDMAC policies intrinsically consider system-effects stemming from structural and statistical dependencies. Through numerical experiments, we demonstrate the efficacy of the proposed method for I&M planning of structural systems exposed to fatigue deterioration. DDMAC policies are compared against state-of-the-art optimized heuristic policies in all analyzed experiments. In particular, the effects of including deterioration dependence and campaign cost models are explored for the case of a 9-out-of-10 system. In the second application studied, of a steel frame structural system featuring 22 fatigue hotspots, the focus is on examining and interpreting the inherent allocation of maintenance

interventions by DDMAC policies according to the element importance to the global structural reliability. Through the results and discussion reported in this work, relevant insights for typical structural fatigue applications can be drawn.

The remainder of the paper is structured as follows: an overview of POMDP methods along with the proposed factored formulation are presented in Section 2. In Section 3, the definition and modeling of Gaussian hierarchical structures are introduced, together with a belief update algorithm, applicable to environments under general deterioration dependence. The integration of the simulator, defined as a factored POMDP, with DDMAC is presented in Section 4. The numerical experiments are then introduced and discussed in Section 5, concluding with some final remarks in Section 6.

2. I&M decision problem formulated as a factored POMDP

2.1. Factored POMDP definition

The inspection and maintenance (I&M) planning decision-making problem is formulated here as a Partially Observable Markov Decision Process (POMDP), whose transition and observation models are defined by Bayesian network structures. POMDPs provide a principled mathematical framework for optimal planning and decision-making under uncertainty, formally specified by the tuple $\langle S, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{Z}, \mathcal{R}, \gamma \rangle$. A decision maker (henceforth agent) interacts with a stochastic environment, described by the state $s \in S$, taking actions $a \in \mathcal{A}$ over a finite or infinite horizon t_N . The dynamics correspond to those in a MDP: At each time step t , an agent takes an action $a_t \in \mathcal{A}$, and the environment evolves from state $s_t \in S$ to state $s_{t+1} \in S$, according to the transition model $\mathcal{T} := p(s_{t+1}|s_t, a_t)$. In a MDP, the agent receives a reward based on the cost model $\mathcal{R} := r_t(s_t, a_t, s_{t+1})$ discounted by factor γ , and the objective is to find the policy π^* that induces the optimal value function $V^*(s_t)$:

$$V^*(s_t) = \max_{a_t \in \mathcal{A}} \left\{ r(s_t, a_t) + \gamma \sum_{s_{t+1} \in S} p(s_{t+1}|s_t, a_t) V^*(s_{t+1}) \right\} \quad (1)$$

In a POMDP, however, states $s \in S$ are not directly observed and, instead, observations $o \in \mathcal{O}$ can be collected according to the observation model $\mathcal{Z} := p(o_{t+1}|s_{t+1}, a_t)$. Note that the observation model is the likelihood of collecting an observation $o_{t+1} \in \mathcal{O}$ after taking an action a_t and having transitioned to state s_{t+1} . In an I&M context, the observation model is often modeled by Probability of Detection (PoD) curves or according to inspection/monitoring measurement noise [9]. A POMDP policy is a mapping of the dynamically updated history of actions and observations, $a_{0:t-1}, o_{0:t}$, to the current action a_t . This history is sufficiently encoded in belief \mathbf{b} , which is the probability over system states, $b(s_t)$. The optimal policy π^* , therefore, corresponds to the value function [37] that satisfies the Bellman equation:

$$V^*(\mathbf{b}_t) = \max_{a_t \in \mathcal{A}} \left\{ \sum_{s_t \in S} r(s_t, a_t) b(s_t) + \gamma \sum_{o_{t+1} \in \mathcal{O}} p(o_{t+1}|\mathbf{b}_t, a_t) V^*(\mathbf{b}_{t+1}) \right\}, \quad (2)$$

where $p(o_{t+1}|\mathbf{b}_t, a_t)$ is the probability of collecting an observation $o_{t+1} \in \mathcal{O}$ given the belief \mathbf{b}_t and action $a_t \in \mathcal{A}$. Assuming a Markovian environment is reasonable in most practical applications with the aid of state augmentation techniques [13], any general I&M planning decision problem can be efficiently formulated as a POMDP. The determination of the optimal I&M policy π^* becomes the main objective, inducing a minimization of the expected life-cycle costs r_{tot} , by balancing structural failure risk against inspection and maintenance costs:

$$\mathbb{E}[r_{tot}] = \sum_{t=0}^{t_N} \mathbb{E}[\gamma^t (r_{ins,t} + r_{rep,t} + r_{F,t})], \quad (3)$$

where r_{ins} , r_{rep} and r_F stand for inspection, repair and failure costs, respectively, defined as negative rewards. In terms of utilities, the failure risk r_F is typically defined in a structural reliability context as the

annual probability of a failure event weighted by the consequence of a structural failure, which might also include environmental and societal consequences, specified in equivalent units. The definition of the failure risk at the system level will be further elaborated in Section 3.

Existing I&M planning applications often model the deterioration evolution d , at the component level, conditional on a set of random variables Θ_d [4,22,36] or as a function of the deterioration rate τ [14,23]. Both formulations are equivalent for modeling deterioration processes, as already discussed and demonstrated in [9] and shown in Fig. 1. When observations are collected, through inspections or monitoring, Bayesian updating can be then conducted. Available algorithms allow exact Bayesian inference if the problem is formulated in a discrete state space [38], as the computation of Bayes' normalization constant is a challenging task in continuous state spaces [39]. In order to utilize discrete state based algorithms, the involved continuous random variables can be discretized. The quality of the discretization has a huge impact and shall be treated carefully [4,9], especially when the problem deals with rare events, e.g., failure events.

In a POMDP, the states cannot be directly observed and the decision maker reasons under partial observability, only informed by a belief \mathbf{b} , which is defined as the probability over states. At each time step, the belief is dynamically updated, based on Bayes' rule, depending on the initial belief, \mathbf{b}_t , the action taken, a_t , and the collected observation, o_{t+1} , following three main steps: (i) the belief evolves according to the transition model $p(s_{t+1}|s_t, a_t)$, (ii) the belief is updated based on the collected observation with probability $p(o_{t+1}|s_{t+1}, a_t)$, and (iii) the belief state is normalized. This belief update operation is denoted as forward pass within the context of hidden Markov models [38]. At the system level, the belief of each component can be updated by implementing the steps listed in Algorithm 1.

State-of-the-art POMDP solvers often require the modeling of the POMDPs in a flat structure, which can be usually encoded by augmenting the state space [13], particularly if the process is described by multiple random variables. However, POMDPs can also be formulated in a factored fashion, exploiting the dependence structure among random variables and thus significantly alleviating the required computational effort. We specify here the transition and observation models based on conditional structures described by dynamic Bayesian networks (DBNs), and while the belief state \mathbf{b} remains the same as that for flat POMDPs, the transition and observation models are now constructed by taking advantage of the involved dependencies. For instance, the deterioration rate model can be constructed as $p(d_{t+1}|d_t, \tau_{t+1}) p(\tau_{t+1}|\tau_t)$ instead of $p(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$, where $s = d_t, \tau_t$.

Algorithm 1 Belief update for a system of N_c uncorrelated components

```

function UPDATEBELIEF( $\mathbf{b}_t, a_t, o_{t+1}$ )
  for 1,  $N_c$  do
     $b(s_{t+1}) \leftarrow b(s_t) p(s_{t+1}|s_t, a_t)$                                 ▷ propagation step
     $b(s_{t+1}) \leftarrow b(s_{t+1}) p(o_{t+1}|s_{t+1}, a_t)$                             ▷ estimation step
     $b(s_{t+1}) \leftarrow b(s_{t+1}) / p(o_{t+1}|\mathbf{b}_t, a_t)$                             ▷ normalization step
  end for
end function

```

This incorporation of conditional structures allows a reduction of the transition model dimensionality from $|S_d|^2 |S_r|^2$ to $|S_d|^2 |S_r| + |S_r|^2$ and can achieve significant computational benefits when multiple random variables are involved. This formulation can be seamlessly applied to simulate the deterioration environment, as will be explained in Section 4, due to the flexibility naturally offered by the proposed deep reinforcement learning approach.

3. System effects in I&M planning

3.1. Deterioration dependence in a hierarchical Gaussian structure

Existing methods model the deterioration correlation among components either via random fields or through common influencing factors.

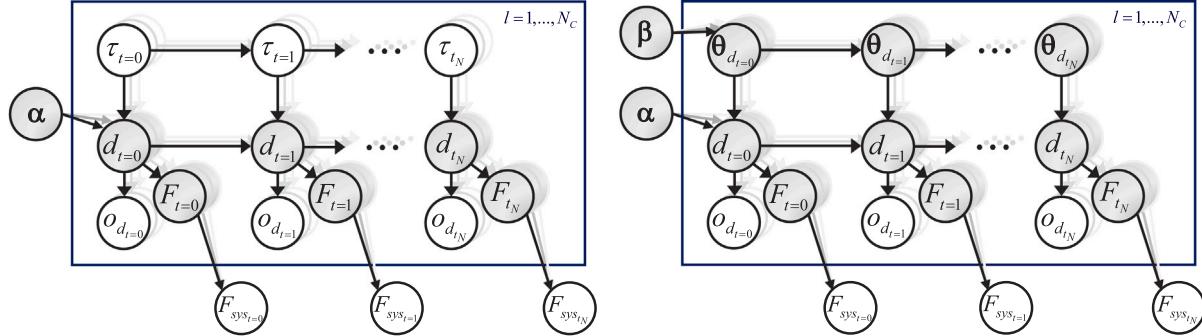


Fig. 1. Graphical representation of dynamic Bayesian networks for modeling deterioration processes. At the component level, damage d_t evolves over time t as a function of the deterioration rate τ_t (left) or conditional on a set of parameters θ_{d_t} (right). While d_t and θ_{d_t} are hidden states, partially observed through o_{d_t} , τ_t is fully observable. The deterioration dependence among components is encoded by the hyperparameter or set of hyperparameters α, β . The binary failure observable state indicates either the survival or failure state of the system F_{sys_t} , depending on the components failure hidden state F_t . Note that other variants of the system failure formulation can also be represented accordingly.

Whereas the former are particularly useful for applications in which the dependence is attributed to the geometrical distance between components, the latter are more suitable for systems in which identical attributes of physical phenomena, e.g., similar manufacturing techniques or similar loading, lead to shared sources of model uncertainties among the components [36]. In a hierarchical structure, the deterioration of each component is defined conditional on a set of common influencing variables, shared among all the components and represented at the highest level of the hierarchy. In theory, the state space of a system under deterioration dependence can be modeled directly as the joint space of all the parameters involved in the deterioration process of the system. In this case, the discretized state space would grow exponentially with the number of components N_C included, into a $|S|^{N_C}$ dimensional space. To overcome this increase in dimensionality, we adopt the hierarchical Gaussian structure previously proposed in [36], in which the belief state of each component is encoded conditional on a hyperparameter (common influencing variable) α or set of hyperparameters α . The central idea behind this hierarchical structure is that component beliefs for a given hyperparameter $b(s|\alpha)$ are independent, enabling an efficient decoupling of the components joint space. This decoupling alleviates the computational complexity from the original joint space $|S|^{N_C}$ to a space $|S| \cdot |S_\alpha| \cdot N_C$ that grows linearly with the number of considered components N_C . Note that the state space includes now the states of the hyperparameter(s), which should also be properly discretized. The increase of the state space due to the incorporation of the hyperparameter(s) is, however, much less significant than when considering the joint state space.

A graphical representation of the proposed hierarchical structure is illustrated in Fig. 1, applicable to deterioration processes modeled either as a function of the deterioration rate or conditional on a set of parameters [9]. In either case, the deterioration process d is encoded conditional on the hyperparameter(s) α , along with the deterioration rate τ or parameters θ_d . Evidence collected through observations o_d , does not only serve for updating the damage state, but also for updating the hyperparameters. Since the hyperparameters are parent nodes for all the components, once a component is inspected, the hyperparameters are also updated, influencing all the other components, even those for which evidence was not directly available. The reliability of the system is represented in Fig. 1 by the binary node F_{sys} , conditional on the failure state of the components $F^{(l)}$. At a component level, the failure state is modeled by the binary variable $F^{(l)}$ and corresponds to the subset of the deterioration space classified as failure $S_F \subseteq S$.

This Gaussian hierarchical structure is a mathematically motivated model induced by the convenient formulation available for normal random variables. Let us first consider the special case in which the marginal probability of each considered component deterioration is defined as a standard normal random variable Y_i . Under correlation,

the parameters Y_i are, however, defined as normal random variables with mean $\lambda_i \alpha$ and standard deviation $\sqrt{1 - \lambda_i^2}$ [40]:

$$Y_i = \sqrt{1 - \lambda_i^2} X_i + \lambda_i \alpha \quad (4)$$

Since both X_i and α are independent standard normal random variables, the covariance of Y_i and Y_j can be formulated as:

$$\begin{aligned} cov(Y_i, Y_j) = & (1 - \lambda_i^2) cov(X_i, X_j) + \sqrt{1 - \lambda_i^2} (\lambda_j) cov(X_i, \alpha) \\ & + \sqrt{1 - \lambda_j^2} (\lambda_i) cov(X_j, \alpha) + \lambda_i \lambda_j cov(\alpha, \alpha) \end{aligned} \quad (5)$$

After removing all the terms associated with zero covariance, i.e., $cov(X_i, X_j)$, $cov(X_i, \alpha)$, and $cov(X_j, \alpha)$, we can define the correlation coefficient between Y_i and Y_j as:

$$\rho(Y_i, Y_j) = \lambda_i \lambda_j \quad (6)$$

If all the components are equi-correlated, then $\lambda_i = \lambda_j = \sqrt{\rho(Y_i, Y_j)}$, for all i, j . This presented Gaussian structure is further generalized in this work for the general case of unequally correlated components, by preserving the validity of Eq. (6). For complex correlation configurations, one hyperparameter α might not be sufficient to satisfy Eq. (6), and in that case, one can incorporate additional hyperparameters α , at the expense of a higher computational cost. When N_h hyperparameters are included, the best fit for $\rho(Y_i, Y_j) = \sum_{k=1}^{N_h} (\lambda_{ik} \lambda_{jk})$ can be found via optimization procedures, e.g., least squares [41]. Once the Gaussian correlation structure is specified through the parameters λ , the cumulative distribution of Y_i conditional on the hyperparameter(s) α can be defined as:

$$F_{Y_i|\alpha}(y_i) = \Phi \left[\frac{y_i - \sum_{k=1}^{N_h} \lambda_{ik} \alpha_k}{\sqrt{1 - \sum_{k=1}^{N_h} \lambda_{ik}^2}} \right] \quad (7)$$

For the cases in which the deterioration process is modeled by random variables other than standard Gaussian and considering that a Nataf transformation is applicable [36], then Eq. (7) can be also redefined as:

$$F_{D_i|\alpha}(d_i) = \Phi \left[\frac{\Phi^{-1}[F_{D_i}(d_i)] - \sum_{k=1}^{N_h} \lambda_{ik} \alpha_k}{\sqrt{1 - \sum_{k=1}^{N_h} \lambda_{ik}^2}} \right] \quad (8)$$

where $F_{D_i|\alpha}(d_i)$ stands for the cumulative distribution function of a variable d_i conditional on the hyperparameter(s) α , and Φ is the standard normal cumulative distribution function. In a discrete state space, the belief conditional on the hyperparameters is equal to the difference between the cumulative distribution function at the upper boundary and at the lower boundary of each belief interval:

$$b(d_i|\alpha) = F_{D_i|\alpha}(d_i^+) - F_{D_i|\alpha}(d_i^-) \quad (9)$$

3.2. Belief update under deterioration dependence

We reformulate here the belief update algorithmic scheme introduced in Section 2 for a system under deterioration dependence among components. All necessary implementation steps are listed in Algorithm 2. Bayesian inference is firstly conducted for the conditional beliefs $b(s_{t+1}|\alpha)$ and hyperparameters $b(\alpha)$, propagating uncertainty according to the transition model $p(s_{t+1}|s_t, a_t)$ and observation model $p(o_{t+1}|s_{t+1}, a_t)$. The likelihood of collecting an observation given the hyperparameter(s) $p(o_{t+1}|\alpha, a_t)$, later necessary to update $b(\alpha)$, can be easily computed by marginalizing out the states other than α :

$$p(o_{t+1}|\alpha, a_t) = \sum_{s_{t+1} \in S} [b(s_{t+1}|\alpha) p(o_{t+1}|s_{t+1}, a_t)] \quad (10)$$

Bayesian inference is then conducted for the hyperparameter(s):

$$p(\alpha|o_{t+1}, a_t) = b(\alpha)p(o_{t+1}|\alpha, a_t)/p(o_{t+1}|a_t) \quad (11)$$

After updating the conditional beliefs and common influencing variables, the marginal deterioration beliefs can be computed by marginalizing out the hyperparameter(s) α as:

$$b(s_{t+1}) = \sum_{\alpha \in \Gamma} [p(s_{t+1}|\alpha) b(\alpha)] \quad (12)$$

The effect of maintenance actions on the Gaussian dependence structure has not been explored in the existing literature [22,36], to the best knowledge of the authors. Whereas the defined deterioration dependence is preserved if no maintenance interventions are planned, structural interventions can potentially disrupt the underlying correlation structure. For instance, if a structural system is specified with a correlated initial crack size among fatigue hotspots, this correlation structure will be perturbed after a component is repaired, along with the correlation reduction naturally experienced by the system over time. The correlation evolution associated with the latter is intrinsically quantified through the uncertainty propagation and updating operations formulated in Eqs. (10)–(11), whereas the correlation disruption associated with the former can be modeled by now defining the transition model $p(s_{t+1}|s_t, a_t, \alpha)$ of the involved components also conditional to the hyperparameter(s) α , enabling therefore the removal or modification of the deterioration dependence by redefining and implementing the relevant correlation coefficients λ_i in Eqs. (7)–(9). Additional discussion and implementation of this aspect is presented in the numerical experiments section.

Algorithm 2 Belief update under deterioration dependence for a system of N_c components

```

function UPDATEBELIEF( $b(s_t|\alpha)$ ,  $b(\alpha)$ ,  $a_t$ ,  $o_{t+1}$ )
  for  $1, N_c$  do
     $b(s_{t+1}|\alpha) \leftarrow b(s_t|\alpha) p(s_{t+1}|s_t, a_t)$             $\triangleright$  propagation step
     $b(s_{t+1}|\alpha) \leftarrow b(s_{t+1}|\alpha) p(o_{t+1}|s_{t+1}, a_t)$         $\triangleright$  estimation step
     $p(o_{t+1}|\alpha) \leftarrow \sum_{s_{t+1} \in S} [b(s_{t+1}|\alpha) p(o_{t+1}|s_{t+1}, a_t)]$     $\triangleright$  likelihood
     $b(s_{t+1}|\alpha) \leftarrow b(s_{t+1}|\alpha)/p(o_{t+1}|b_t, a_t)$             $\triangleright$  normalization step
     $b(\alpha) \leftarrow b(\alpha) p(o_{t+1}|\alpha, a_t)/p(o_{t+1}|a_t)$             $\triangleright$  hyperparameter(s)
    update
  end for
  for  $1, N_c$  do
     $b(s_{t+1}) \leftarrow \sum_{\alpha \in \Gamma} [b(s_{t+1}|\alpha) b(\alpha)]$             $\triangleright$  marginalizing out
    hyperparameter(s)
  end for
  return  $b(s_{t+1})$ 
end function

```

3.3. System structural reliability and system cost model

As input to the I&M decision-making problem (Section 2), utilities, r_{ins} and r_{rep} , are assigned to inspection and repair actions, respectively,

specified according to available options and settings in each problem. The annual risk of a system failure, r_F , is defined accounting for two consecutive time steps (e.g., years) and its associated system failure cost, r_f , as:

$$r_F = (p_{F_{sys,t+1}} - p_{F_{sys,t}})r_f \quad (13)$$

The system structural failure event, as illustrated in Fig. 1 by the node F_{sys} , is specified by a binary variable $p_{F_{sys}}$, indicating the failure and survival states, conditional on the belief state $b(s)$ of the structural components, as these are determined by the performed I&M actions in time. Within the deep reinforcement learning approach presented in Section 4, $P_{F_{sys}}$ can be computed via closed-form procedures and/or supported by efficient matrixial algorithms [41]; or it can be computed following a general scheme, obtaining $p_{F_{sys}}$ through a simulator [42]. By assigning utilities to the system state, the importance of each structural element to the global risk of a system failure is implicitly accounted. To illustrate the effect of defining the failure risks at the system level, I&M strategies for a redundant 2-dimensional frame structure are later explored in Section 5.

In most structural systems, inspection and repair actions are not planned separately for each structural element. Maintenance campaigns are instead scheduled, collecting information or performing repairs on a group of structural components. The cost model can thus be adapted from Eq. (3), to include a fixed campaign cost, r_{camp} , incurred every time a campaign is planned, along with inspection, $r_{ins}^{(l)}$, and repair, $r_{rep}^{(l)}$, costs assigned to the individual components according to any linear/nonlinear function of choice, $\mathcal{H}(\cdot)$, e.g., simple linear sum operator, as:

$$r_{tot} = r_{camp} + \mathcal{H}(r_{ins}^{(l)}, r_{rep}^{(l)}) + r_F \quad (14)$$

4. Optimal I&M planning via deep reinforcement learning

I&M planning decision problems, formulated as POMDPs (as explained in Sections 2 and 3), can be solved by dynamic programming algorithms, e.g., via exact alpha-vector value iteration [43]. In practice, however, exact value iteration can be applied to only very small state space problems due to the complexity associated with the exponential increase in the number alpha vectors with the number of observations at every iteration. Recently, I&M planning decision problems, at the component level, formulated as POMDPs and characterized by multiple states have been efficiently solved via point-based POMDP algorithms [9,14,44]. Point-based solvers exploit the fact that the value function $V(\mathbf{b})$ (Eq. (2)) is piece-wise linear and convex and can be thus parameterized by a set of $\mathbf{v}_p \in \mathcal{V}_p$ vectors, each of which is associated with a specific action $a \in \mathcal{A}$. The optimal value function, $V^*(\mathbf{b})$, can be therefore defined in terms of a set of \mathbf{v}_p vectors [45].

State-of-the-art point-based POMDP solvers mainly differ on their approach of sampling reachable belief points, and the way Bellman backup operations are executed, e.g., in [37,46,47]. The reader is directed to [44] for a detailed comparison of point-based solvers applied to infrastructure I&M settings. While point-based solvers are able to efficiently provide optimal policies at the component level and for realistically large systems, the dimensionality still becomes a limiting factor in high-dimensional state, action, and observation space settings, typical in structural systems. Deep Reinforcement Learning (DRL) provides then a powerful solution in such settings, as the value or policy function can be parameterized with deep artificial neural networks. Thereby, the planning task reduces to finding a number of parameters that is much lower than the number of original states and actions of the problem. The interested reader is directed to [48,49] for a well elaborated introduction and discussion on DRL. In our proposed approach, we integrate the factored POMDP formulation introduced in Section 2 with a Decentralized Deep Multi-agent Actor-Critic (DDMAC) scheme, adopted from [23]. This combination provides an efficient algorithmic platform for inspection and maintenance planning of structural systems

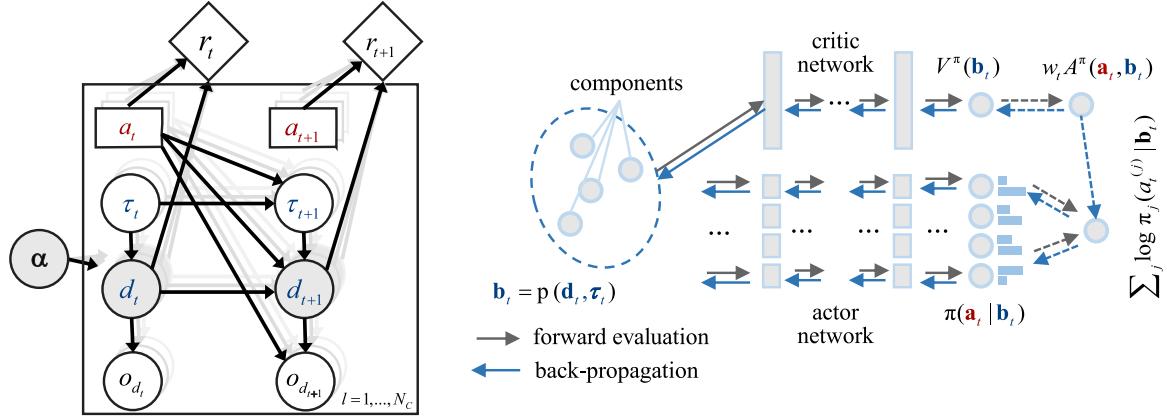


Fig. 2. On the left: Representation of a factored POMDP derived from the deterioration rate dynamic Bayesian network introduced in Fig. 1. The deterioration process d_t , influenced by the deterioration rate τ_t , conditional on the hyperparameters α and partially observed through o_{d_t} , is controlled by the action decision node a_t . A reward r_t is collected as a result of taking action a_t at state d_t . On the right: Deep Decentralized Multi-Agent Actor Critic (DDMAC) featuring the critic network at the top, and a group of actor networks, one for each component, at the bottom.

under deterioration, reliability and cost dependencies, in large-scale multi-component environments.

Each component of the system is controlled by the stochastic policy $\pi(a^{(l)}|\mathbf{b}, \Theta^\pi)$ provided by a group of multi-agent actor networks, defined as a function of parameters Θ^π , as illustrated on the right side of Fig. 2 with light blue bars. In many applications, DRL policies after training are nearly deterministic, suggesting one action in particular, whereas stochastic policies are often optimal in constraint environments [27]. In our implementation, we consider agents acting as independent units in a decentralized manner, i.e., the actions taken by one actor are naturally not affected by the actions taken by other actors:

$$\pi(\mathbf{a}|\mathbf{b}) = \prod_{l=1}^{N_c} \pi_l(a^{(l)}|\mathbf{b}) \quad (15)$$

The input to the actor networks corresponds to the marginal belief states of all components along with the deterioration rate and the time step encoded as a zero-one vector. For instance, if the environment is described by the factored POMDP represented on the left side of Fig. 2, the actor networks receive the deterioration belief states $\mathbf{b}_{(s_d)}$ and deterioration rate states $\mathbf{b}_{(s_\tau)}$ for all components (in $|S_d| \cdot N_c$ and $|S_\tau| \cdot N_c$ matrix formats, respectively), plus an input indicating the time step $t \in t_N$. If deterioration dependence is included through a hierarchical Gaussian model, as explained in Section 3, then conditional beliefs $\mathbf{b}_{(s_d|\alpha)}$ and hyperparameters beliefs $\mathbf{b}(\alpha)$ should also be used while simulating the deterioration environment.

Even for environments under deterioration dependence, the neural networks only receive as input the components' marginal beliefs $\mathbf{b}^{(l)}$, for all l , computed by following the steps listed in Algorithm 2. ReLU activation functions are used for the hidden layers of the actor networks, and the output layer is activated by a softmax function, generating the output policy as a probability distribution over the available actions.

The actor network weights are adjusted/updated according to the noisy rewards collected from a batch of previous experiences, following an off-policy training approach that offers more sample efficiency than on-policy training algorithms. A replay buffer [50] stores beliefs \mathbf{b}_t , actions \mathbf{a}_t , rewards $r(\mathbf{b}_t, \mathbf{a}_t)$ and behavior policies μ_t , experienced during the simulations of the deterioration environment. The off-policy gradient estimator is thus formulated with samples generated by a behavior policy μ , different from π , and corrected with the truncated importance sampling weight $w_t = \min\{c, \pi(\mathbf{a}_t|\mathbf{b}_t)/\mu(\mathbf{a}_t|\mathbf{b}_t)\}$, with $c > 0$ [23]:

$$\mathbf{g}_{\Theta^\pi} = \mathbf{E}_{\mathbf{a}_t \sim \mu} \left[w_t \left\{ \sum_{i=1}^{N_c} \nabla_{\Theta^\pi} \log \pi_i(a_t^{(i)}|\mathbf{b}_t, \Theta^\pi) \right\} A^\pi(\mathbf{b}_t, \mathbf{a}_t|\Theta^\pi) \right] \quad (16)$$

The advantage function $A^\pi(\mathbf{b}_t, \mathbf{a}_t)$ indicates how optimal is action \mathbf{a}_t with respect to the current estimated value function $V^\pi(\mathbf{b}_t)$ and defined

Algorithm 3 Deep Decentralized Multi-agent Actor Critic (DDMAC)

```

Initialize replay buffer
Initialize actor and critic network weights  $\Theta^\pi, \Theta^V$ 
for episode = 1, M do
    for t = 1,  $t_N$  do
        Select action  $\mathbf{a}_t$  at random according to exploration noise
        Otherwise select action  $\mathbf{a}_t \sim \mu_t = \{\pi_j(\cdot|\mathbf{b}_t, \Theta^\pi)\}_{j=1}^{N_c}$ 
        Collect reward  $r(\mathbf{b}_t, \mathbf{a}_t)$ 
        Observe  $o_{t+1}^{(l)} \sim p(o_{t+1}^{(l)}|\mathbf{b}_t, \mathbf{a}_t)$  for  $l = 1, 2, \dots, N_c$ 
        Compute beliefs  $\mathbf{b}_{t+1}$ :  $\text{UPDATEBELIEF}(\mathbf{b}_t, \mathbf{a}_t, o_{t+1})$ 
        Store experience  $(\mathbf{b}_t, \mathbf{a}_t, \mu_t, r(\mathbf{b}_t, \mathbf{a}_t), \mathbf{b}_{t+1})$  in replay buffer
        Sample batch of  $(\mathbf{b}_i, \mathbf{a}_i, \mu_i, r(\mathbf{b}_i, \mathbf{a}_i), \mathbf{b}_{i+1})$  from replay buffer
        If  $\mathbf{b}_{i+1}$  is terminal state  $A_i^\pi = r(\mathbf{b}_i, \mathbf{a}_i) - V^\pi(\mathbf{b}_i, \Theta^V)$ 
        Otherwise  $A_i^\pi = r(\mathbf{b}_i, \mathbf{a}_i) + \gamma V^\pi(\mathbf{b}_{i+1}, \Theta^V) - V^\pi(\mathbf{b}_i, \Theta^V)$ 
        Update actor parameters  $\Theta^\pi$  according to gradient:
             $\mathbf{g}_{\Theta^\pi} \simeq \sum_i w_i \{ \sum_{j=1}^{N_c} \nabla_{\Theta^\pi} \log \pi_j(a_i^{(j)}|\mathbf{b}_i, \Theta^\pi) \} A_i^\pi$ 
        Update critic parameters  $\Theta^V$  according to gradient:
             $\mathbf{g}_{\Theta^V} \simeq \sum_i w_i \nabla_{\Theta^V} V^\pi(\mathbf{b}_i|\Theta^V) A_i^\pi$ 
    end for
end for

```

in a temporal difference learning fashion as:

$$A^\pi(\mathbf{b}_t, \mathbf{a}_t|\Theta^V) \simeq r(\mathbf{b}_t, \mathbf{a}_t) + \gamma V(\mathbf{b}_{t+1}|\Theta^V) - V(\mathbf{b}_t|\Theta^V) \quad (17)$$

The value function is approximated by the critic network, defined as a function of parameters Θ^V , as illustrated on the right side of Fig. 2. Whereas the critic network receives the same input as the actor network (components marginal beliefs, deterioration rates, and a time step indicator), the output of the critic is the value function, i.e., one scalar value that indicates the expected reward of the system. The critic network provides the value function used by the advantage function $A^\pi(\mathbf{b}_t, \mathbf{a}_t|\Theta^V)$, acting, therefore, as a critic who is determining how good the action taken by the actor network is. The training of the critic network also follows a temporal difference approach, collecting experiences from the replay buffer, and adjusting the critic parameters Θ^V according to the gradient:

$$\mathbf{g}_{\Theta^V} = \mathbf{E}_{\mathbf{a}_t \sim \mu} \left[w_t \nabla_{\Theta^V} V^\pi(\mathbf{b}_t|\Theta^V) A^\pi(\mathbf{b}_t, \mathbf{a}_t|\Theta^V) \right] \quad (18)$$

All the algorithmic steps are described in Algorithm 3. With our proposed method, we are able to find optimal I&M policies for structural systems featuring very high dimensional state, action and observation spaces. Moreover, the obtained DDMAC policies intrinsically account for system-effects (Section 3) as the actor network is adjusted according

to the rewards collected by simulating the deteriorating environment at the system level. Specifically, the integration of DDMAC with a deterioration environment simulated with a factored POMDP (Section 2) enables the identification of optimal I&M policies that concurrently consider the following system effects:

- (i) Deterioration dependence among components (statistical dependence): A Gaussian hierarchical model efficiently captures the deterioration dependence, e.g., initial crack size, or loading. The belief of each component is conditional on the common hyperparameter(s) $\mathbf{b}(s_d|\alpha)$. Under statistical deterioration dependence, information collected by inspecting one component informs belief updates for other components as per the observed deterioration. The influence of this system-effect on the policy is explored via numerical experiments in Section 5, for a 9-out-of-10 system and for a steel frame structural system subject to fatigue deterioration.
- (ii) System structural reliability (structural dependence): Failure risk is computed at the system level, by multiplying annual risk with a negative reward r_F that is defined as a function of the components structural health, as shown in Eq. (13). The actors, even though acting individually, are all conditioned on all component beliefs, thus knowing the system reliability. DDMAC is able to intrinsically adjust the policy according to the relative importance of each component to the system structural reliability, as demonstrated with the numerical experiments conducted for the steel frame structural system (Section 5).
- (iii) Inspection and maintenance cost model (cost dependence): A campaign cost r_{camp} is included, in the applications as a base cost, if at least one component is inspected or repaired, on top of the inspection or repair cost for each inspected or repaired component, as shown in Eq. (14). Since DDMAC collects rewards at the system level, the campaign cost model affects the resulting I&M policies, concentrating inspection and repair actions at particular time steps, as observed in the numerical experiments conducted for the 9-out-of-10 system (Section 5).

5. Numerical experiments

DDMAC inspection and maintenance policies are tested for a 9-out-of-10 system under fatigue deterioration, exploring the different statistical, structural, and cost dependencies. A second set of numerical experiments is conducted to investigate the efficiency of DDMAC policies for a 2D steel frame, also known as Zayas frame, used as a benchmark structural system for offshore engineering collapse analyses [51]. The numerical experiments are conducted on an Intel Core i9-7900X processor with a clock speed of 3.30 GHz.

Fatigue deterioration model

The components explored throughout the numerical investigations are assumed to be exposed to a similar fatigue deterioration, described according to the Markovian model, originally proposed in [52]:

$$d_{t+1} = \left[\left(1 - \frac{m}{2}\right) C_{FM} S_R^m \pi^{m/2} n + d_t^{1-m/2} \right]^{2/(2-m)} \quad (19)$$

where the crack depth, d , evolution over time, t , follows a linear-elastic fracture mechanics law with material parameters C_{FM} and m , stress range S_R , and n annual stress cycles. At the component level, failure occurs if the crack depth, d , exceeds a critical size, d_c , that corresponds to the plate thickness. In a stochastic environment, the initial crack depth, d_0 , along with fracture mechanics model parameters are either represented by random variables or deterministic parameters as listed in Table 1. The failure probability p_{F_t} , defined as $p_{F_t} = Pr[g_t \leq 0]$, can be computed following, for instance, a through-thickness failure criterion [10] by formulating the failure limit state at time step t as:

$$g_t = d_c - d_t \quad (20)$$

Table 1

Random variables and deterministic parameters utilized to model the fatigue deterioration of the components in the numerical experiments.

Variable	Distribution	Mean	SD
$\ln(C_{FM})$	Normal	-35.2	0.5
S_R (N/mm ²)	Normal	70	10
d_0 (mm)	Exponential	1	1
m	Deterministic	3.5	-
n (cycles)	Deterministic	10^6	-
t_N (yr)	Deterministic	30	-
d_c (mm)	Deterministic	20	-

Table 2

Description of the discretization scheme implemented for the factored deterioration rate POMDP.

Variable	Interval boundaries
Deterioration rate model	
S_d	$0, \exp\left\{\ln(10^{-4}) : \frac{\ln(d_c) - \ln(10^{-4})}{28} : \ln(d_c)\right\}, \infty$
S_r	0 : 1 : 30

The fatigue deterioration is encoded in a deterioration rate DBN model, ultimately shaping a factored POMDP, as shown on the left side of Fig. 2, and presented in Section 2. The continuous crack depth, d , is adequately discretized into $|S_d| = 30$ states conditional on $|S_r| = 31$ fully observable deterioration rates states. The intervals and state space utilized for this deterioration rate model are listed in Table 2. In terms of observation model, the inspection quality is quantified with a Probability of Detection curve $PoD(d) \sim Exp[\mu = 8]$. Further details on the fatigue deterioration or observation model, including an extensive investigation of the discretization scheme can be found in [9].

5.1. I&M planning for a 9-out-of-10 system

The system explored in this application is composed of ten components, each of which is subjected to a non-stationary fatigue deterioration, as described earlier in this Section. The system is assumed to be functional if at least 9-out-of-10 components are operational (not failed), thus characterized with a single step change in terms of redundancy with respect to a series system, which would correspond to the case of a 10-out-of-10 system. The system failure probability $p_{F_{sys}}$ is efficiently computed here, as a function of the failure state of all components, by following the recursive method proposed in [53].

Description of the I&M decision problem

A total of eight I&M planning scenarios are investigated, exploring different deterioration, risk, and cost dependencies among components. In terms of deterioration dependence, some environments are specified with an equally correlated initial crack size, d_0 , among components, defined by an equal correlation $\rho_{eq} = 0$, $\rho_{eq} = 0.4$ and $\rho_{eq} = 0.8$, respectively. Additionally, a deterioration environment is examined with an unequally correlated ρ_{uq} initial crack size, d_0 , among components. The unequal deterioration dependence case is originally specified with a different correlation among components of either $\rho = 0.4$, $\rho = 0.6$, or $\rho = 0.8$, as shown on the left side of Fig. 3. After a Gaussian hierarchical structure with two hyperparameters is optimized, by computing the λ parameters with the objective of satisfying Eq. (6), an approximated correlation structure is obtained with relatively small errors, as shown on the right side of Fig. 3. The approximated correlation structure with two hyperparameters is deemed to be sufficiently accurate for the conducted experiments. Otherwise, a more accurate correlation structure can be achieved by adding more hyperparameters, at the expense of additional computational cost, as explained in Section 3. For each of the aforementioned environments, specified with different deterioration dependencies, two I&M cost models are further investigated, i.e., an I&M cost model that incurs inspection and repair costs individually;

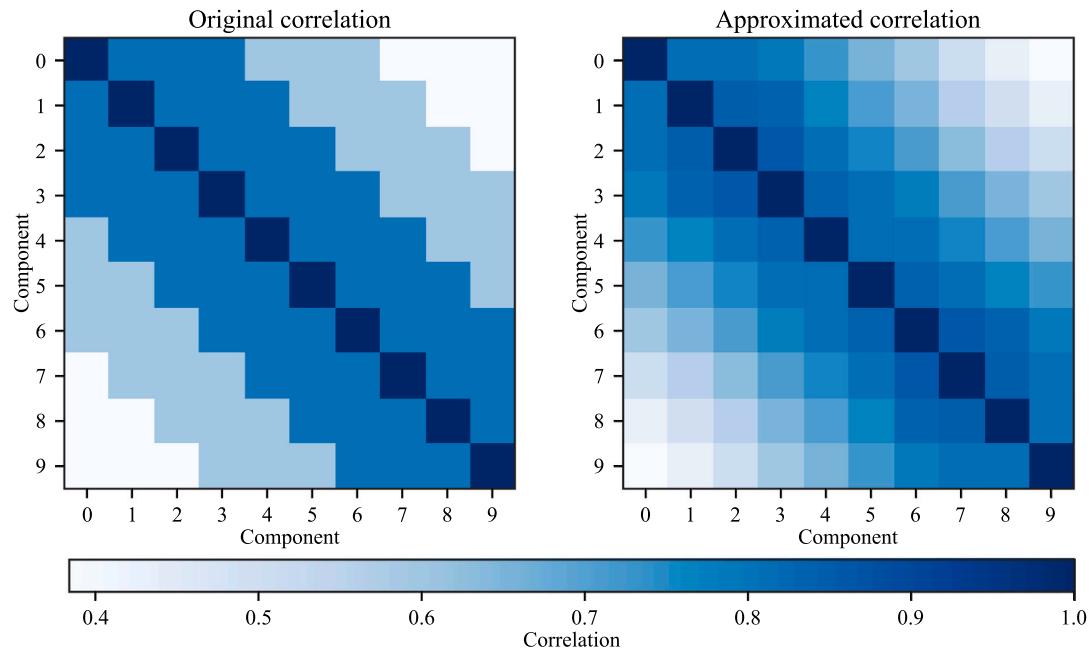


Fig. 3. Representation of the initial crack size dependence among the components of the unequally correlated 9-out-of-10 system. The original deterioration correlation is represented by the colored matrix on the left. The approximated correlation structure, resulting from the derived Gaussian hierarchical model with two hyperparameters, is displayed on the right colored matrix.

and an I&M cost model in which an initial campaign cost exists, if at least one component is inspected or repaired, plus a cost surplus per inspected or repaired component.

Since each component, herein denoted as fatigue hotspot, contains 930 states, defined by the joint space of 30 crack states S_d and 31 deterioration rate states S_r , these are a total of 9,300 states in the experiments that do not consider deterioration correlation ($\rho_{eq} = 0$). For experiments under equal correlation (ρ_{eq}) the total states become 744,080, rising up to $5.95 \cdot 10^7$ for deterioration environments under unequal correlation (ρ_{uq}). The increase of the state space corresponds to the incorporation of the Gaussian hierarchical model, in which crack and deterioration rate states are formulated conditional on the hyperparameter(s) states. When the deterioration correlation is modeled equally for all components, only one hyperparameter is sufficient to satisfy Eq. (6), while two hyperparameters are added for the case of unequal correlation, as explained earlier. Each hyperparameter is discretized into 80 states, initially prescribed with equal probability for each state. Note here the importance of optimizing the number of hyperparameters included in the model, as the state space grows exponentially with the number of considered random variables. By formulating the POMDPs' transition model as dynamic Bayesian networks, the dimensionality is reduced from $|S_d|^2|S_r|^2$, in a flat structure, to $|S_d|^2|S_r| + |S_r|^2$ for the uncorrelated scenario. In that case, the transition model of only one component is reduced from 864,900 to 28,861 elements. Moreover, the formulation of the environment through a hierarchical deterioration dependence model importantly enables the decoupling of the joint state space at the system level, which would grow exponentially for a flat POMDP structure, but instead grows linearly now. For instance, in the setting under unequal deterioration dependence, the joint space would be described by $\{|S_d||S_r|\}^{N_C}$, equaling 930^{10} states, while it is now instead defined by $\{|S_d||S_r|N_C|S_\alpha||S_\beta| + |S_\alpha| + |S_\beta|\}$, thus resulting in $930 \cdot 10 \cdot 80^2 + 80 \cdot 2 \simeq 5.95 \cdot 10^7$ states in the hierarchical model, with two hyperparameters (α and β) discretized into 80 states.

In terms of the neural networks' architecture, DDMAC is laid out in this application with two hidden fully-connected layers of 100 neurons for each actor network, and two hidden fully-connected layers with 200 neurons for the critic network. The learning rate is adjusted during the

training of the networks from 10^{-4} to 10^{-5} for the actor, and from 10^{-3} to 10^{-4} for the critic. The exploration is set up initially with a 100% random noise, decreasing linearly over the first 20,000 episodes to a random noise of 1%, held constant for the remaining episodes. A more stable and efficient training was found when a prioritization of do-nothing actions is implemented at the beginning of the training, because this allows visitation of more states at various horizon time steps, thus better exploration. The computational time usually sufficient for convergence ranged from 30 to 46 hours, with the setting under unequal deterioration correlation demanding a longer computation due to its higher-dimensional state space. Please note that the reported computational time is influenced by the specific computational resources used and the code efficiency.

Following typical fatigue I&M planning settings, inspection and repair decisions are combined into three available actions per component: *do-nothing/no-inspection*, *do-nothing/inspection*, and *perfect repair/no-inspection*. The action *perfect-repair/inspection* is considered a priori suboptimal, without loss of generality, as it would be unusual to plan an inspection just after a component returns to its initial state. Inspections provide binary indications, i.e., *detection* or *no-detection* of a crack according to the observation model. In terms of costs, two different scenarios are considered. In the first case, inspection and repair costs are incurred independently per component, i.e., $r_{ins} = -1$ and $r_{rep} = -20$, respectively. In the second case, a campaign cost of $r_{camp} = -5$ is incurred if at least one component is inspected or repaired, plus a surplus per inspected or repair component of $r_{ins} = -0.2$ and $r_{rep} = -20$ money units, respectively. The consequence of a system structural failure is $r_F = -10,000$ money units for both cases, and the discount factor γ is 0.95 in all the experiments.

In order to verify the optimality of the obtained DDMAC policies, predefined heuristic decision rules, adopted from [22], are optimized and compared against the results provided by DDMAC strategies. The investigated heuristic-based policies are dictated by (i) the interval between equidistant inspections Δ_{ins} , (ii) how many components n_{ins} are inspected at each campaign, in which the n_{ins} components with higher failure probability p_F are prioritized, and (iii) a perfect repair action is undertaken after a crack is detected. Initially, all the combinations of heuristics, i.e., interval between inspections Δ_{ins} and number of

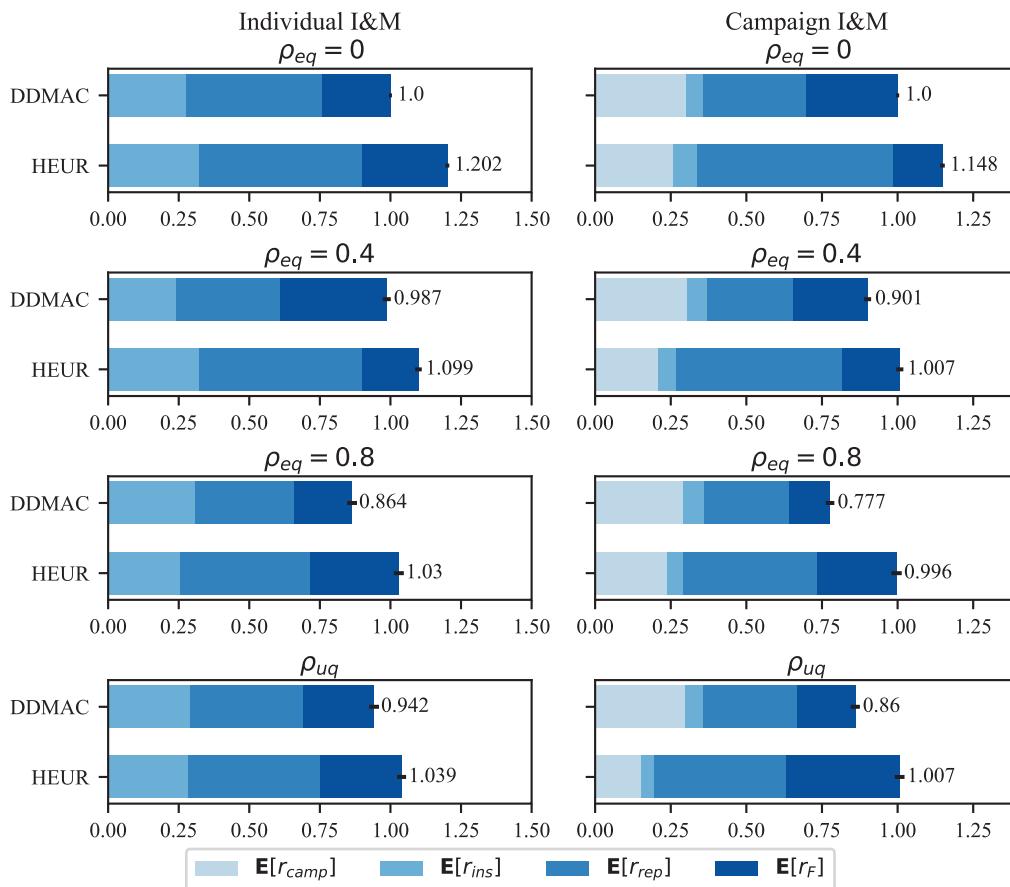


Fig. 4. Expected cost results of all the numerical experiments conducted for the 9-out-of-10 system, divided into campaign $E[r_{camp}]$, inspection $E[r_{ins}]$, perfect-repair $E[r_{rep}]$ and failure $E[r_f]$ expected costs. On the left, DDMAC and heuristic policies, specified with an I&M cost model, are compared for different deterioration correlation environments. Likewise, on the right, DDMAC and heuristic policies are compared for different levels of deterioration dependence, yet specified with a campaign I&M cost model.

components inspected per campaign n_{ins} , are evaluated over 3,000 policy realizations. Then, the 5 sets of heuristic rules that yielded the minimum expected costs are evaluated again, this time over 10,000 policy realizations, and at the end, the set of heuristics that minimized the expected total costs are selected for comparison against DDMAC-based policies, also evaluated over 10,000 policy realizations. The resulting set of optimized heuristics is listed in Appendix A (Table A.1).

Results and discussion

The life-cycle expected costs obtained by evaluating the investigated policies are displayed in Fig. 4, sorted in two main categories according to the specified cost model, comparing DDMAC and optimized heuristic policies and investigating the effect of adding campaign I&M costs. For each category, four degrees of deterioration correlation are compared, i.e., no correlation ($\rho_{eq} = 0$), equal correlation with ($\rho_{eq} = 0.4$), equal correlation with ($\rho_{eq} = 0.8$) and unequal correlation (ρ_{uq}). In all explored numerical examples, DDMAC outperforms the optimized heuristics, yielding life-cycle cost reductions ranging from 9.7% to 21.9%. The difference is more predominant for the case in which inspections and repairs are planned separately because the explored heuristic decision rules plan inspections for a group of n_C components, being thus more tailored to the campaign I&M setting. A closer examination reveals that DDMAC policy provides lower inspection, repair, and failure expected costs, with respect to heuristics-based policies, for the uncorrelated deterioration experiment specified with the individual I&M cost model. In this case, the savings on repairs are more significant probably because the heuristic policy prescribes a repair anytime a crack detection is observed, while DDMAC-based policy usually requires more evidence than a single detection instance. With

regard to deterioration dependence, highly correlated environments result generally in lower expected total cost, as observed in Fig. 4.

In order to facilitate the interpretation of the identified policies, a snapshot corresponding to a specific policy realization is additionally displayed in Fig. 5, illustrating the failure probability of the analyzed systems along with each component failure probability. Evidently, the resulting strategies importantly influence the results reported in terms of system reliability, as they control components' inspection and maintenance actions. Information collected on one component, in environments under deterioration correlation, also provides information to other components' states in the system. One can clearly observe the effect of inspection and repairs on the resulting component and system failure probabilities as well as the oscillation of the hyperparameters around an uncorrelated mean of 0 when inspections among components are not consistent, i.e., some inspections indicate damage and some others do not, whereas the opposite is observed when all inspections are consistent, e.g., no damage indicators.

For instance, a crack detection observed on component 9 for the case of equal correlation $\rho_{eq} = 0.8$, leads to an incremental increase on the failure probability of other non-repaired components, as indicated inside the green rectangles on the lower-left corner of Fig. 5. This effect can be also visualized when observing the impact of a crack detection on component 4, for the case under unequal deterioration dependence, marked by green rectangles on the lower-right plot of Fig. 5. In this case, components 3 and 5, highly correlated with component 4, as indicated in Fig. 3, are clearly affected by the observed crack detection.

Moreover, highly correlated deteriorating environments induce higher variability on the expected total costs, as shown by the black error bars in Fig. 4. The variability can be attributed to the very different resulting policy paths following inspections, depending on whether

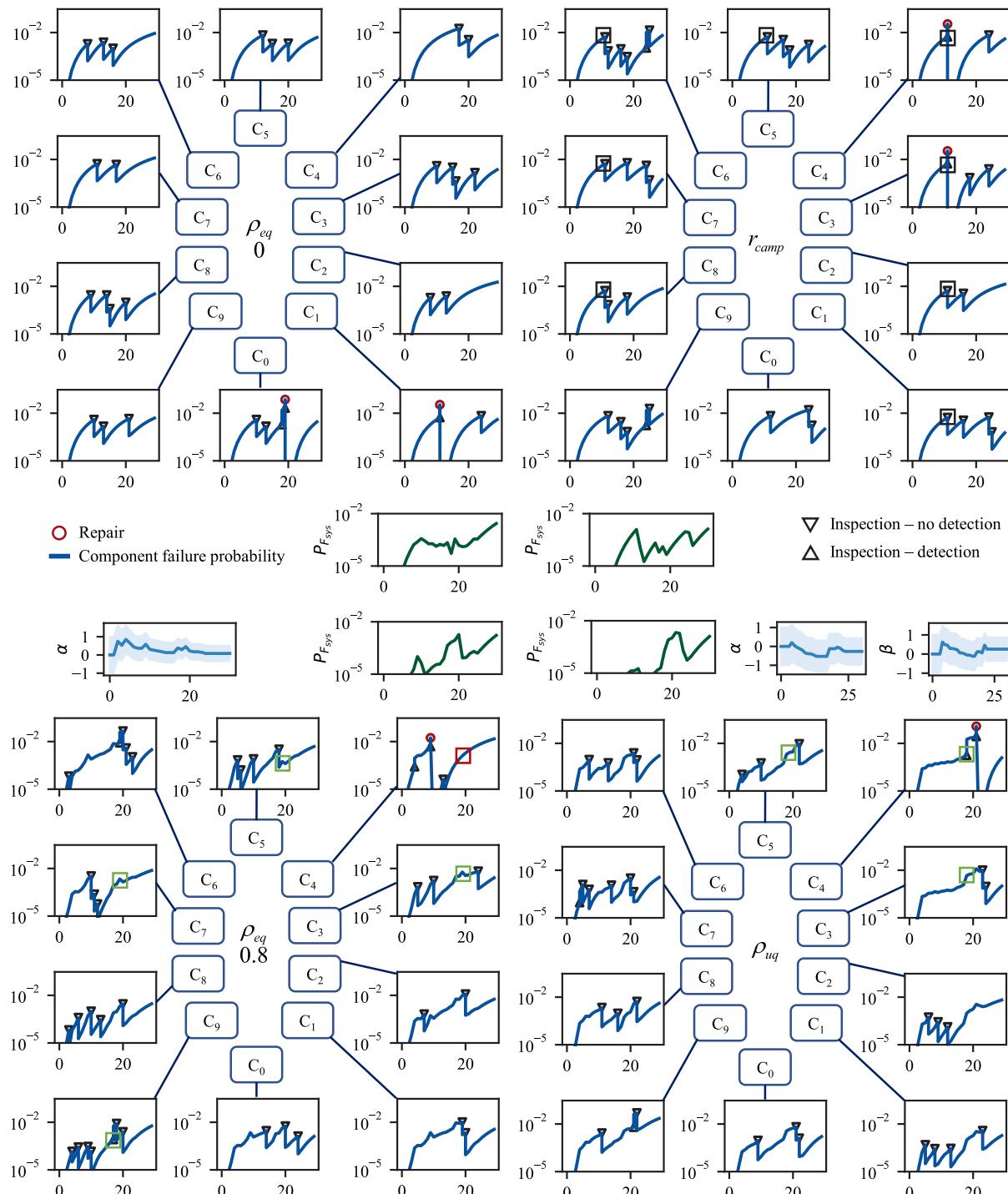


Fig. 5. 9-out-of-10 system policy realizations: (Upper-left) uncorrelated deterioration and individual cost model; (Upper-right) uncorrelated deterioration and campaign cost model; (Lower-left) equally correlated environment ($\rho_{eq} = 0.8$) and individual cost model; (Lower-right) unequally correlated deterioration and individual cost model. Failure probabilities at the component level are depicted by blue lines, inspection indications are represented by upwards (detection) or downwards (no-detection) triangles and repairs are circled in red. At the system level, the failure probability is represented by green diagrams and the evolution of the hyperparameters, under correlated deterioration, is described by light-blue graphs, in which the expected value is bounded by ± 1 standard deviation. As mentioned in the text, specific policy and inference effects are marked with colored rectangles.

the collected observations indicate cracks or not. If a crack is detected on one component, the other components' failure probabilities increase, and repair actions or additional inspections will be planned thereafter. Conversely, if a crack is not detected, the failure probability of all the correlated components will decrease, inducing less repair actions in the future. Interestingly, policies under dependent environments do not always plan fewer inspections, as it could be expected due

to the additional information gained through the underlying correlation among components, but instead highly correlated environments might plan more inspections, often resulting in significant failure risk reductions, as displayed for the case with $\rho_{eq} = 0.8$ in Fig. 4. The effect of repair actions on the deterioration dependence structure can also be clearly observed. Once an element is repaired, its damage belief becomes independent from the global hyperparameter(s), and thus inspection outcomes from other components do not influence the

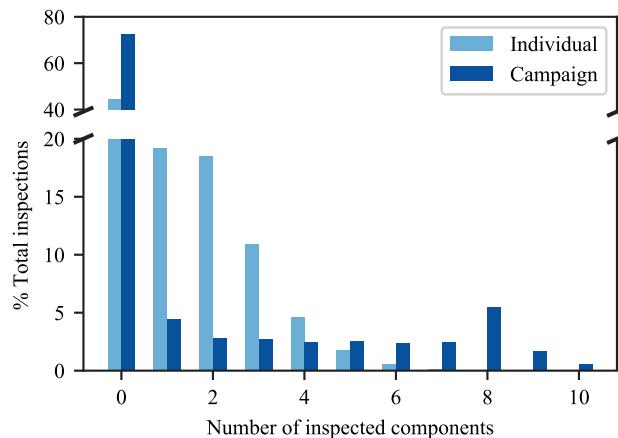


Fig. 6. Comparison of DDMAC policies specified with either individual (light blue) or campaign (dark blue) I&M cost models. For each case, the number of inspected components per time step is represented in a histogram based on 10,000 policy realizations.

repaired component, and vice versa. As illustrated with a red rectangle at the lower-left corner of [Fig. 5](#) (i.e., equal correlation $\rho_{eq} = 0.8$ setting), after component 4 is repaired, its failure probability is not influenced now by inspection results retrieved from other components, e.g., a crack detection observed on component 9.

To further investigate the effect of including campaign utilities within the cost model, a histogram over 10,000 policy realizations is shown in [Fig. 6](#) for a DDMAC policy in which inspections and repairs are incurred separately (light blue) and another DDMAC policy considering the expense of campaign actions (dark blue). The deterioration environment for both DDMAC policies does not consider deterioration correlation among components. The emphasis of [Fig. 6](#) is on the number of components inspected on every occasion an inspection is planned. If inspections and repairs are paid individually for each component, information on only one or two components usually suffices, whereas inspection of more than five components per time step is rare. In contrast, eight-component inspections become the predominant inspection decision if an initial campaign cost is included in the cost model. This system effect can be visualized, for example, in the policy realizations shown at the top of [Fig. 5](#), in which black rectangle-marked components are inspected at the same year for the DDMAC policy under campaign costs. The policy for the campaign cost model, therefore, tends to group inspection and repair actions at the same year, avoiding, if possible, unnecessary campaign costs associated with one or two inspected components. In some cases, campaigns are however planned for only one or two components, contrasting with the static inspection decision rules imposed by heuristics, where a specific number of inspections is fixed for all the inspection campaigns. Based on the above, we observe that DDMAC is able to devise dynamic I&M policies according to the specified cost model, whether under campaign costs or individual inspection and repairs, and to provide an advanced, flexible, and adaptive decision-making framework.

5.2. I&M planning for Zayas frame

In the first set of numerical experiments, conducted for a 9-out-of-10 system, the focus was mainly directed to the investigation of the deterioration dependence among identical components and the effects of including a campaign cost within the cost model. In this second application, we further explore how I&M DDMAC policies are able to inherently capture the relative importance of each element with respect to the system structural reliability. The structural system of study, in this case, is the 2-dimensional Zayas frame, well studied in many structural reliability analysis applications [[51,54,55](#)]. Zayas frame is composed of two columns, which along with 13 braces, sustain a rigid

beam at the top. The geometry and material properties used in this work are presented in [Appendix B](#) ([Fig. B.1](#)) and are the same as the ones used in [[55](#)].

Description of the I&M decision problem

In this application, DDMAC policies are identified for two I&M settings: (i) under equal deterioration dependence among components with $\rho_{eq} = 0.4$, and (ii) assuming independence among components' deterioration. The state space for the latter includes 30 crack states along with 31 deterioration rate states, for each of the 22 hotspots (i.e., components), resulting therefore in a total of 20,460 input variables; while the input variables for the former climbs to approximately $1.6 \cdot 10^6$ states, including 80 states for the one discretized hyperparameter. The benefits associated with the proposed decoupled hierarchical structure are very significant, since the state space has a dimension of 930^{22} if the joint states of all hotspots are explicitly considered.

Similarly to the experiments reported in [Section 5.1](#), the decision maker is here able to select three actions per hotspot at each time step: *do-nothing/no-inspection*, *do-nothing/inspection*, and *perfect repair/no-inspection*. Again, inspections provide binary crack indications (detected or no-detected), equally modeled for each component by the observation model described in the beginning of this Section. As for the cost model, inspections and repairs (planned individually for each component) cost $r_{ins} = -1$ and $r_{rep} = -15$ money units, respectively, while the system failure cost is defined as $r_F = -50,000$ money units. All costs are discounted to the present value by a $\gamma = 0.95$ factor. DDMAC's architecture is similar to the first application, featuring two hidden fully-connected layers of 150 neurons for each actor, and two hidden layers of 300 neurons for the critic network. Learning rate, prioritization of actions and additional exploration settings are equally defined as for the first application. In this case, the computational time usually sufficient for convergence ranged from 85 to 96 h for the uncorrelated and equally correlated settings, respectively, influenced again by the used computational resources and the code efficiency. The investigated heuristic-based policies rely here also on the same set of decision rules introduced in the former numerical experiments, accounting in this case, for inspections intervals Δ_{ins} and inspected hotspots per campaign n_C . Both DDMAC and heuristic policies are evaluated over 30,000 episodes and the results, in terms of expected total costs, are showcased in [Fig. 7](#).

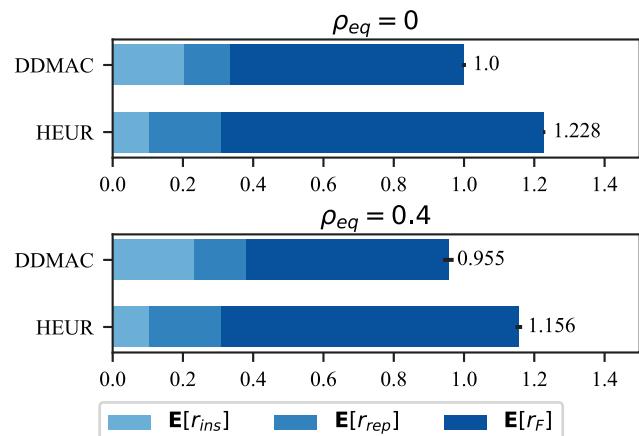


Fig. 7. Expected cost results for the numerical experiments conducted for the Zayas frame, divided into inspection $E[r_{ins}]$, perfect-repair $E[r_{rep}]$ and failure $E[r_F]$ expected costs. DDMAC and heuristics are compared under an uncorrelated deterioration environment at the top, and under an equally correlated environment ($\rho_{eq} = 0.4$) at the bottom.

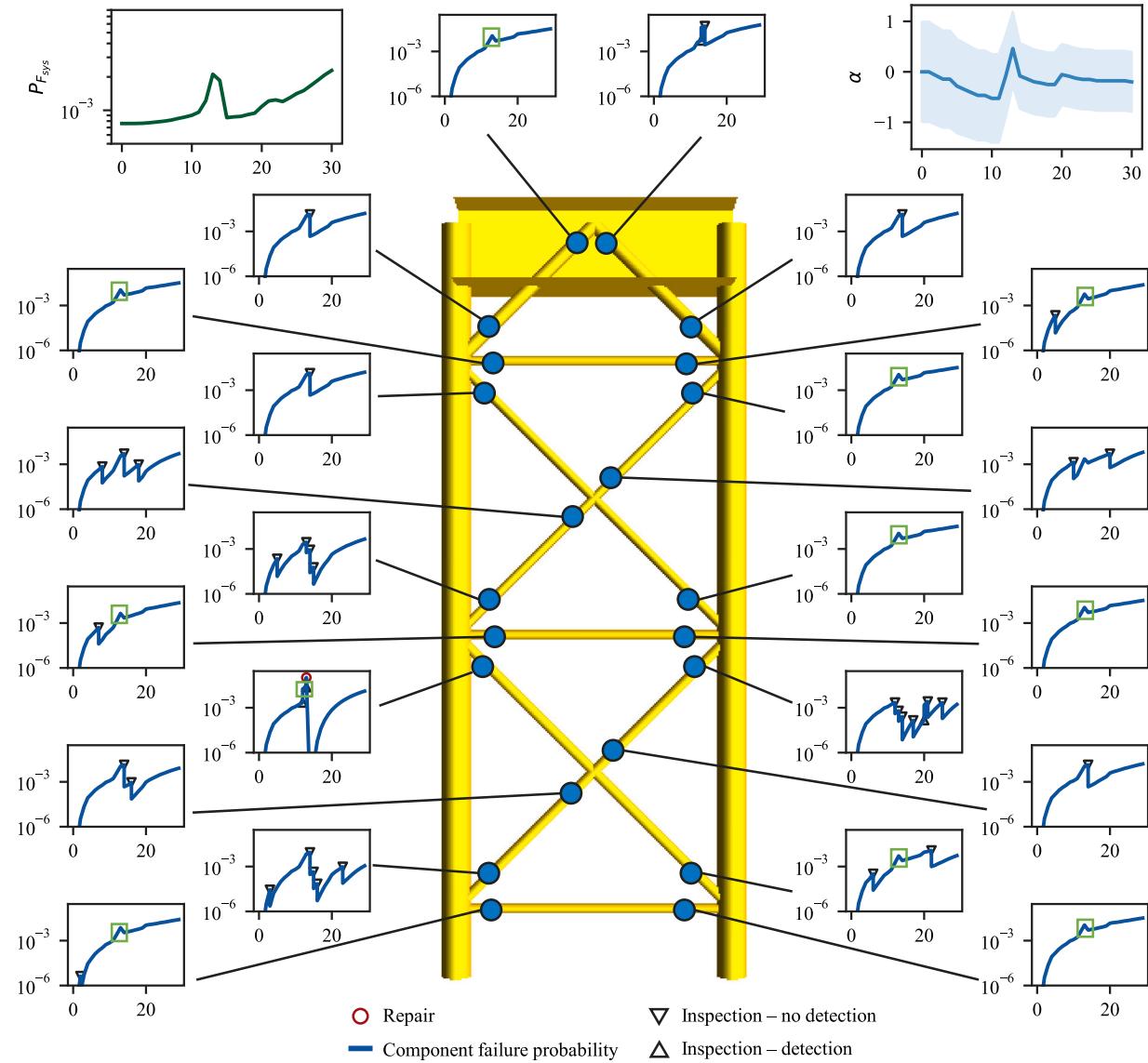


Fig. 8. Zayas frame sample policy realization in an equal deterioration correlation environment ($\rho_{eq} = 0.4$). The failure probability of each hotspot is depicted by a blue line, inspection indications are represented by upwards (detection) or downwards (no-detection) triangles, and perfect repairs are circled in red. At the system level, the failure probability, at the top-left corner, and system-effects, as these are explained in the main text, are represented by a green line and squares, respectively. The evolution of the hyperparameters over time is plotted in a light blue diagram, at the top-right corner.

System failure probability

Offshore structures are exposed to fatigue and corrosion deterioration due to the combined cyclic effect of waves and wind in a harsh marine environment. Initial defects at geometric discontinuities or at welded regions (hotspots) grow over time, becoming critical if maintenance actions are not timely undertaken. In this study, and following the experiments conducted in [36,55], a total of 22 hotspots are considered, located at the joints at the braces or columns. Each brace is associated with either one or two hotspots, as illustrated in Fig. 8, at critical locations for fatigue deterioration. The fatigue deterioration is assumed similar for all hotspots, modeled by the same deterioration process as for the 9-out-of-10 structural system (Section 5.1).

The failure of the system is defined here as the incapacity of the frame to withstand the concentrated horizontal load applied at the upper-left corner. At the component level, the health of each hotspot is described by the vector \mathbf{F}_h , in which $F_h = 0$ indicating a hotspot failure and $F_h = 1$ corresponding to a

hotspot survival. The failure probability of a hotspot $p_F^{(h)}$ corresponds thus to the probability of being in state $F_h = 0$. At an element level, the state of each brace is represented by a vector \mathbf{x}_{el} , considering $x_{el} = 0$ if the element has failed and $x_{el} = 1$ otherwise. Assuming that a brace fails if any of its associated hotspots fail, the failure probability of an element $p_F^{(el)}$, i.e., $Pr[x_{el} = 0]$, can be therefore computed as a series system:

$$p_F^{(el)} = 1 - \prod_{h \in N_h} \left[1 - p_F^{(h)} \right] \quad (21)$$

At the system scale, the health of the frame depends on the state of all its constitutive elements, i.e., 13 braces, and the failure probability of the system $p_{F_{sys}}$ is computed herein as a function of all the element state combinations. A total of 8,192 ($= 2^{13}$) non-linear static push-over simulations have been run with the assistance of the computer code 'USFOS' (available within the software package Sesam) [56], before the training of DDMAC, so that the failure probability of the system conditional on all element state combinations is explicitly and directly defined. The element configuration for each push-over simulation is

arranged according to the element state vector \mathbf{x}_{el} , removing the braces associated with a failed state $x_{el} = 0$. The resistance $L_{col}(\mathbf{x}_{el})$ of each element state combination cases is retrieved from the conducted push-over simulations.

The collapse event of the frame is defined as the probability of the external horizontal load exceeding the structural system resistance $Pr[L > L_{col}]$. In this case, the horizontal load is modeled as a lognormal random variable with mean $\mu_L = 70$ kN and 25% coefficient of variation, while no uncertainty is associated with the resistance, a reasonable assumption when the external load is highly uncertain in comparison with the resistance [55]. The failure probability of the system, p_{Fsys} , conditional on the element state vector \mathbf{x}_{el} can be then defined directly from the probability density function of the load, f_L :

$$p_{Fsys} = \int_{L_{col}}^{\infty} f_L(x) dx \quad (22)$$

In the undamaged case, i.e., no elements are removed from the original configuration, the collapse load is 247 kN, resulting in a failure

probability of approximately 10^{-4} . The state of the frame is, however, computed conditional on the state of all the elements, and to do so the probability of being in each state combination should be computed. We follow the iterative procedure proposed in [41] to compute the probability of being in each element state $\mathbf{q} \doteq p(\mathbf{x}_{el})$ as a function of the element failure probability $p_F^{(el)}$ and the element survival probability $\bar{p}_F^{(el)}$:

$$\begin{aligned} \mathbf{q}_{[1]} &= \left[p_F^{(1)} \quad \bar{p}_F^{(1)} \right]^T \\ \mathbf{q}_{[i]} &= \left[\mathbf{q}_{[i-1]} \cdot p_F^{(i)} \right] \\ &= \left[\mathbf{q}_{[i-1]} \cdot \bar{p}_F^{(i)} \right] \end{aligned} \quad (23)$$

Finally, the system failure probability p_{Fsys} is equal to the system failure probability conditional on the element state $p_{Fsys}^{(\mathbf{x}_{el})}$ multiplied by the probability of being in that state $q^{(\mathbf{x}_{el})}$:

$$p_{Fsys} = \sum_{\mathbf{x}_{el} \in \mathcal{X}_{el}} \left[p_{Fsys}^{(\mathbf{x}_{el})} \cdot q^{(\mathbf{x}_{el})} \right] \quad (24)$$

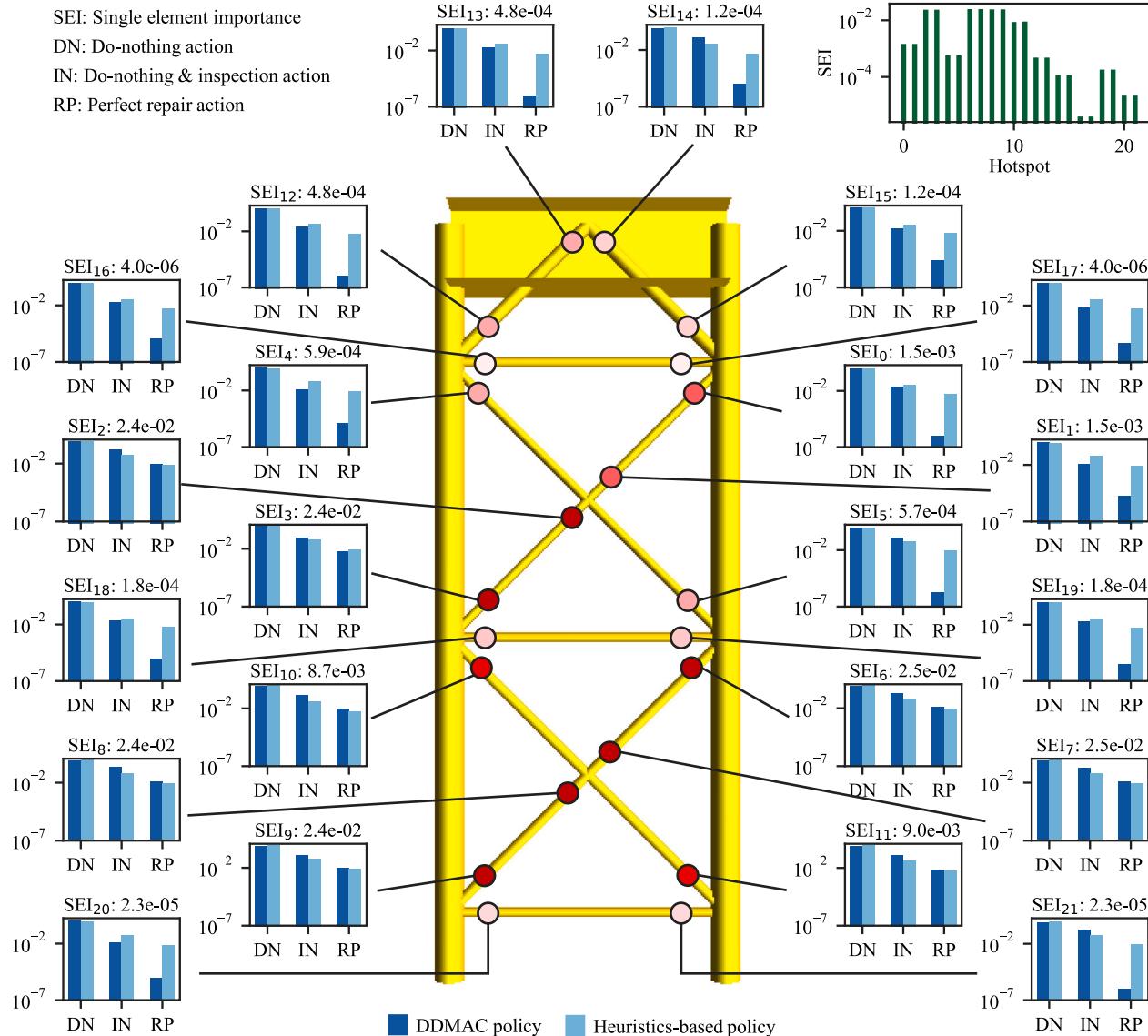


Fig. 9. Histograms of DDMAC and heuristics-based policy actions of 30,000 realizations of this Zayas frame setting for a 30-year policy period. The Single Element Importance metric (SEI) associated with each fatigue hotspot is indicated at the top of each histogram and summarized at the green top-right diagram. The relative importance of each hotspot is also represented by color, with a darker red being a more critical element for the structural reliability of the system.

Results and discussion

The comparison between DDMAC and optimized heuristics follows the same trend as that of the 9-out-10 structural system experiments. In terms of expected life-cycle costs, DDMAC policies outperform heuristic-based policies in the two tested settings, as shown in Fig. 7, with cost savings ranging from 20.1% to 22.8%. A slight decrease in the expected life-cycle costs can also be observed for the case under deterioration correlation, as a result of the reduction of failure risk. This is expected, since under deterioration dependence, i.e., the initial crack size among the hotspots is correlated, an observation collected at one hotspot also provides information to other hotspots, accordingly updating their damage state belief.

As seen in Fig. 8, where the evolution of components' and system failure probabilities are illustrated over time, beliefs are updated for both detection and no detection observation outcomes. At year 12, a crack is detected at the lower X-brace, and this observation shows up as a failure probability update for the other components, marked with a green rectangle in the plots, an effect that can also be observed clearly in the updated mean of the hyperparameter, α . The effect of deterioration dependencies among components on the resulting system failure probability can also be visualized in Fig. 8, e.g., a crack detection at hotspot 6 ultimately induces a kink in the system failure probability around year 20. Essentially, DDMAC is able to discover the importance of each hotspot regarding the structural reliability of the frame. To explore this system effect, the Single Element Importance (SEI) measure is calculated for each hotspot. The concept of SEI, as defined in [57], determines the importance of each element to the system structural reliability by subtracting the undamaged system failure probability $p_{F_{sys}}$ from the system failure probability with the element removed ($\sim el$). In this case, and since each element is defined as a series system of hotspots, the SEI can be directly computed for each hotspot h , determining each hotspot importance as:

$$SEI_h = p_{F_{sys}}^{(\sim h)} - p_{F_{sys}} \quad (25)$$

The SEI of a vital element for the structural system is thus higher than the SEI of a less important component. The structural element importance (SEI) of each hotspot is shown in Fig. 9, along with histograms of the actions taken at each component during 30,000 DDMAC (dark blue) and optimized-heuristic (light blue) realizations. As represented by the dark green bar diagram at the top-right corner of Fig. 9, and in agreement with the findings reported in [55], the critical hotspots are located at the X-braces, whereas the less critical hotspots are the ones connecting the horizontal braces.

While the *do-nothing* action is dominant and *inspection* actions are distributed similarly among components, the distribution of *repair* actions among hotspots differs for DDMAC and heuristics-based decision rules. DDMAC plans repairs mainly for important elements with respect to the global structural reliability, i.e., with a high SEI, such as hotspots 6 and 7, whereas less important components for the system structural reliability are less frequently repaired. In contrast, the heuristic-based policy plans component repairs nearly evenly, disregarding the influence of each hotspot to the reliability of the system. We can therefore conclude that DDMAC policies are able to inherently identify the system effects attributed to the structural and reliability importance of each element for the entire system.

6. Concluding remarks

This paper introduces an efficient algorithmic framework for inference and optimal decision-making under uncertainty for engineering systems exposed to deteriorating environments. In terms of inference, a Gaussian hierarchical structure is presented, within a dynamic Bayesian network model, and further formalized here with the objective of enabling the treatment of engineering systems under general, unequal

deterioration correlation settings, considering also the effect of maintenance actions. The proposed efficient inference framework is then seamlessly integrated with principled optimization methods by formulating the decision-making problem as a factored Partially Observable Markov Decision Process (POMDP), with its dynamics encoded by Bayesian network conditional structures. The system life-cycle realizations collected by simulating the specified POMDP are employed by a multi-agent actor-critic deep reinforcement learning algorithm, able to identify optimal strategies in very high dimensional state, action, and observation spaces, commonly found in practical structural and engineering systems. In particular, we demonstrate through numerical experiments that the proposed approach provides efficient inspection and maintenance (I&M) policies, outperforming state-of-the-art policies, and enables a systematic consideration of system effects, that is autonomously and intrinsically reflected in the identified strategies.

POMDP-based policies, parameterized in high-dimensional settings through the Deep Decentralized Multi-agent Actor Critic (DDMAC) algorithm, map the current belief state of the system to a probability distribution of possible actions. These stochastic policies are thus prescribed as a function of the belief state, which is a sufficient statistic of the history of actions and observations. Constructing the policies based on a sufficient statistic enables more effective optimal decision-making strategies than static optimization approaches, which are constrained by the limited space explored during the policy search. POMDP-based policies also provide additional flexibility to the decision maker, who can opt for an alternative decision at some point, for any reason, and the policy through the updated belief state will be automatically adapted thereafter, yielding near-optimal results.

DDMAC policies are approximated by actor neural networks, whose weights are learned according to noisy rewards collected at the system level. By including deterioration dependence among components in the simulated environment, and by formulating the cost model at the system level, DDMAC policies are able to intrinsically consider the following system effects:

- In deterioration dependent environments, observing the state of one component provides indirect information to the other components of the system, modulated by their degree of correlation. In the tested I&M planning scenarios, environments with higher correlation resulted in a reduction of expected costs, usually characterized by lower expected failure risk. As structural systems are designed according to high reliability standards, demanding a low failure risk, observations mostly indicate sound structural states, which in highly correlated environments results in a global reduction of failure risk. As opposed to independent deterioration settings, higher variability in the expected costs is observed in dependent environments, in which very different I&M policy scenarios can be experienced based on the acquired observations.
- A clustering effect on inspections and repairs is observed in settings that include a campaign cost model, i.e., a fixed cost is activated if at least one component is repaired or inspected. In this case, policies seek to avoid planning single or few inspections and repairs at one time step. Instead, inspection and maintenance actions are generally grouped, saving the additional campaign cost associated with inspecting and repairing only few components within one campaign.
- Maintenance actions are influenced by the relative importance of the components to the system structural reliability. As observed in the steel frame application, repairs were mostly allocated to critical elements, whereas components less important to the global reliability were less often repaired.

In this work, the proposed factored POMDP approach importantly enables closed-form Bayesian inference on the discrete state space, developed by properly discretizing the involved continuous random variables, to ensure accurate computation of system reliability metrics. Yet, the dimensionality of this resulting discrete space might

significantly increase computational efforts in certain settings and applications. To allow for additional state/action space dimensionality reductions in such cases, further future research on the development of continuous state POMDPs is encouraged.

CRediT authorship contribution statement

P.G. Morato: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Investigation, Data curation, Writing – original draft, Writing – review & editing, Funding acquisition. **C.P. Andriots:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – review & editing. **K.G. Papakonstantinou:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Supervision, Writing – review & editing, Funding acquisition. **P. Rigo:** Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research is funded by the National Fund for Scientific Research in Belgium F.R.I.A. - F.N.R.S. This support is gratefully acknowledged. Dr. Papakonstantinou would further like to acknowledge that this material is also based upon work supported by the U.S. National Science Foundation under Grants No. 1751941 and 2053620. Dr. Andriots would like to acknowledge the support of the TU Delft AI Labs program. Finally, the authors would like to acknowledge the support provided by DNV GL Digital Solutions for granting access to the software package 'Sesam'.

Appendix A. Optimized heuristic decision rules

Table A.1 lists the optimized heuristic decision rules used in all presented numerical experiments.

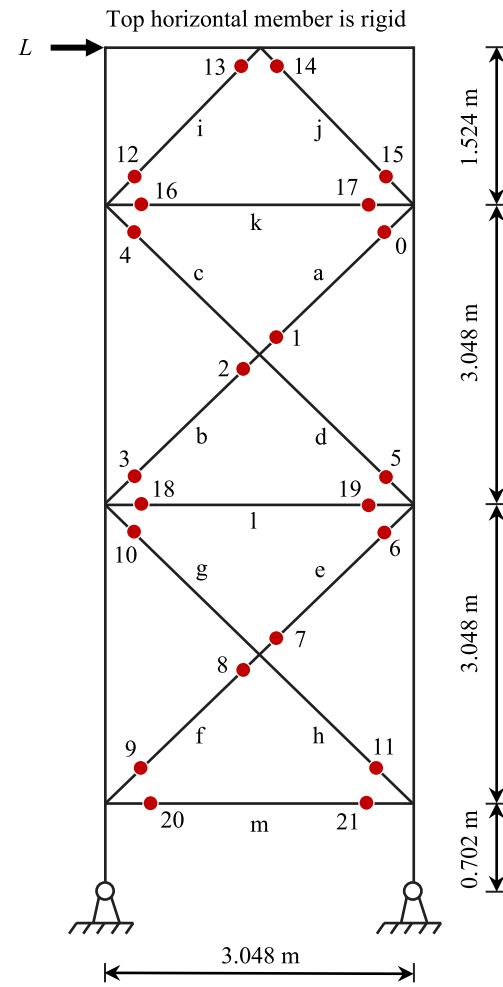
Table A.1

List of optimized heuristic decision rules employed in the numerical experiments. For each considered setting, the resulting heuristic decision rules dictate inspections for n_c components at equidistant intervals of Δ_{ins} years. ρ indicates equal (eq) or unequal (uq) deterioration correlation among components.

Setting	Deterioration correlation	Cost model	Δ_{ins}	n_c
9-out-of-10 system	$\rho_{eq} = 0$	Individual	6	10
9-out-of-10 system	$\rho_{eq} = 0.4$	Individual	6	10
9-out-of-10 system	$\rho_{eq} = 0.8$	Individual	6	8
9-out-of-10 system	ρ_{uq}	Individual	5	7
9-out-of-10 system	$\rho_{eq} = 0$	Campaign	5	10
9-out-of-10 system	$\rho_{eq} = 0.4$	Campaign	6	10
9-out-of-10 system	$\rho_{eq} = 0.8$	Campaign	5	7
9-out-of-10 system	ρ_{uq}	Campaign	8	10
Zayas frame	$\rho_{eq} = 0$	Individual	10	16
Zayas frame	$\rho_{eq} = 0.4$	Individual	10	16

Appendix B. Zayas frame geometry and material properties

Fig. B.1 displays characteristics of the Zayas frame used in this paper.



Section properties

Elements a-d & k-m: OD 102 mm; WT 2 mm
 Elements e-h: OD 127 mm; WT 3 mm
 Elements i & j: OD 152 mm; WT 3.2 mm
 Legs: OD 324 mm; WT 7.1 mm

Yield strength

Elements a-m: 284 N/mm²
 Legs: 324 N/mm²

Fig. B.1. Zayas frame representation [51,55]. Elements are denoted with lower-case letters and fatigue hotspots are designated with numbers. The outer diameters, OD, wall thicknesses, WT, and mechanical properties corresponding to each element of the frame are specified in the diagram. An external horizontal load, L , is applied at the upper-left corner of the frame.

References

- [1] Faber MH, Stewart MG. Risk assessment for civil engineering facilities: Critical overview and discussion. Reliab Eng Syst Saf 2003;80(2):173–84. [http://dx.doi.org/10.1016/S0951-8320\(03\)00027-9](http://dx.doi.org/10.1016/S0951-8320(03)00027-9).
- [2] Frangopol DM, Soliman M. Life-cycle of structural systems: recent achievements and future directions. Struct Infract Eng 2016;12(1):1–20. <http://dx.doi.org/10.1080/15732479.2014.999794>.
- [3] Straub D. Generic approaches to risk based inspection planning for steel structures [Ph.D. thesis], Swiss Federal Institute of Technology Zürich (ETH); 2004, p. 247.
- [4] Straub D. Stochastic modeling of deterioration processes through dynamic Bayesian networks. J Eng Mech 2009;135(10):1089–99. [http://dx.doi.org/10.1061/\(asce\)jem.1943-7889.0000024](http://dx.doi.org/10.1061/(asce)jem.1943-7889.0000024).

[5] Papakonstantinou KG, Shinozuka M. Probabilistic model for steel corrosion in reinforced concrete structures of large dimensions considering crack effects. *Eng Struct* 2013;57:306–26. <http://dx.doi.org/10.1016/j.engstruct.2013.06.038>.

[6] Yang DY, Frangopol DM. Probabilistic optimization framework for inspection/repair planning of fatigue-critical details using dynamic Bayesian networks. *Comput Struct* 2018;198:40–50. <http://dx.doi.org/10.1016/j.compstruc.2018.01.006>.

[7] Nielsen JS, Sørensen JD. Computational framework for risk-based planning of inspections, maintenance and condition monitoring using discrete Bayesian networks. *Struct Infract Eng* 2018;14(8):1082–94. <http://dx.doi.org/10.1080/15732479.2017.1387155>.

[8] Long L, Mai QA, Morato PG, Sørensen JD, Thöns S. Information value-based optimization of structural and environmental monitoring for offshore wind turbines support structures. *Renew Energy* 2020;159:1036–46. <http://dx.doi.org/10.1016/j.renene.2020.06.038>.

[9] Morato PG, Papakonstantinou KG, Andriots CP, Nielsen JS, Rigo P. Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes. *Struct Saf* 2022;94:102140. <http://dx.doi.org/10.1016/j.strusafe.2021.102140>.

[10] Hlaing N, Morato PG, Nielsen JS, Amirafshari P, Kolios A, Rigo P. Inspection and maintenance planning for offshore wind structural components: integrating fatigue failure criteria with Bayesian networks and Markov decision processes. *Struct Infract Eng* 2022;18(7):983–1001. <http://dx.doi.org/10.1080/15732479.2022.2037667>.

[11] Bismut E, Straub D. Optimal adaptive inspection and maintenance planning for deteriorating structural systems. *Reliab Eng Syst Saf* 2021;215:107891. <http://dx.doi.org/10.1016/j.ress.2021.107891>.

[12] Corotis RB, Hugh Ellis J, Jiang M. Modeling of risk-based inspection, maintenance and life-cycle cost with partially observable Markov decision processes. *Struct Infract Eng* 2005;1(1):75–84. <http://dx.doi.org/10.1080/15732470412331289305>.

[13] Papakonstantinou KG, Shinozuka M. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory. *Reliab Eng Syst Saf* 2014;130:202–13. <http://dx.doi.org/10.1016/j.ress.2014.04.005>.

[14] Papakonstantinou KG, Shinozuka M. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation. *Reliab Eng Syst Saf* 2014;130:214–24. <http://dx.doi.org/10.1016/j.ress.2014.04.006>.

[15] Papakonstantinou KG, Shinozuka M. Optimum inspection and maintenance policies for corroded structures using partially observable Markov decision processes and stochastic, physically based models. *Probab Eng Mech* 2014;37:93–108. <http://dx.doi.org/10.1016/j.probengmech.2014.06.002>.

[16] Andriots CP, Papakonstantinou KG, Chatzi EN. Value of structural health information in partially observable stochastic environments. *Struct Saf* 2021;93:102072. <http://dx.doi.org/10.1016/j.strusafe.2020.102072>.

[17] Memarzadeh M, Pozzi M, Kolter JZ. Hierarchical modeling of systems with similar components: A framework for adaptive monitoring and control. *Reliab Eng Syst Saf* 2016;153:159–69. <http://dx.doi.org/10.1016/j.ress.2016.04.016>.

[18] Memarzadeh M, Pozzi M, Zico Kolter J. Optimal planning and learning in uncertain environments for the management of wind farms. *J Comput Civ Eng* 2015;29(5):04014076. [http://dx.doi.org/10.1061/\(ASCE\)CP.1943-5487.0000390](http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000390).

[19] Thoft-Christensen P, Dalsgård Sørensen J. Reliability of structural systems with correlated elements. *Appl Math Model* 1982;6(3):171–8. [http://dx.doi.org/10.1016/0307-904X\(82\)90006-3](http://dx.doi.org/10.1016/0307-904X(82)90006-3).

[20] Ito S, Deodatis G, Fujimoto Y, Asada H, Shinozuka M. Non-periodic inspection by Bayesian method II: structures with elements subjected to different stress levels. *Probab Eng Mech* 1992;7(4):205–15. [http://dx.doi.org/10.1016/0266-8920\(92\)90024-C](http://dx.doi.org/10.1016/0266-8920(92)90024-C).

[21] Straub D, Papaioannou I. Bayesian updating with structural reliability methods. *J Eng Mech* 2015;141(3):34. [http://dx.doi.org/10.1061/\(ASCE\)EM.1943-7889.0000839](http://dx.doi.org/10.1061/(ASCE)EM.1943-7889.0000839).

[22] Luque J, Straub D. Risk-based optimal inspection strategies for structural systems using dynamic Bayesian networks. *Struct Saf* 2019;76:68–80. <http://dx.doi.org/10.1016/j.strusafe.2018.08.002>, URL <https://doi.org/10.1016/j.strusafe.2018.08.002>.

[23] Andriots CP, Papakonstantinou KG. Managing engineering systems with large state and action spaces through deep reinforcement learning. *Reliab Eng Syst Saf* 2019;191:106483. <http://dx.doi.org/10.1016/j.ress.2019.04.036>.

[24] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;529(7587):484–9. <http://dx.doi.org/10.1038/nature16961>.

[25] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. *Nature* 2017;550(7676):354–9. <http://dx.doi.org/10.1038/nature24270>.

[26] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. 2013, arXiv preprint [arXiv:1312.5602](http://arxiv.org/abs/1312.5602).

[27] Andriots CP, Papakonstantinou KG. Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. *Reliab Eng Syst Saf* 2021;212:107551. <http://dx.doi.org/10.1016/j.ress.2021.107551>.

[28] Wei S, Bao Y, Li H. Optimal policy for structure maintenance: A deep reinforcement learning framework. *Struct Saf* 2020;83:101906. <http://dx.doi.org/10.1016/j.strusafe.2019.101906>.

[29] Zhang N, Si W. Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. *Reliab Eng Syst Saf* 2020;203:107094. <http://dx.doi.org/10.1016/j.ress.2020.107094>.

[30] Nguyen V-T, Do P, Vosin A, Iung B. Artificial-intelligence-based maintenance decision-making and optimization for multi-state component systems. *Reliab Eng Syst Saf* 2022;228:108757. <http://dx.doi.org/10.1016/j.ress.2022.108757>.

[31] Fan L, Su H, Wang W, Zio E, Zhang L, Yang Z, et al. A systematic method for the optimization of gas supply reliability in natural gas pipeline network based on Bayesian networks and deep reinforcement learning. *Reliab Eng Syst Saf* 2022;225:108613. <http://dx.doi.org/10.1016/j.ress.2022.108613>.

[32] Zhou Y, Li B, Lin TR. Maintenance optimisation of multicomponent systems using hierarchical coordinated reinforcement learning. *Reliab Eng Syst Saf* 2022;217:108078. <http://dx.doi.org/10.1016/j.ress.2021.108078>.

[33] Mohammadi R, He Q. A deep reinforcement learning approach for rail renewal and maintenance planning. *Reliab Eng Syst Saf* 2022;108615. <http://dx.doi.org/10.1016/j.ress.2022.108615>.

[34] Yang DY. Deep reinforcement learning-enabled bridge management considering asset and network risks. *J Infract Syst* 2022;28(3):04022023. [http://dx.doi.org/10.1061/\(ASCE\)IS.1943-555X.0000704](http://dx.doi.org/10.1061/(ASCE)IS.1943-555X.0000704).

[35] Saifullah M, Andriots CP, Papakonstantinou KG, Stoffels SM. Deep reinforcement learning-based life-cycle management of deteriorating transportation systems. In: *Bridge safety, maintenance, management, life-cycle, resilience and sustainability*. CRC Press; 2022, p. 293–301.

[36] Luque J, Straub D. Reliability analysis and updating of deteriorating systems with dynamic Bayesian networks. *Struct Saf* 2016;62:34–46. <http://dx.doi.org/10.1016/j.strusafe.2016.03.004>.

[37] Kurniawati H, Hsu D, Sun Lee W. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In: *Proceedings of robotics: Science and systems IV*. Zurich, Switzerland, 2008, <http://dx.doi.org/10.15607/RSS.2008.IV.009>.

[38] Murphy KP. *Dynamic Bayesian networks: Representation, inference and learning* [Ph.D. thesis], Berkeley: University of California; 2002.

[39] Papakonstantinou KG, Amir M, Warn GP. A scaled spherical simplex filter (S3F) with a decreased $n+2$ sigma points set size and equivalent $2n+1$ Unscented Kalman Filter (UKF) accuracy. *Mech Syst Signal Process* 2022;163:107433. <http://dx.doi.org/10.1016/j.ymssp.2020.107433>.

[40] Dunnett CW, Sobel M. Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika* 1955;42:258–60. <http://dx.doi.org/10.2307/2333441>.

[41] Song J, Kang W-H. System reliability and sensitivity under statistical dependence by matrix-based system reliability method. *Struct Saf* 2009;31(2):148–56. <http://dx.doi.org/10.1016/j.strusafe.2008.06.012>.

[42] Der Kiureghian A. *Structural and system reliability*. Cambridge University Press; 2022, <http://dx.doi.org/10.1017/9781108991889>.

[43] Pineau J, Gordon G, Thrun S. Point-based value iteration: An anytime algorithm for POMDPs. In: *IJCAI international joint conference on artificial intelligence*. 2003, p. 1025–30.

[44] Papakonstantinou KG, Andriots CP, Shinozuka M. POMDP and MOMDP solutions for structural life-cycle cost minimization under partial and mixed observability. *Struct Infract Eng* 2018;14(7):869–82. <http://dx.doi.org/10.1080/15732479.2018.1439973>.

[45] Puterman ML. *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons; 2014.

[46] Spaan MTJ, Vlassis N. Perseus: Randomized point-based value iteration for POMDPs. *J Artificial Intelligence Res* 2005;24:195–220. <http://dx.doi.org/10.1613/jair.1659>.

[47] Smith T, Simmons R. Focused real-time dynamic programming for MDPs: Squeezing more out of a heuristic. *Proc Natl Conf Artif Intell* 2006;2(January):1227–32.

[48] Sutton RS, Barto AG. *Reinforcement learning: An introduction*. MIT Press; 2018.

[49] Li Y. Deep reinforcement learning: An overview. 2017, arXiv preprint [arXiv:1701.07274](http://arxiv.org/abs/1701.07274).

[50] Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. 2015, arXiv preprint [arXiv:1511.05952](http://arxiv.org/abs/1511.05952).

[51] Popov EP, Zayas VA, Mahin SA. Inelastic cyclic behavior of tubular braced frames. *J Struct Div* 1980;106(12):2375–90. <http://dx.doi.org/10.1061/JSDEAG.0005590>.

[52] Ditlevsen O, Madsen HO. Structural reliability methods. Department of Mechanical Engineering. Technical University of Denmark; 2007, p. 361.

[53] Barlow RE, Heidtmann KD. Computing k-out-of-n system reliability. *IEEE Trans Reliab* 1984;33(4):322–3.

[54] Moan T, Amdahl J, Granli T, Hellan O. Collapse behaviour of offshore structural systems. *Adv Mar Struct-2* 1991;469–94.

[55] Schneider R, Thöns S, Straub D. Reliability analysis and updating of deteriorating systems with subset simulation. *Struct Saf* 2017;64:20–36. <http://dx.doi.org/10.1016/j.strusafe.2016.09.002>.

[56] Søreide T, Amdahl J, Eberg E, Holmås T, Hellan Ø. USFOS—A computer program for progressive collapse analysis of steel offshore structures. In: Theory manual. Trondheim, Norway: SINTEF; 1993.

[57] Straub D, Der Kiureghian A. Reliability acceptance criteria for deteriorating elements of structural systems. *J Struct Eng* 2011;137(12):1573–82. [http://dx.doi.org/10.1061/\(ASCE\)ST.1943-541X.0000425](http://dx.doi.org/10.1061/(ASCE)ST.1943-541X.0000425).