Optimal Inspection and Maintenance Planning for Deteriorating Structures via Markov Decision Processes and Deep Reinforcement Learning

Application to Offshore Wind Substructures

Pablo G. Morato

Supervisor: Prof. Philippe Rigo

Co-supervisor: Prof. Konstantinos G. Papakonstantinou

Applied Sciences
University of Liège

This dissertation is submitted for the degree of
Doctor of Philosophy in Engineering Sciences

August 2021

PhD Committee

Maarten Arnst (chairman)
Professor
University of Liège, Belgium

Philippe Rigo (supervisor)
Professor
University of Liège, Belgium

Konstantinos G. Papakonstantinou (co-supervisor)
Professor
The Pennsylvania State University, USA

Vincent Denoël
Professor
University of Liège, Belgium

Damien Ernst
Professor
University of Liège, Belgium

Dan M. Frangopol
Professor
Lehigh University, USA

Charalampos P. Andriotis
Professor
Delft University of Technology, Netherlands

Jannie S. Nielsen
Professor
Aalborg University, Denmark

I would like to dedicate this thesis to Maribel, who cleverly inspired me to always remain curious, seeking answers in a sea of endless questions.

Acknowledgements

Throughout this PhD journey, I have received the support and assistance from many individuals.

I would first like to thank my supervisor, Prof. Philippe Rigo, for his guidance and encouragement along this investigation and for sharing with me his broad research experience. His constant support is greatly appreciated.

This work could not have been possible without the feedback and constructive comments provided by my co-supervisor Konstantinos G. Papakonstantinou (The Pennsylvania State University) and Prof. Charalampos P. Andriotis (Delft University of Technology). Our fruitful conversations, along with their expert scientific advice, have notably contributed and inspired this research.

I gratefully recognize the knowledge transfer and scientific suggestions I have acquired from Prof. Jannie S. Nielsen (Aalborg University), Prof. John D. Sørensen (Aalborg University), Prof. Maarten Arnst (University of Liège), Prof. Daniel Straub (Technical University of Munich), Prof. Michael H. Faber (Aalborg University) and Prof. Sebastian Thöns (Lund University). Their talks and inspirational lessons have widened my interest in the application of decision-making methods for managing deteriorating structural systems.

I would also like to thank my colleagues at ANAST for supplying me with doses of moral and scientific support during the development of this work. I acknowledge, especially, the help that Ms. Nandar Hlaing has selflessly donated me during the final years of this PhD. Also, Mr. Lucas Marquez, Mr. Felipe Giro, and Mr. Radomir Jasic should take the credit not only for their technical assistance, but for inducing a comfortable and amicable atmosphere at the office. In a more personal note, I am also very grateful to my family for illuminating the dark passages of these four years.

Finally, I would like to acknowledge the financial support granted by the National Fund of Scientific Research (FRNS).

Abstract

Safeguarding engineering infrastructures in a healthy condition is of paramount importance to sustain the economic and societal growth of most countries. Deterioration mechanisms and mechanical stressors have a detrimental effect on structural performance, inducing a risk of failure that might lead, in some cases, to considerable economic, societal, and environmental consequences. Although the estimation of deterioration processes is associated with significant uncertainties, information from inspections and monitoring can be collected, at a cost, in order to dictate more informed maintenance decisions. Inspection and Maintenance (I&M) planning thus demands methods capable of identifying optimal management strategies in stochastic environments and under imperfect information. Addressing the aforementioned needs, this thesis is devoted to the exploration of efficient methods with the objective of controlling the risk of adverse events by timely planning inspections and optimally dictating maintenance actions. Throughout the investigation, the I&M planning decision-making problem is formally formulated as a Partially Observable Markov Decision Process (POMDP), constituting the underlying principled mathematical foundation of the stochastic control optimization. From medium to high-dimensional state space settings, infinite and finite horizon policies are efficiently computed via POMDP point-based solvers, whereas for higher dimensional state, action, and observation space settings, POMDPs are integrated with a multi-agent actor critic deep reinforcement learning approach. Besides overcoming dimensionality limitations by approximating both policy and value functions with artificial neural networks, the formulation of the POMDPs through conditional formations enables the treatment of structural systems under deterioration, reliability, and cost dependence. Sequential monitoring decisions, influenced by the condition of the sensors, can also be aptly allotted through the proposed approach. Extensive numerical experiments have been conducted for both traditional and detailed I&M planning settings with a strong emphasis on offshore wind substructures, thoroughly comparing POMDP policies against corrective, calendar, and heuristic-based strategies. The results reveal that POMDP-based policies offer substantial savings compared to their counterparts in all the tested settings.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Rationale and motivation of the research

Civil and maritime engineering structures are exposed to mechanical and environmental stressors throughout their operational lifetime. Such deterioration mechanisms impose a detrimental effect on the infrastructures condition, reducing their structural resistance and inducing a failure risk, which is turn associated with economical, societal, and environmental consequences. Fatigue deterioration is, for instance, experienced by bridge structures, due to the exerted cyclic traffic loads, or by offshore wind substructures, as a result of the combined action of wind and wave loads. Along with fatigue, engineering structures are also subjected to corrosion deterioration, leading not only to a thickness reduction of structural members, but also accelerating the growth of existing fatigue cracks.

In contrast to reliability assessments of mechanical or electronic components, statistics of structural failures are very scarce, as civil and maritime structures are robustly designed due to the huge consequences associated with a failure event. And even if statistics could be gathered for a particular structure, the collected data will most likely not be applicable to other similar structure, inherently different in its design. Structural reliability thus resorts to analytical and/or numerical models. The estimation of deterioration processes through engineering models is, however, associated with significant uncertainties. For instance, Miner's cumulative damage law, often adopted for fatigue assessments at the design stage, provides fatigue damage estimates with a 30 % coefficient of variation. Loading predictions also contain uncertainties, especially when long-term estimations are involved, as for the computation of offshore structures fatigue loading.

Within a Bayesian probabilistic framework, both aleatory uncertainties arising from the intrinsic randomness of natural phenomena, and epistemic uncertainties introduced by the limited information available, can be jointly quantified, and the latter can be reduced based on observed evidence. In practice, information collected through inspections and/or monitoring can be utilized to reassess the condition of the structural components, enabling

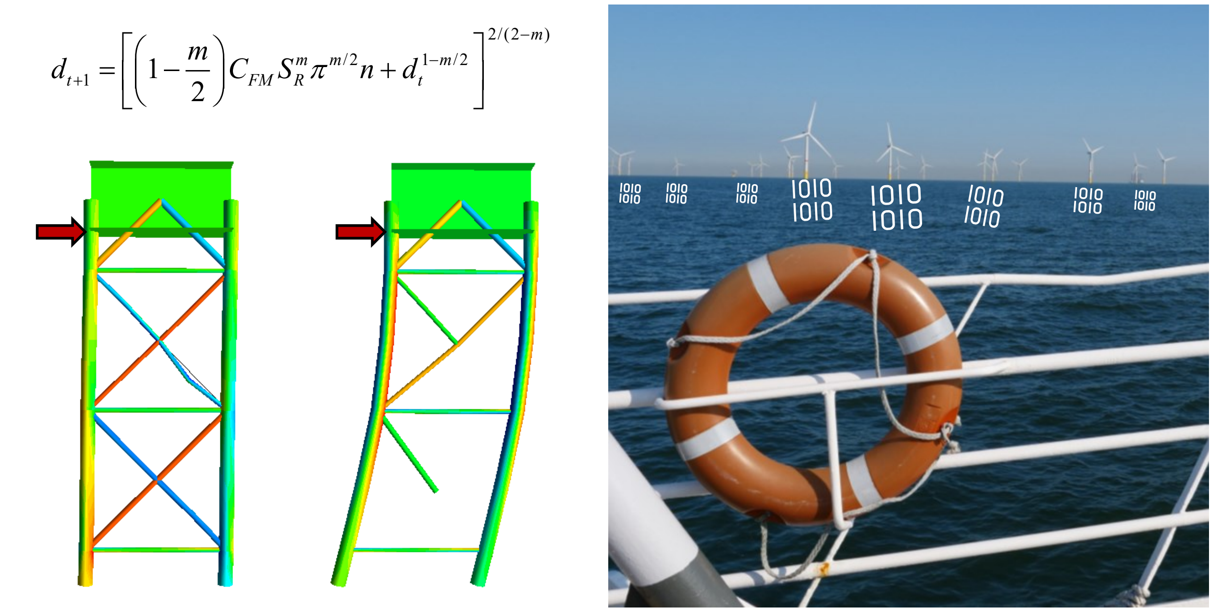


Fig. 1.1. Integration of engineering models with operational data.

operators to select more informed and optimal decisions, e.g. timely scheduling maintenance interventions. Fiber-optic load cells, drones, and remotely operated underwater vehicles are examples of the increasingly available modern techniques for retrieving operational data. In certain applications, a vast amount of data can be acquired, e.g. wind turbine operational data controlled by a SCADA system. Inspections and monitoring activities are, however, charged with costs that vary depending on the structure, accessibility, and measurement quality. An eddy current non-destructive experiment yields, for instance, more accurate crack detection indications than visual inspections, albeit at a probably higher expense.

In some cases, the management of infrastructures rely on calendar-based inspection policies and maintenance interventions dictated according to the most recent collected observation. More rational methods, such as risk-based inspection planning, schedule observations upon exceeding a predefined failure probability threshold or other prescribed heuristic decision rules, yet the resulting strategies are limited by the number of evaluated heuristic decision rules out of an immense policy space. Recent societal concerns on sustainability and stricter environmental regulations additionally demand management strategies able to provide an optimal allocation of the available resources. This is reflected, for instance, on the sustainable development goals proposed by the United Nations, in which economic growth, reliable and clean infrastructure, clean energy and sustainable communities are within the main focus.

In summary, there is a need for efficient inspection and maintenance planning methods to rationally manage engineering structures, controlling economic, societal, and environmental

Fig. 1.2. Updating of epistemic uncertainties throughout the life-cycle of engineering structures.

risks by optimally scheduling inspections and maintenance interventions. In principle, the devised inspection and maintenance strategies should be able to answer the following questions:

- What is the target of the inspections? Crack size in a weld under the action of fatigue or thickness reduction due to corrosion.
- What type of repairs should be conducted? A minor repair consisting in grinding a weld, a major welding interventions or a repainting of an area affected by corrosion.
- What inspection technique should be employed? A thorough eddy current non-destructive examination or a visual inspection.
- When inspection and repairs should be planned?
- Which components (or fatigue hotspots) should be inspected or repaired?

Besides providing answers to the aforementioned questions, the identified strategy should be, ideally, the optimal policy, resulting in the minimum expected life-cycle cost. This thesis is devoted to explore and propose inspection and maintenance methods able to address the modern societal and economic demands.

1.2 Inspection and maintenance planning: state-of-the-art

Inspection and Maintenance (I&M) planning consists not only in identifying optimal maintenance actions based on information collected from inspection or monitoring, but decisions on when and where to collect further information should also be planned. Dictating actions once information has already been collected corresponds to a posterior decision analysis, and if decisions on information collection are also planned, then the analysis is denoted as pre-posterior [1]. In the latter, the decision-maker faces a complex decision-making problem under uncertainty with the objective, as already introduced in Section 1.1, to minimize the expected total life-cycle cost. Since inspections are associated with measurement uncertainties, the environment becomes partially observable in practical applications, making the decision problem even more challenging. In such environments, simplifications to the decision problem are often introduced by state-of-the-art inspection and maintenance planning methods in order to identify policies within a reasonable computational time. Table 1.1 provides an overview of inspection and maintenance methods available in the literature, classified in terms of decision optimization, modelling approach, and applications. Specific mention to reference sources can be additionally found at the introduction of each chapter in the remainder of this thesis.

Policy optimization

Traditional risk-based inspection planning (RBI) methods [2] formulate the inspection and maintenance decision-making problem with basis on classical decision theory. In theory, the optimal policy can be identified by conducting a pre-posterior analysis, in which the probabilities and consequences of all the potential events are formally estimated. In practice, the horizon of the decision problem spans over time, and the extensive solution of the exponentially growing decision tree becomes computationally intractable. To alleviate the computational complexity, RBI approaches resort to policies based on predefined heuristic decision rules, planning inspections at equidistant time intervals or upon exceeding a specific failure probability threshold, and scheduling repair interventions after an observation (e.g. crack detection) is collected. The optimization problem is, therefore, simplified to identifying the optimized set of heuristics from all the predefined decision rules. The reader is directed to [3] for a detailed overview on RBI methods. These methods have often been applied to the I&M planning of offshore structures [4] or offshore wind turbines [5]. Whereas structural reliability methods, i.e. FORM, SORM or Monte Carlo simulations, are often utilized to compute risk metrics in RBI approaches, modern methods rely on dynamic Bayesian networks for modeling the underlying deterioration process. In any case, RBI policies are still defined according to predefined heuristics decision rules, usually specified with basis on an engineering or prior understanding of the problem at hand[6, 7].

Table 1.1. Survey of inspection and maintenance planning methods available in the literature.

	Optimization method		Modeling approach		Application		
Reference	Heuristics	POMDP	Component	System	O&G	OWT	Other
[2]	✓		✓		✓		
[3]	✓		✓		✓		
[4]	✓		✓		✓		
[5]	✓		✓			✓	
[8]	✓		✓			✓	
[9]	✓		✓			✓	
[10]	✓		✓			✓	
[11]	✓		✓				✓
[12]	✓		✓				✓
[13]	✓			✓	✓		
[14]	✓			✓	✓		
[15]	✓			✓	✓		
[16]	✓			✓	✓		
[17]	✓			✓	✓		
[6]	✓			✓	✓		
[18]	✓			✓	✓		
[19]	✓			✓	✓		
[20]	✓			✓	✓		
[21]		✓	✓				✓
[22]		✓	✓				✓
[23]		✓	✓		✓		
[24]		✓	✓		✓		
[25]		✓		✓	✓		
[26]		✓		✓	✓		
[27]		✓		✓	✓		
[28]		✓		✓	✓		
[29]		✓		✓	✓		

The main shortcoming of risk-based inspection planning methods is that the policy optimization is restricted to the evaluated heuristic decision rules, while the remaining policy space remains unexplored.

The I&M planning decision-making problem can, instead, be formulated as a Partially Observable Markov Decision Process (POMDP). In general, POMDPs constitute a principled mathematical framework for decision-making in stochastic environments and under partial observability. At their initial stage, POMDPs were restricted to small state space problems due to the computational complexity associated with solving high-dimensional POMDPs by, for instance, exact value iteration algorithms. With the advent of point-based solvers [30], initially applied to robotic navigation problems [31, 32], the inherent complexities of the solution process have been alleviated, thus enabling the treatment of medium to large state space settings. POMDP methods for inspection and maintenance planning of engineering systems can also be found in the literature [21, 23, 25, 26]. In contrast to RBI policies, based on specified heuristic decision rules, POMDP policies are defined as a function of the belief state, i.e. the probability distribution over states. Since the belief state is a sufficient statistic corresponding to the dynamically updated history of actions and observations, POMDP policies are intrinsically adaptive and result optimal if solved exactly.

I&M methods at the system level

In most of the reported I&M methods, the decision-making problem is formulated at the component level due to the additional computational complexities that arise when the optimization is extended to the system. I&M policies at the system level also dictate which components should be inspected or repaired at each time step, complicating, therefore, the policy search not only due to the higher dimensional state space involved, but also as a result of the high-dimensional action and observation spaces. Even if the computational complexity is alleviated by modeling the I&M decision problem at the component level, the policies might result suboptimal in most practical applications. System risk metrics are, for example, not considered if the optimization is only approached at the component level. Moreover, deterioration or cost dependencies among the constitutive components cannot be included if the decision problem is approached at the component level. The need for I&M methods capable of identifying policies at the 'system level' is a recurrent claim within the scientific community.

Early I&M planning methods approached at the system level include [13, 14, 15, 16]. In [17], fatigue hotspots are classified into categories according to their fatigue design factor, thereby constituting a proxy for identifying policies at the system level. More recent risk-based I&M planning approaches [6, 33] engineer strategies based on a set of optimized heuristic decision rules, supported by dynamic Bayesian networks for the efficient modeling of deterioration processes under deterioration, cost, and reliability dependence. As mentioned, the I&M policies are identified by optimizing a set of predefined heuristic decision rules: (i) equidistant inspection interval, (ii) number of inspected components, (iii) inspection of the components with higher failure probability, and (iv) repairs are planned after a detection indication. As for other heuristics-based optimization methods, the policies are restricted to the set of explored heuristic decision rules.

Addressing the complexities of managing large engineering systems, a deep reinforcement learning method has been introduced in [28], motivated by the success of deep reinforcement learning algorithms in complex game environments, e.g. in [34, 35, 36]. In particular, a multi-agent actor-critic is developed in [28], relying on (PO)MDPs, 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 I&M decision-making problems under constraints, e.g. imposed risk thresholds or budget limitations.

1.3 Objectives of the research

The overarching aim of this thesis is to investigate and contribute to the development of Inspection and Maintenance (I&M) methods able to identify optimal and rational strategies. The explored inspection and maintenance planning methods should be able to control the risk of failure events by optimally allocating inspections and maintenance interventions. Specifically, this investigation targets the achievement of the following objectives:

- **Objective 1:** To propose the necessary formulation for specifying the component level (I&M) decision-making problem as a Partially Observable Markov Decision Process (POMDP);
- **Objective 2:** To extend the POMDP-based component level formulation to the system level, estimating the reliability of the involved engineering system, enabling the treatment of deterioration dependent environments, and defining the cost model at a global scale;
- **Objective 3:** To investigate the underlying system effects of I&M strategies in environments under deterioration, structural reliability, and cost dependence;
- **Objective 4:** To thoroughly compare the resulting POMDP-based strategies against policies provided by conventional and state-of-the-art I&M methods;

- **Objective 5:** To apply the proposed methods for the optimal management of offshore wind turbine substructures subjected to fatigue deterioration.

This work has been financially supported by the FRNS and conducted in close collaboration with The Pennsylvania State University (USA).

1.4 Outline of the thesis

The developments and analyses conducted in this thesis are organized in a paper-based structure. After this introductory chapter, each following chapter corresponds to a submitted - or under review - paper, culminating with concluding remarks and further research directions in Chapter 7. Addressing the objectives listed in Chapter 1.3, the work is organized as illustrated in Fig. 1.3, providing methods for inspection and maintenance planning of deteriorating structures, with applications to the management of offshore wind substructures, both at component and system levels. In all the selected papers, the author of this thesis is also the first author and holds a primary responsibility for the conceptualization of the study, formal analysis, interpretation of the results, and writing. Additionally, the contributions of each co-author are listed at the end of each chapter.

	Methodology	Application
Component level	**Chapter 2** Optimal Inspection and Maintenance Planning for Deteriorating Structural Components through Dynamic Bayesian Networks and Markov Decision Processes	**Chapter 3** Managing Offshore Wind Turbines through Markov Decision Processes and Dynamic Bayesian Networks **Chapter 4** POMDP based Maintenance Optimization of Offshore Wind Substructures including Monitoring
System level	**Chapter 5** Optimal Management of Deteriorating Structures considering System Effects: a Deep Reinforcement Learning Approach	**Chapter 6** Optimal Management of Offshore Wind Structural Systems via Deep Reinforcement Learning and Bayesian Networks

Fig. 1.3. Organization of the PhD dissertation.

An efficient algorithmic platform for optimal decision-making under uncertainty is proposed in Chapter 2, integrating dynamic Bayesian networks with Partially Observable Markov Decision Processes (POMDPs) and providing the necessary formulation for deriving infinite or finite horizon POMDPs based on standard parametric or deterioration rate dynamic Bayesian networks. The proposed DBN-POMDP algorithmic scheme is implemented and tested for both traditional and detailed I&M planning settings. The results of the study reveal that POMDP-based policies offer substantial benefits compared to heuristic-based strategies.

In Chapter 3, the DBN-POMDP inspection and maintenance planning approach introduced in Chapter 2 is further implemented and tested for the case of an offshore wind structural detail subject to fatigue deterioration. Within the numerical experiments, the deterioration environment is specified according to typical offshore wind standards and the optimal decisions are identified for both a 20-year finite horizon I&M planning setting and for a lifetime extension planning application.

A POMDP-based approach is introduced, in Chapter 4, to quantify the expected benefits of installing a monitoring system, relying on value of information theoretical principles. The value of information is assessed by treating separately the I&M decision problem and estimating the difference in expected total cost of each considered monitoring scheme. In a representative numerical example, the value of monitoring is quantified for a traditional I&M planning setting.

The I&M decision problem at the system level is formulated, in Chapter 5, as a factored POMDP, whose transition and observation models are specified based on Bayesian networks. The proposed approach enables an efficient treatment of engineering systems under deterioration, structural reliability, and cost dependence. In terms of policy optimization, a deep decentralized multi-agent actor-critic (DDMAC) scheme is adopted, approximating POMDP policies by actor neural networks, guided by a critic neural network during the policy search. Since DDMAC adjusts the neural networks weights according to the collected system costs, POMDP-DDMAC policies intrinsically consider the underlying system effects, as demonstrated through the conducted numerical experiments.

In Chapter 6, monitoring choices are incorporated into the sequential I&M decision-making problem by specifying the monitoring observation model conditional on the sensors' health. Following the formulation introduced in Chapter 5, the decision problem is formulated as a POMDP, adopting also a DDMAC approach for computing I&M strategies at the system level. The proposed POMDP-DDMAC algorithmic scheme is implemented for the optimal monitoring, inspection, and maintenance planning of offshore wind substructures, both at the offshore wind turbine and offshore wind farm levels. The results show that DDMAC policies provides substantial benefits compared to corrective, calendar, and state-of-the-art heuristic-based strategies.

1.5 List of papers

Papers included in this thesis

The following papers are included in this thesis, selected as the most direct and relevant contributions to address the overall objectives stated in Chapter 1.3:

- **Paper 1:** Morato, P. G., Papakonstantinou, K. G., Andriotis, C. P., Nielsen, J. S. and Rigo P. (2021). Optimal Inspection and Maintenance Planning for Deteriorating Structural Components through Dynamic Bayesian Networks and Markov Decision Processes. *Structural Safety*, under review.

- **Paper 2:** Morato, P. G., Papakonstantinou, K. G., Andriotis, C. P. and Rigo P. (2021). Managing Offshore Wind Turbines through Markov Decision Processes and Dynamic Bayesian Networks. In *Proc. ICOSSAR 2021*, under internal review.

- **Paper 3:** Morato, P. G., Nielsen, J. S., Mai, A. Q. and Rigo P. (2019). POMDP-based Maintenance Optimization of Offshore Wind Substructures including Monitoring. In *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 2019)*, 270-277.

- **Paper 4:** Morato, P. G., Andriotis, C. P., Papakonstantinou, K. G. and Rigo P. (2021). Optimal Management of Deteriorating Structures considering System Effects: a Deep Reinforcement Learning Approach. *Reliability Engineering and System Safety*, under internal review.

- **Paper 5:** Morato, P. G., Andriotis, C. P., Papakonstantinou, K. G., Hlaing, N. and Rigo, P. (2021). Optimal Management of Offshore Wind Structural Systems via Deep Reinforcement Learning and Bayesian Networks. *Engineering Structures*, under internal review.

Other papers

The following additional contributions, also pertaining to the thematic of decision-making under uncertainty, are not included in this thesis:

- **Paper 6:** Hlaing, N., Morato Dominguez, P. G., Nielsen, J. S., Amirafshari, P., Kolios, A. and Rigo, P. (2021). The effect of failure criteria on risk-based inspection and maintenance planning of offshore wind support structures. *Structure and Infrastructure Engineering*, under review.

- **Paper 7:** Long, L., Mai, Q. A., Morato, P. G., Sørensen, J. D. and Thöns, S. (2020). Information value-based optimization of structural and environmental monitoring for offshore wind turbines support structures. *Renewable Energy, 159*, 1036-1046.

- **Paper 8:** Mai, Q. A., Weijtjens, W., Devriendt, C., Morato, P. G., Rigo, P. and Sørensen, J.D. (2019). Prediction of remaining fatigue life of welded joints in wind turbine support structures considering strain measurement and a joint distribution of oceanographic data. *Marine Structures, 66*, pp.307-322.

- **Paper 9:** Dang, T. V., Morato, P. G., Mai, Q. A. and Rigo, P. (2019). Optimal inspection and repair scheduling for mitre lock gates. In *Proceedings of the Institution of Civil Engineers-Maritime Engineering*, 172 (3), 95-103. Thomas Telford Ltd.

- **Paper 10:** Morato, P. G., Mai, Q. A., Rigo, P. and Nielsen, J. S. (2019). Point-based POMDP Risk Based Inspection of Offshore Wind Substructures. In *The Sixth International Symposium on Life-Cycle Civil Engineering (IALCCE 2019)*, 3069-3076. CRC Press.

- **Paper 11:** Hlaing, N., Morato Dominguez, P. G., Rigo, P., Amirafshari, P., Kolios, A., and Nielsen, J. S. (2021). The effect of failure criteria on risk-based inspection planning of offshore wind support structures. *InProc. 7th International Symposium on Life-Cycle Civil Engineering (IALCCE 2021).*

Chapter 2

Optimal Inspection and Maintenance Planning for Deteriorating Structural Components through Dynamic Bayesian Networks and Markov Decision Processes

Paper Morato, P. G., Papakonstantinou, K. G., Andriotis, C. P., Nielsen, J. S. and Rigo P. (2021). Optimal Inspection and Maintenance Planning for Deteriorating Structural Components through Dynamic Bayesian Networks and Markov Decision Processes. *Structural Safety*, under review.

Abstract Civil and maritime engineering systems, among others, from bridges to offshore platforms and wind turbines, must be efficiently managed, as they are exposed to deterioration mechanisms throughout their operational life, such as fatigue and/or corrosion. Identifying optimal inspection and maintenance policies demands the solution of a complex sequential decision-making problem under uncertainty, with the main objective of efficiently controlling the risk associated with structural failures. Addressing this complexity, risk-based inspection planning methodologies, supported often by dynamic Bayesian networks, evaluate a set of pre-defined heuristic decision rules to reasonably simplify the decision problem. However, the resulting policies may be compromised by the limited space considered in the definition of the decision rules. Avoiding this limitation, Partially Observable Markov Decision Processes (POMDPs) provide a principled mathematical methodology for stochastic optimal control under uncertain action outcomes and observations, in which the optimal actions are prescribed as a function of the entire, dynamically updated, state probability distribution. In this paper, we combine dynamic Bayesian networks with POMDPs in a joint framework for optimal inspection and maintenance planning, and we provide the relevant formulation for developing both infinite and finite horizon POMDPs in a structural reliability context. The proposed methodology is implemented and tested for the case of a structural component subject to fatigue deterioration, demonstrating

the capability of state-of-the-art point-based POMDP solvers of solving the underlying planning stochastic optimization problem. Within the numerical experiments, POMDP and heuristic-based policies are thoroughly compared, and results showcase that POMDPs achieve substantially lower costs as compared to their counterparts, even for traditional problem settings.

2.1 Introduction

Preserving infrastructures in a good condition, despite their exposure to diverse deterioration mechanisms throughout their operational life, enables, in most countries, a stable economic growth and societal development [37]. For instance, a bridge structural component may experience a thickness reduction due to corrosion effects [38, 39, 40, 41]; or a surface crack at an offshore platform might drastically propagate due to fatigue deterioration [42, 43, 44]; or the structural resistance of an offshore welded joint can be reduced due to the combined cyclic actions of wind and ocean waves [45, 46, 47]. The prediction of such deterioration processes involves a probabilistic analysis in which all relevant uncertainties are properly quantified.

Information about the condition of structural components can be gathered during their operational life through inspections or monitoring, allowing the decision maker to take more informed and rational actions [48, 49]. However, both maintenance actions and observations are associated with certain costs which must be optimally balanced against the risk of structural failure. As suggested by [50, 51] and others, inspections and/or maintenance actions should be planned with the objective of optimizing the structural life-cycle cost. Besides economic consequences associated with structural failures or maintenance interventions, societal and environmental aspects can also be included within a decision-making context in terms of utilities, defined in monetary units. A decision maker should, therefore, identify the decisions that result in the minimization of the total expected costs over the lifetime of the structure [52, 53].

In the context of Inspection and Maintenance (I&M) planning, the decision maker faces a complex sequential decision-making problem under uncertainty. This sequential decision-making problem is illustrated in Fig. 2.1, showcasing the involved random events and decision points, and can be formulated either from the perspective of the classical applied statistical decision theory [1], or through artificial intelligence [54] conceptions, or a combination thereof. In all cases, the main objective of a decision maker, or an intelligent agent, is to identify the optimal policy that minimizes the total expected costs.

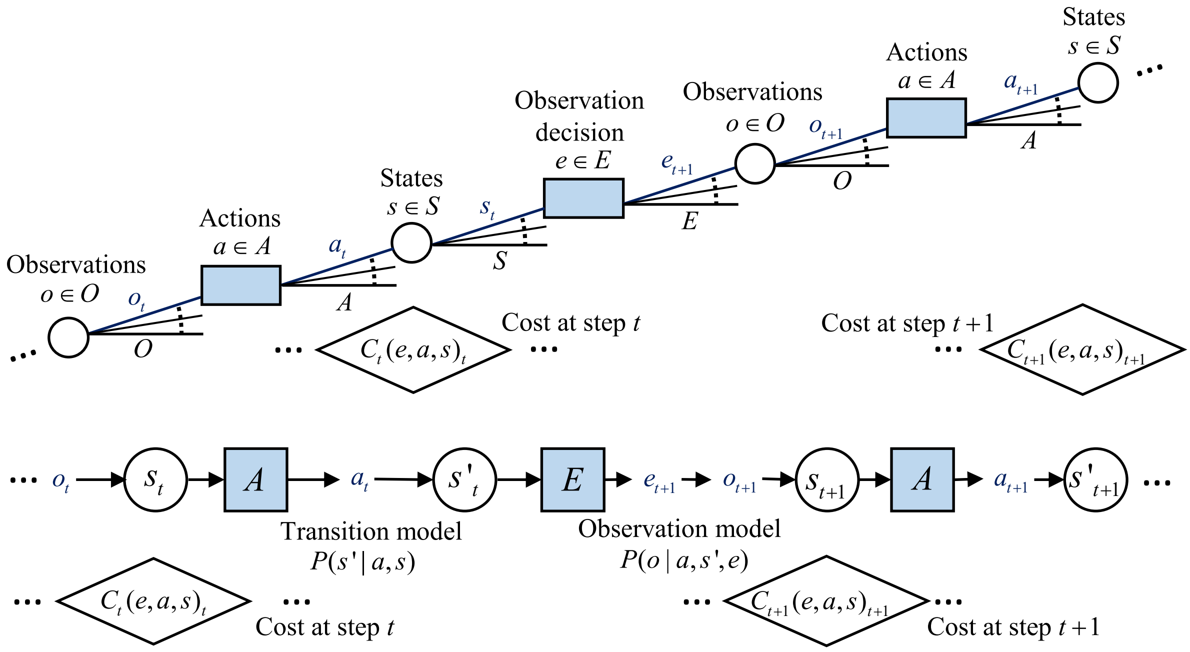


Fig. 2.1. (Top) Inspection and Maintenance (I&M) planning decision tree. Maintenance actions and observation decisions are represented by blue boxes and chance nodes are depicted by white circles. At every time step, the cost C_t depends on the action a, observation decision e, and state s of the component. (Bottom) An I&M POMDP sequence is represented where at each step t, the cost C_t depends on the action a, observation decision e, and state s of the component. In both representations, an observation outcome o is collected according to the current state, taken action and observation decision.

With the aim of addressing this complex decision-making problem, Risk-Based Inspection (RBI) planning methodologies have been traditionally proposed [55] and have often also been applied to the I&M planning of offshore structures [4, 5]. By imposing a set of heuristic decision rules, RBI methodologies are able to simplify and solve the decision-making problem within a reasonable computational time, while structural reliability methods are often employed within this framework, to quantify and update the reliability and risk metrics.

More recently, RBI methodologies have also been integrated with Dynamic Bayesian Networks (DBNs) [6, 7, 56, 57, 58]. DBNs provide an intuitive and robust inference approach to Bayesian updating; however, they do not tackle the decision optimization problem by themselves. In the proposed methodologies, heuristic decision rules, usually based on engineering principles and understanding of the problem, are still utilized to simplify the decision problem. Albeit their practical advantages, the main shortcoming of heuristic-based policies is the limited policy space exploration due to the prior, ad-hoc prescription of decision rules.

In this paper, we thus present how DBNs describing deterioration processes can be instead combined with Markov decision processes and dynamic programming [59], and be used to define transition and emission probabilities in such settings. Partially Observable Markov Decision Processes (POMDPs) provide a principled mathematical methodology for planning in stochastic environments under partial observability. In the past, POMDPs were only applicable for small state space problems due to the difficulty of finding appropriate solutions in a reasonable computation time. However, starting with the development of point-based solvers [30], which managed to efficiently alleviate the inherent complexities of the solution process, POMDPs have been increasingly used for planning problems, especially, in the field of computer science and robot navigation [31, 32]. POMDPs have also been proposed for I&M of engineering systems [21, 23, 25, 26, 60]. In the reported POMDP methodologies, either the condition of the structural component has been modeled with less than five discrete states or the rewards have not been defined in a structural reliability context. This different POMDP approach to the I&M problem, as compared with typical RBI applications, has raised some misconceptions in the literature about their use, which we formally rectify herein.

In this work, POMDPs are successfully combined with dynamic Bayesian networks in a joint framework, for optimal inspection and maintenance planning, in order to take advantage of both the modeling flexibility of DBNs and the advanced optimization capabilities of POMDPs. In particular, this paper originally derives the POMDP dynamics from DBNs, enabling optimal control of physically-based stochastic deterioration processes, modeled either through a conditional set of time-invariant parameters or as a function of the deterioration rate. We further provide all relevant formulations for deriving both infinite and finite horizon POMDPs within a structural reliability context. The proposed framework is analyzed, implemented, and tested for the case of a structural component subject to a fatigue deterioration process, and the capability of state-of-the-art point-based POMDP value iteration methods to efficiently solve challenging I&M optimization problems is verified. POMDP and typical heuristic risk-based and/or periodic policies are thoroughly analyzed and compared, in a variety of problem settings, and results demonstrate that POMDP solutions achieve substantially lower costs in all cases, as compared to their counterparts.

2.2 Background: Risk-based inspection planning

A typical Inspection and Maintenance (I&M) sequential decision problem under uncertainty is illustrated in Fig. 2.1. The optimal strategy can be theoretically identified by means of a pre-posterior decision analysis [1].

Assuming the costs at different times to be additive independent, the pre-posterior analysis prescribes the observation decisions $e \in E$ and actions $a \in A$ that minimize the total expected cost $C_T(a,e) = C_{t_0}(e,a,s)_{t_0} + ... + C_{t_N}(e,a,s)_{t_N}\gamma^{t_N}$, i.e. the sum over the lifetime t_N of the discounted costs received at each time step t, with γ being the discount factor. Note that societal and environmental consequences, specified in monetary units, can also be included within the definition of the total expected cost.

If the probabilities associated with the random events, as well as the costs, are assigned to each branch of the decision tree, then the branch corresponding to the optimal cost $C^*_T(a,e)$ can be identified. This analysis is denoted backwards induction or extensive analysis. Alternatively, a normal analysis can also be conducted by identifying the optimal decision rule, $h^*_{a,e}$, from all possible decision rules. In any case, the exact solution of a pre-posterior analysis very quickly becomes computationally intractable for practical problems because the number of branches increases exponentially with the number of time steps, actions, and observations.

2.2.1 RBI assumptions and heuristic rules

Risk-Based Inspection (RBI) planning methodologies [61] introduce simplifications to the I&M decision-making problem in order to be able to identify strategies in a reasonable computational time. To simplify the problem, the expected cost is computed only for a limited set of pre-defined decision rules $h_{a,e}$. The best strategy among them is then identified as the decision rule with the minimum cost.

Within an I&M planning context, the total expected cost $\mathbf{E}[C_T(h,t_N)]$ is the combination of expected costs from inspections $\mathbf{E}[C_I(h,t_N)]$, repairs $\mathbf{E}[C_R(h,t_N)]$, and failures $\mathbf{E}[C_F(h,t_N)]$, as a function of the imposed decision rules $h_{a,e}$. This expectation for a structural component designed for a lifetime of t_N years is simply computed as:

$$\mathbf{E}[C_T(h,t_N)] = \mathbf{E}[C_I(h,t_N)] + \mathbf{E}[C_R(h,t_N)] + \mathbf{E}[C_F(h,t_N)] \tag{2.1}$$

The simplifications introduced to the I&M decision-making problem by pre-defining a set of decision rules are listed below:

i) Observations (inspections) are planned according to a pre-defined heuristic rule. Two heuristic rules are commonly employed in the literature [62]:

- Equidistant inspections: inspections are planned at constant intervals of time Δt.
- Failure probability threshold: inspections are planned just before a pre-defined annual failure probability ΔP_F threshold is reached.

ii) If the outcome of an inspection indicates damage detection ($d > d_{det}$), a repair action is immediately performed. In that case, the repair probability is equal to the probability of detection $P_R = P(d > d_{det})$. Alternatively, other heuristic rules can also be imposed (adding computational complexity), such as that a repair is performed if an inspection indicates detection ($d > d_{det}$) and a pre-defined failure probability threshold P_F is simultaneously exceeded.

iii) After a component is repaired, it is assumed that it behaves like a component with no damage detection, i.e. the remaining life can be computed as if the inspection at the time of repair indicates no damage detection. With these assumptions, the decision tree represented in Fig. 2.1 can be simplified to a single branch. Alternatively, if a repair is performed at time t and it is assumed to be perfect, the component returns to its initial damage state at the beginning of a new decision tree with a lifetime equal to $\bar{t}_N = t_N - t$.

Summarizing, one can simplify the problem to one decision tree branch by assuming that: (i) inspections are to be planned according to a heuristic rule, (ii) a repair is to be performed if an inspection indicates detection, and (iii) after a repair is performed, the inspection at that time is considered as a no detection event. In this case, the individual contributions to the total expected cost in Eq. 2.1 can be computed analytically.
The expected inspection cost $\mathbf{E}[C_I(h, t_N)]$ is computed as the sum of all conducted inspections I_n, with individual inspection cost C_i, and discounted by the factor $\gamma \in [0, 1]$:

$$\mathbf{E}[C_I(h, t_N)] = \sum_{t_I = t_{I_1}}^{t_{I_n}} C_i \gamma^{t_I} \tag{2.2}$$

The expected repair cost $\mathbf{E}[C_R(h, t_N)]$ corresponding to a heuristic scheme h is calculated as the repair cost C_r multiplied by the probability of repair P_R at each inspection year t_I:

$$\mathbf{E}[C_R(h, t_N)] = \sum_{t_I = t_{I_1}}^{t_{I_n}} C_r P_R(h, t) \gamma^{t_I} \tag{2.3}$$

The expected risk of failure $\mathbf{E}[C_F(h, t_N)]$ is computed as the sum of discounted annual failure risks, in which ΔP_F is the annual failure probability and C_f is the cost of failure:

$$\mathbf{E}[C_F(h, t_N)] = \sum_{t=1}^{t_N} C_f \Delta P_F(h, t) \gamma^t \tag{2.4}$$

2.2.2 Probabilistic deterioration model and reliability updating

Structural reliability methods and general sampling based methods [63] can be used to compute the probabilities associated with the random events represented in the I&M decision tree (Fig. 2.1). In a simplified decision tree, the main random events are the damage detections during inspections and the structural failure.

The failure event is defined through a limit state $g_F(t) = d_c - d(t)$, in which d_c represents the failure criteria, such as the critical crack size, and $d(t)$ is related to the temporal deterioration evolution. Uncertainties involved in the deterioration process are incorporated by defining $d(t)$ as a function of a group of random variables or random processes. The probability of failure $P_F(t)$ can be then computed as the probability of the limit state being negative $P_F = P\{g_F(t) \leq 0\}$, and the reliability index is inversely related to the failure probability, usually defined in the standard normal space as $\beta(t) = -\Phi^{-1}\{P_F(t)\}$, in which Φ is the standard normal cumulative distribution function. The probability of the failure event can also be defined over a reference period, e.g. the annual failure probability can be computed as $\Delta P_F(t) = \{P_F(t) - P_F(t-1)\}$.

The measurement uncertainty of the available observations (inspections) is often quantified by means of Probability of Detection (PoD) curves. A PoD indicates the probability of detection as a function of the damage size d and depends on the employed inspection method, i.e. the function of the detectable damage size can be modeled by an exponential distribution $F(d_d) = F_0\Big[1 - exp(-d/\lambda)\Big]$, where F_0 and λ are parameters determined by experiments. The event of no detection at time t_I is then modeled by the limit state function $g_{I_{nd}}(t_I) = d(t_I) - d_d(t_I)$. Similarly, the event of detection at time t_I is modeled by the limit state $g_{I_d}(t_I) = d_d(t_I) - d(t_I)$. Both detection and no detection events are evaluated as inequalities, for instance, the probability of no detection is assessed as the probability of the limit state being negative $P_{I_{nd}} = P\{g_{I_{nd}}(t_I) \leq 0\}$. Alternatively, a discrete damage measurement d_m can be collected and the limit state is modeled in this case as $g_m(t_I) = d(t_I) - (d_m - \epsilon_m)$, where ϵ_m is a random variable that represents the measurement uncertainty, and the equality event $P_m = P\{g_m(t) = 0\}$ can be estimated equal to some limit, as explained in [63, 64, 65].

The additional information gained by observations can be used to update the structural reliability or failure probability P_F by computing a failure event conditional on inspection events [66], as:

$$P_{F|I_1,\dots,I_N}(t) = \frac{P\Big[g_F(t) \leq 0 \cap g_{I_1}(t) \leq 0 \cap \dots \cap g_{I_N}(t) \leq 0\Big]}{P\Big[g_{I_1}(t) \leq 0 \cap \dots \cap g_{I_N}(t) \leq 0\Big]} \tag{2.5}$$

The conditional failure probability introduced in Eq. 2.5 can be computed by structural reliability methods (FORM, SORM) or by Monte Carlo sampling methods [63].

2.3 Stochastic deterioration processes via Dynamic Bayesian Networks

A brief overview on the adoption of dynamic Bayesian networks (DBNs) for structural deterioration and reliability problems is presented here, with the objective of demonstrating that the main principles underlying BNs inference tasks are fundamentally invariant to those employed by POMDPs. Bayesian networks (BNs) are directed acyclic graphical models particularly suited for inference tasks in probabilistic environments. A DBN is a template model of a Bayesian network evolving over time and in the context of structural reliability and related problems, DBNs have played an important role [6, 7, 33]. For a detailed background of probabilistic graphical models and BNs, the reader is directed to [67].

To allow DBNs based inference within a reasonable computational time for practical problems, the following assumptions are often imposed:

i) Discrete state space: Exact inference algorithms are limited to discrete random variables [68]. A discretization operation must thus be performed to convert the original continuous random variables to the discrete space. The unknown error introduced by the discretization operation converges to zero in the limit of an infinitesimal interval size. However, the computational complexity of the inference task grows linearly with the number of states and exponentially with the number of random variables.

ii) Markovian assumption: The state space S is the domain of all random variables involved in the description of the deterioration process, and the conditional probabilities $P(s_{t+1}|s_t)$ associated with the random variables at time step $t+1$ depend only on the random variables at the current time step t, and are independent of all past states.

The transition probability matrix $P(s_{t+1}|s_t)$ can also be assumed as stationary for some applications, thus facilitating the formulation of the problem. This can however be easily relaxed without entailing additional computational efforts [9].

2.3.1 Parametric DBN

A stochastic deterioration process can be represented by the DBN shown in Fig. 2.2. The deterioration is represented through the damage node d_t which is influenced by a set of time-invariant random variables $\boldsymbol{\theta_t}$. The model is denoted as parametric DBN as the damage d_t is influenced by the parameters $\boldsymbol{\theta_t}$. Imperfect observations are added

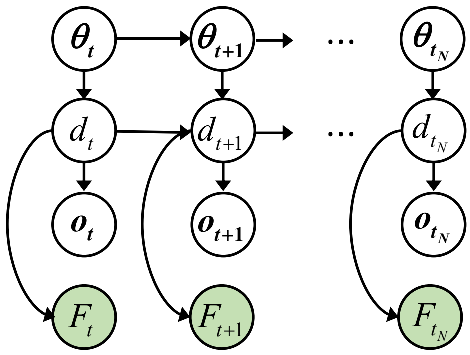


Fig. 2.2. Parametric dynamic Bayesian network, adapted from [7]. The evolution of a stochastic deterioration process is represented by the nodes d_t influenced by a set of time-invariant random variables $\boldsymbol{\theta_t}$. Imperfect observations are added through the nodes $\boldsymbol{o_t}$, and F_t binary node indicates the probability of failure and survival events.

into the DBNs by means of the node $\boldsymbol{o_t}$. This DBN can be extended by incorporating time-variant random variables as proposed by [7]; yet, we consider only time-invariant random variables here as they are widely used in the literature and to avoid unnecessary presentation complications. Finally, the binary node F_t provides an indication of the failure and survivability.

Within the context of structural reliability and related problems, DBNs are often employed to propagate and update the uncertainty related to a deterioration process, incorporating evidence from inspections or monitoring. Filtering becomes the preferred inference task for inspection and maintenance planning problems, as a decision is taken at time t supported by evidence gathered from the initial time step t_0 up to time t. The belief state, defined as the probability distribution over states, can be propagated and updated by applying the forward operation from the forward-backward algorithm [68]. The transition algorithmic step of the forward operation is assumed to be Markovian, being therefore equivalent to the underlying transition model of a POMDP. More details on the formulation of POMDP transition models are introduced in Section 2.4.1.

At time step t_0, the initial belief corresponds to the joint probability of the initial damage and time-invariant parameters $P(d_{t_0}, \boldsymbol{\theta_{t_0}})$. The forward operation is then applied for the subsequent time steps, comprised of the following steps:

1. Transition step: the belief propagates in time according to a pre-defined conditional probability distribution or transition matrix $P(d_{t+1}, \boldsymbol{\theta_{t+1}} | d_t, \boldsymbol{\theta_t})$, as:

$$P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) = \sum_{d_t} \sum_{\boldsymbol{\theta_t}} P(d_{t+1}, \boldsymbol{\theta_{t+1}} | d_t, \boldsymbol{\theta_t}) \, P(d_t, \boldsymbol{\theta_t} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{2.6}$$

2. Estimation step: the belief is now updated based on the obtained evidence by means of Bayes' rule, as:

$$P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_{t+1}}) \propto P(\boldsymbol{o_{t+1}} | d_{t+1}) P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) \qquad (2.7)$$

   The quality of the observation is quantified by the likelihood $P(\boldsymbol{o_{t+1}}|d_{t+1})$. This likelihood can be directly obtained from probability of detection curves or by discretizing a direct measurement. Since the random variables are discrete, a normalization of $P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_{t+1}})$ can be easily implemented.

The failure probability assigned to the node F_t corresponds to the probability of being in a failure state. As the failure states are defined based on the damage condition d_t, the time invariant parameters $\boldsymbol{\theta_t}$ can be marginalized out to compute the failure probability. Disregarding the discretization error, the resulting structural reliability is equivalent to the one computed in Eq. 2.5.

In terms of computational complexity, note that the belief is composed of $(|\theta_1| \cdot ... \cdot |\theta_k| |d|)$ states, defined by the damage d along with k time-invariant random variables. Thus, the transition matrix includes $(|\theta_1| \cdot ... \cdot |\theta_k| \, |d|)^2$ elements. Since $P(\theta_{t+1}|\theta_t)$ is defined by an identity matrix, the transition is prescribed by a very sparse, block-diagonal matrix with a maximum density of $\rho_P = 1/(|\theta_1| \cdot ... \cdot |\theta_k|)$.

2.3.2 Deterioration rate DBN

We present herein an alternative DBN in which a stochastic deterioration process is represented as a function of the deterioration rate. This model is adopted from [69] and denoted here as deterioration rate DBN. Fig. 2.3 graphically illustrates the model. In this case, the stochastic deterioration process is described in time t by the nodes d_t, conditional on the deterioration rate τ_t. If the stochastic process is stationary, the deterioration evolution will vary equally over time, and thus the deterioration rate τ_t is not utilized. The deterioration does not, however, progress equally over time in a non-stationary process, and in that case, the parameter τ_t needs to be incorporated to effectively model the varying deterioration effects over time. After collecting experimental or physically-based simulated data (e.g. Monte Carlo simulations) from a non-stationary deterioration process, the transition probabilities can be calculated, for each deterioration rate τ_t, by counting the number of transitions from d_t to d_{t+1} over the total data available in d_t. Additional methods to compute the transition model are described in [69]. As illustrated in Fig. 2.3, imperfect observations are added through the nodes $\boldsymbol{o_t}$ and the structural reliability is indicated through the node F_t.

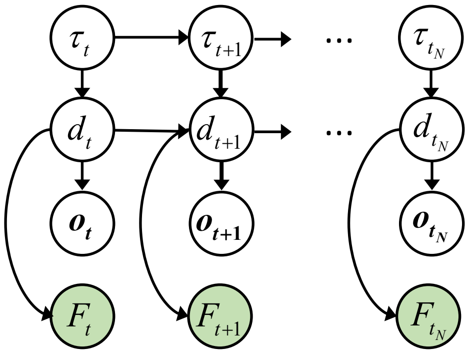


Fig. 2.3. Deterioration rate dynamic Bayesian network, derived from [69]. The evolution of a stochastic deterioration process is represented by the nodes d_t dependent on the deterioration rate τ_t. Imperfect observations are included through the nodes $\boldsymbol{o_t}$, and F_t binary node indicates the probability of failure and survival events.

To ensure compliance with the DBNs time invariant property, the belief incorporates both the damage condition and deterioration rate through the joint probability $P(d_t, \tau_t)$. Yet, the node τ_t is a zero-one vector (one-hot) that transitions each time step from one deterioration rate τ_i to the next τ_{i+1}. The deterioration evolution is computed by a forward operation in a similar manner as for the parametric DBN. Initially, the belief corresponds to the joint probability $P(d_0, \tau_0)$. Subsequently, the belief experiences a transition according to the transition matrix $P(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$:

$$P(d_{t+1}, \tau_{t+1}|\boldsymbol{o_0}, ..., \boldsymbol{o_t}) = \sum_{d_t}\sum_{\tau_t} P(d_{t+1}, \tau_{t+1}|d_t, \tau_t)\, P(d_t, \tau_t|\boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{2.8}$$

Based on the gathered observations, the beliefs are then updated by applying Bayes' rule. The likelihood $P(\boldsymbol{o_{t+1}}|d_{t+1})$ can be directly defined from probability of detection curves or other observation uncertainty measures:

$$P(d_{t+1}, \tau_{t+1}|\boldsymbol{o_0}, ..., \boldsymbol{o_{t+1}}) \propto P(\boldsymbol{o_{t+1}}|d_{t+1}) P(d_{t+1}, \tau_{t+1}|\boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{2.9}$$

The computational complexity is influenced by the belief size. For the case of a deterioration rate DBN, the belief $P(d_t, \tau_t)$ is composed of $|\tau| \cdot |d|$ states and its sparse transition matrix $P(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$ accounts for $(|\tau|\,|d|)^2$ elements. Since the only non-zero probabilities of the transition matrix $P(\tau_{t+1}|\tau_t)$ are the ones to define the transition from deterioration rate τ_t to the next deterioration rate τ_{t+1}, the maximum density of $P(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$ is $\rho_{DR} = 1/|\tau|$.

Advantages between a parametric DBN and a deterioration rate one are case dependent. If the deterioration process can be modeled by just few parameters or it evolves over a long time span, the parametric DBN is recommended. However, if the deterioration modeling involves many parameters or complex random processes spanning over a short time horizon, the deterioration rate DBN should be preferred. If both DBN models are applied for the same problem, the results should be equivalent and differences are only affected by the discretization error.

Risk-based inspection planning and DBNs

While DBNs can be successfully used for reliability updating, they do not possess by themselves intrinsic optimization capabilities. To this end, modern RBI methodologies include a combination of DBNs and heuristic rules to identify the optimal strategy [6, 7]. The methodologies often follow a similar logic as the theoretical scheme presented in Section 2.2, where the decision tree is simplified.

Alternatively, the optimal I&M strategy among different alternatives can be identified with the support of DBNs in a simulation environment. Any of the proposed DBN types (Sections 2.3.1 and 2.3.2) can be generalized to an influence diagram by adding utility and decision nodes [6]. The total cost C_T for a set of pre-defined heuristic rules $h_{a,e}$ can be computed by simulating one episode ep of length t_N as:

$$C_{T_{ep}}(h) = \sum_{t=t_0}^{t_N} \left[C_i(t)\gamma^t + C_r(t)\gamma^t + \Delta P_F(t) C_f \gamma^t \right] \tag{2.10}$$

The total expected cost $\mathbf{E}[C_T(h)]$ is then computed with a Monte Carlo simulation of n_{ep} episodes (policy realizations):

$$\mathbf{E}[C_T(h)] = \frac{\sum_{ep=1}^{n_{ep}} \left[C_{T_{ep}}(h) \right]}{n_{ep}} \tag{2.11}$$

One can compute the costs of all pre-defined heuristic rules and identify the strategy with the minimum expected cost as the optimal policy. However, the resulting optimal policies might be compromised due to the limited space covered by the imposed heuristic rules, out of all possible decision rules.

2.4 Optimal I&M planning through POMDPs

We propose herein a methodology for optimal I&M planning of deteriorating structures under uncertainty based on Partially Observable Markov Decision Processes (POMDPs). The methodology is adopted by similar frameworks, as studied in [22]. While the damage evolution was modeled in [22] as function of its deterioration rate, following the formulation presented in Section 2.3.2, we extend here the methodology to deterioration mechanisms modeled as functions of time-invariant parameters, formulated according to Section 2.3.1. In addition, the user penalty is defined in this work as a consequence of the annual failure probability experienced by the component.

A Markov decision process (MDP) is a 5-tuple $\langle S, A, T, R, \gamma \rangle$ controlled stochastic process in which an intelligent agent acts in a stochastic environment. The agent observes the component at state $s \in S$ and takes an action $a \in A$, then the state randomly transitions to state $s' \in S$ according to a transition probability model $T(s, a, s') = P(s'|s, a)$, and finally the agent receives a relevant reward $R_t(s, a)$, where t is the current decision step.

As described in Section 2.1, the optimal decisions result in a minimum expected cost. The expected cost, or value function, is expressed for a finite horizon MDP as the summation of the decomposed rewards $V(s_0) = R_{t_0} + ... + R_{t_{N-1}}\gamma^{t_{N-1}}$, from time step t_0 up to the final time step t_{N-1}. For an infinite or unbounded horizon MDP, the rewards are infinitely summed up ($t_N = \infty$). Note that the rewards are discounted by the factor γ. From an economic perspective, the discount factor converts future rewards into their present value. Computationally, discounting is also necessary to guarantee convergence in infinite horizon problems.

An MDP policy ($\pi : S \to A$) prescribes an action as a function of the current state. The main goal of an MDP is the identification of the optimal policy $\pi^*(s)$ which maximizes the value function $V^*(s)$. There exist efficient algorithms that compute the optimal policy using the principles of dynamic programming and invoking Bellman's equation. Both value and policy iteration algorithms can be implemented to identify the optimal policy $\pi^*(s)$ [70]. While the state of the component in an MDP is known at each time step, imperfect observations are usually obtained in real situations, e.g. noise in the sensor of a robot, measurement uncertainty of an inspection, etc. POMDPs are a generalization of MDPs in which the states are perceived by the agent through imperfect observations. The POMDP becomes a 7-tuple $\langle S, A, O, T, Z, R, \gamma \rangle$. While the dynamics of the environment are the same as for an MDP, an agent collects an observation $o \in O$ in the state $s' \in S$ with emission probability $Z(o, s', a) = P(o|s', a)$, after an action $a \in A$ is taken. Fig. 2.4 shows the dynamic decision network of a POMDP, which is built based on a parametric model. A deterioration rate POMDP can be equally represented if one replaces the time-invariant parameters $\boldsymbol{\theta}$ by a deterioration rate variable τ. Since an agent is uncertain about the current state, the decisions should in principle be planned based on the full history of observations $o_1 : o_t$, up to the current decision step t. Instead, a belief state $b(s)$ is tracked to plan the decisions.

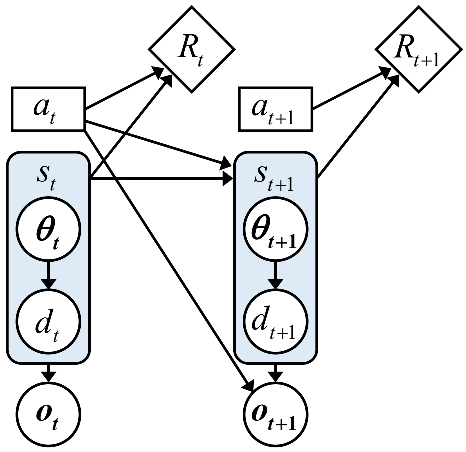


Fig. 2.4. Graphical representation of a Partially Observable Markov Decision Process (POMDP). The states S_t are modeled as the joint distribution of the time-invariant parameters $\boldsymbol{\theta_t}$ and the damage size d_t. The imperfect observations are modeled by the node $\boldsymbol{o_t}$. Actions a_t are represented by rectangular decision nodes and rewards R_t are drawn with diamond shape nodes. A deterioration rate POMDP can be graphically modeled by adding a deterioration rate variable τ_t instead of the time-invariant parameters $\boldsymbol{\theta_t}$.

A belief state $b(s)$ is a probability distribution over states and it is updated as a function of the transition $T(s', a, s)$ and collected observation $Z(o, s', a)$:

$$b'(s') \propto P(o|s', a) \sum_{s \in S} P(s'|s, a) b(s) \tag{2.12}$$

The normalizing constant $P(o|\mathbf{b}, a)$ is the probability of collecting an observation $o \in O$ given the belief state $\mathbf{b}$ and action $a \in A$.

One can see in Eq. 2.12 that for a specific action $a \in A$, updating a belief is equivalent to the forward operation described for DBNs in Eqs. 2.6-2.9. Yet, the main objective of a POMDP is to identify the optimal policy $\pi^*(\mathbf{b})$ as a function of the belief state $\mathbf{b}$. Since the belief state is a sufficient statistic equivalent to the history of all taken actions and gathered observations, a policy $\pi^*(\mathbf{b})$ as function of $\mathbf{b}$ will always be optimal, as compared to a policy $\pi(h)$ constrained by a limited set of heuristic rules $h_{a,e}$. This is also demonstrated through numerical experiments in Section 2.5. In Section 2.4.1, POMDP implementation details are provided and in Section 2.4.2, we explain how point-based solvers are able to solve high-dimensional state space POMDPs and find the optimal strategies.

2.4.1 POMDP model implementation

A systematic scheme for building a POMDP model in the context of optimal inspection and maintenance planning is provided in this section. A POMDP is built by defining all the elements of the tuple $\langle S, A, O, T, Z, R, \gamma \rangle$. While most of the reported applications of POMDPs for infrastructure planning employed a deterioration rate model [22], a parametric model as presented in Section 2.3.1 is originally implemented here.

States

For the typical discrete state MDP/POMDP cases, a discretization should be first performed for continuous random variables, transforming them to the discrete state space. As mentioned in Section 2.3, an efficient discretization has to balance model fidelity and computational complexity.

To construct an infinite horizon POMDP equivalent to the DBN parametric model presented in Section 2.3.1, the states $S_t = S_{d_t} \times S_{\boldsymbol{\theta}}$ are assigned as the domain instances of the joint probability $P(d_t, \boldsymbol{\theta})$. POMDPs are often represented in robotics applications by Markov hidden models containing only one hidden random variable. This has induced some confusion in the literature, where it is reported that POMDPs cannot handle deterioration mechanisms as function of time-invariant parameters [71]. However, a deterioration represented by time-invariant parameters can be easily modeled with POMDPs by augmenting the state space to include the joint probability distribution $P(d_t, \boldsymbol{\theta})$. While state-augmentation techniques have been already proposed in the literature [22, 72, 73], we particularly augment the state space here in order to specify the POMDP dynamics based on deterioration processes modeled as parametric DBNs that also include time-invariant parameters. This approach can also accommodate formulations with model updating. Naturally, augmenting the state space implies an increase of computational complexity, as is the case for both DBNs and POMDPs.

If the deterioration rate model (Section 2.3.2) is instead preferred, the states $S_t = S_{d_t} \times S_{\tau_t}$ are defined directly from the domain of the joint probability $P(d_t, \tau_t)$. The implementation for this case is documented in [22]. At the initial time step, one can prescribe the initial belief $\mathbf{b_0}$ as the joint probability $P(d_{t=0}, \boldsymbol{\theta})$ or $P(d_{t=0}, \tau_0)$.

Action-observation combinations

Actions $a \in A$ correspond to maintenance actions, such as “do-nothing", “perfect-repair" or “minor-repair", and observation action $e \in E$ are defined based on the available inspection or monitoring techniques, such as “no-observation", “visual-inspection" or “Nondestructive Evaluation (NDE)-inspection".

Since rewards are assigned as a result of an agent who takes an action and perceives an observation, it is recommended to combine actions and observations into groups [22]. For instance, one can combine the action "do-nothing" with two inspections, resulting in the two combinations: "do-nothing / visual-inspection" or "do-nothing / NDE-inspection" and a relevant reward will be assigned to each combination.

Transition probabilities

A transition matrix $T(s, a, s')$ models the transition probability $P(s'|s, a)$ of a component from state $s \in S$ to state $s' \in S$ after taking an action $a \in A$. Therefore, the transition matrix is constructed as a function of the maintenance actions:

- Do-nothing (DN) action: there is no maintenance action planned in this case and the state evolves according to the stochastic deterioration process. For an infinite horizon POMDP, the transition probability $T(s, a_{DN}, s')$ is equal to the transition matrix $P(d_{t+1}, \boldsymbol{\theta_{t+1}}|d_t, \boldsymbol{\theta_t})$ or $P(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$, derived in Section 2.3.

- Perfect repair (PR) action: a maintenance action is performed and the component returns from its current damage belief $\mathbf{b_t}$, at time step t, to the belief $\mathbf{b_0}$, associated with an intact status. In a belief space environment, a perfect repair transition matrix is defined as:

$$\mathbf{P}(s'|s, a_{PR}) = \begin{pmatrix} b_0(s_0) & b_0(s_1) & \cdots & b_0(s_k) \\ b_0(s_0) & b_0(s_1) & \cdots & b_0(s_k) \\ \vdots & \vdots & \ddots & \vdots \\ b_0(s_0) & b_0(s_1) & \cdots & b_0(s_k) \end{pmatrix} \tag{2.13}$$

  Since the belief state is a probability distribution, the summation over all the states is equal to one ($\sum b_t(s) = 1$). If one multiplies a belief state by the transition matrix defined in Eq. 2.13, the current belief returns to the belief $\mathbf{b_0}$, independently of its current condition as:

$$b_0(s) = b_t(s)\, \mathbf{P}(s'|s, a_{PR}) \tag{2.14}$$

  If the states are fully observable, the belief state becomes a zero-one vector and a perfect repair matrix can be formulated as $\mathbf{P}(s_0|s_t, a_{PR}) = 1$, transferring any state s_t to the intact state s_0.

- Imperfect repair (IR) action: a maintenance action is performed and the component returns from a damage belief $\mathbf{b_t}$ to a healthier damage state or more benign deterioration rate. The definition of the repair transition matrix $\mathbf{P}(s_{t+1}|s_t, a_{IR})$ is thus case dependent. Some examples can be found in [22].

Observation probabilities

An observation matrix $Z(o, s', a)$ quantifies the probability $P(o|s', a)$ of perceiving an observation $o \in O$ in state $s' \in S$ after taking action $a \in A$. Note that we denote the observation action as a to be coherent with usual POMDP formulation; yet the observation action could be also named as e to be consistent with the nomenclature used in Section 2.2.1. The relevant observation actions considered here are:

- No observation (NO): the belief state should remain unchanged after the transition as no additional information is gathered. The emission probability $\mathbf{P}(o|s', a_{NO})$ can be modeled as a uniform distribution over all observations. Alternatively, it can be modeled as $\mathbf{P}(o_0|s', a_{NO}) = 1$. The former is recommended as it will speed up the computation [22].

- Discrete indication (DI): the likelihood $P(o|s', a_{DI})$ is modeled as a discrete event, for instance, a binary indication: detection or no-detection. The likelihood is usually quantified for the binary case by a Probability of Detection (PoD) curve. A $PoD(s')$ is equivalent to the probability $P(o_D|s')$ of collecting an observation $o_D \in O$ as function of the state $s' \in s$, and the emission probability can be directly implemented as $P(o_D|s', a_{DI}) = PoD(s')$. The implementation can be equally applied for a higher dimensional discrete observation space.

- Continuous indication (CI): the likelihood $P(o|s', a_{CI})$ is modeled as a continuous distribution, for example, a direct measure of a crack. In this case, the observation space must be discretized into a finite set of observations.

Rewards

An agent having a belief $\mathbf{b}$, receives a reward $R(\mathbf{b}, a)$ after taking an action $a \in A$ and collecting an observation $o \in O$. In a MDP, the reward $R(s, a)$ is defined as a function of the state, while in a POMDP, the reward $R(s, a)$ is weighted over the belief state $\mathbf{b}$ to finally obtain $R(\mathbf{b}, a)$:

$$R(\mathbf{b}, a) = \sum_{s \in S} b(s) R(s, a) \tag{2.15}$$

For ease of notation, the reward model is formulated hereafter based on the same notation used for the definition of the RBI cost model in Section 2.2. If desired, societal, environmental, and other consequences can also be incorporated to the reward model. In the context of infrastructure planning, the state cost $C(s, a, s')$ is defined depending on the action-observation combination. Some recommendations are listed below:

- Do-nothing/no-observation (DN/NO): this case corresponds to computing the failure risk. Once the failure state subspace $S_F \subseteq S$ is defined, the annual failure probability is the probability $P(S'_F|S)$ of reaching any state in the failure state subspace S'_F from the state space S. Alternatively, Eq. 2.16 defines the cost $C_F(s, a_{DN-NO})$ only as a function of the initial state $s \in S$, if the transition matrix $P(s'|s, a)$ is implicitly considered. This option leads to a faster computation with a point-based solver, as explained subsequently. The cost value $\bar{C}(s, a_{DN-NO})$ is equal to the failure cost C_f if $s \in S_F$, and equal to 0, otherwise:

$$C_F(s, a_{DN-NO}) = \sum_{s' \in S_F} \{P(s'|s, a_{DN-NO})\, C_f\} - \bar{C}(s, a_{DN-NO}) \tag{2.16}$$

- Do-nothing/observation (DN/O): the cost is equal in this case to the one related failure risk plus one inspection cost. Both discrete and continuous indications can be included in this category. One can therefore compute the cost $C_O(s, a_{DN-O})$ just by further considering the inspection cost C_i:

$$C_O(s, a_{DN-O}) = C_F(s, a_{DN-NO}) + C_i \tag{2.17}$$

- Repair/no-observation (R/NO): the cost $C_R(s, a_{R-NO})$ is equal to the repair cost C_r:

$$C_R(s, a_{R-NO}) = C_r \tag{2.18}$$

  The cost $C_R(s, a_{R-O})$ for a repair/inspection combination can be similarly defined by including also the inspection cost C_i along with the repair cost $C_R(s, a_{R-NO})$.

2.4.2 Point-based POMDP solvers

In principle, one could apply a value iteration algorithm [74] to solve a POMDP. While value updates are computed in a $|S|$-dimensional discrete space for an MDP, value updates for POMDPs should be instead computed in a $(|S| - 1)$-dimensional continuous space. The computation thus scales up considerably with the number of dimensions, increasing the computational complexity. This fact is denoted as the curse of dimensionality.

Moreover, planning in a horizon t_N also suffers from the curse of history, as the number of potential action-observation histories scales exponentially with the number of time steps. Hence, solving POMDPs by applying a value iteration algorithm to the whole belief state space $\mathbb{B}$, or even to a discretized belief space grid, becomes computationally intractable for practical problems.

Relatively recent, however, point-based solvers have emerged able to solve high-dimensional state space POMDPs. Point-based solvers compute value updates only based on a representative set of belief points. Several point-based solvers [31, 32, 75] have been presented in the literature. Their main difference is their basis for selecting the set of representative belief points. The reader is directed to [24] for a detailed analysis of point-based solvers applied to infrastructure planning problems.

In an I&M planning context, the main objective is to identify the optimal policy, as explained in Section 2.2. Instead of constraining the policy space with pre-defined decision rules, POMDPs' main objective is to find the sequence of actions $a_0, ..., a_t$ that maximizes the expected sum of rewards for each belief $\mathbf{b} \in \mathbb{B}$. The value function is then formulated as a function of beliefs:

$$V^*(\mathbf{b}) = \max_{a \in A} \left[\sum_{s \in S} b(s) R(s,a) + \gamma \sum_{o \in O} P(o|\mathbf{b}, a) V^*(\mathbf{b_{s'}}) \right] \tag{2.19}$$

It is demonstrated in [76] that the value function is piece-wise linear and convex when it is solved exactly. The piece-wise linearity property is related to an effective value function parametrization by a set of hyper-planes or α-vectors $\in \Gamma$, each of them associated with an action $a \in A$. The optimal policy $\pi^*(b)$ can be selected by identifying the α-vectors that maximize the value function $V^*(\mathbf{b})$:

$$V^*(\mathbf{b}) = \max_{\alpha \in \Gamma} \sum_{s \in S} \alpha(s) b(s) \tag{2.20}$$

The convexity property now is associated with the value of information theory [77], i.e. lower-entropy states result in better decisions and as such have higher expected values than higher-entropy states. Both of these properties of piece-wise linearity and convexity can be easily visualized in up to 4D state spaces, e.g. in [21]. Naturally, in applications where the state space is augmented, as explained in Section 2.4.1, the belief still remains a probability over states and the value function preserves its piece-wise linearity and convexity at this newly defined, enhanced state space.

Finite horizon POMDPs

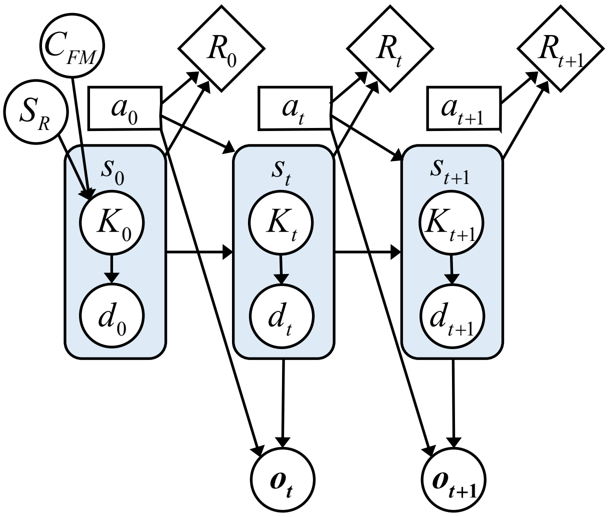
Existing point-based solvers are mostly able to solve large state space problems for infinite horizon POMDPs [78]. However, an infinite horizon POMDP can be transformed to a finite horizon one by augmenting the state space, as proposed by [21, 22, 48]. In this case, the time must be encoded in the state space and a terminal state is required. Note that the resulting transition, observation and reward matrices will be very sparse. Yet, it remains essential to augment the space efficiently by taking into consideration the nature of the decision-making problem. Some recommendations are listed below:

- Parametric model: the transition model is stationary. Then, the same transition matrix built for an infinite horizon POMDP can be reused for any time step of the augmented, finite horizon POMDP. To ensure a finite horizon, the last time step must include an absorbing state. An infinite horizon POMDP with $|S|$ states and $|A|$ actions can be augmented to a $|A|\,|S|\,t_N + |S| + 1$ finite horizon one with horizon t_N.

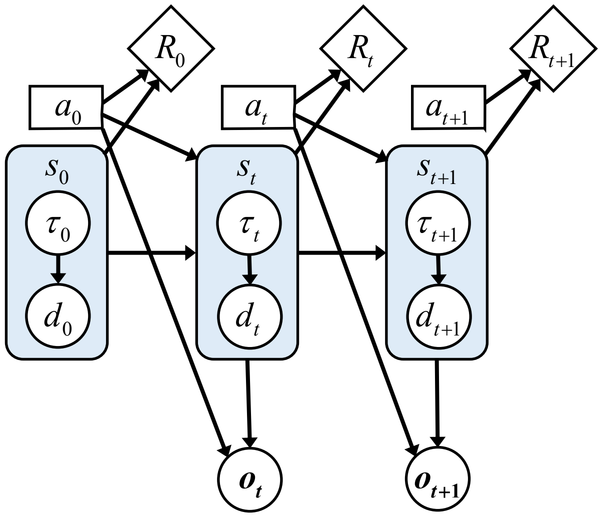
- Deterioration rate model: the state space can be efficiently formatted if the component experiences only one deterioration rate per time step. This way, one deterioration rate is considered at the first time step, two deterioration rates at the second time step, and so on, incorporating one additional deterioration rate per step until the last time step is reached. An absorbing state must also be included at the end. A deterioration rate model with $|S_d|$ states, spanning over a t_N horizon and two actions (do-nothing and one maintenance action) becomes a finite horizon POMDP with $\{(t_N+1)^2|S_d| + (t_N+1)|S_d|\}/2+1$ states. Additional maintenance actions can be included without an increase of the state space if they do not introduce additional/new deterioration rates.

2.5 Numerical experiments: Crack growth represented by time-invariant parameters.

With the main objectives of providing implementation details for the two presented POMDP formulations, as well as quantifying the differences in policies and costs between POMDP and heuristic-based I&M approaches, a set of numerical experiments is performed in this section. All computations are conducted on an Intel Core $i9-7900X$ processor with a clock speed of $3.30\,GHz$. The experiments consist in identifying the optimal I&M strategy for a structural component subjected to fatigue deterioration.



(a) Parametric POMDP



(b) Deterioration rate POMDP

Fig. 2.5. Graphical representation of the POMDPs utilized for the numerical experiments. A parametric POMDP and a deterioration rate POMDP are created from the DBNs displayed in Fig. 2.2 and Fig. 2.3, respectively. Note that the random variables C_{FM} and S_R are combined into the variable K.

In particular, the first presented I&M planning setting (in Section 2.5.2) is inspired by an earlier investigation of risk-based maintenance planning methods [71]. In that study, the fatigue deterioration model was approximated by a 2-parameter Weibull distribution, whereas a physically-based crack growth model is directly utilized here. According to this fracture mechanics model, the crack size d_{t+1} is computed as a function of the crack size at the previous time step d_t:

$$d_{t+1} = \left[\left(1 - \frac{m}{2}\right) C_{FM} S_R^m \pi^{m/2} n + d_t^{1-m/2}\right]^{2/(2-m)} \tag{2.21}$$

This Markovian model is derived from Paris' law, as shown in [63]. The process uncertainty is incorporated through the random variables listed in Table 2.1, where S_R stands for stress range, C_{FM} corresponds to a crack growth parameter, and d_0 represents the initial crack size. While the crack distribution evolves over time, the parameters C_{FM} and S_R are time-invariant random variables. The remaining parameters, i.e. the crack growth parameter m and the number of cycles n are considered deterministic. The component fails once the crack exceeds the plate thickness d_c and its considered life spans over a finite horizon t_N of 30 years.

Table 2.1. Random variables and deterministic parameters utilized to model fatigue deterioration.

Variable	Distribution	Mean	Standard Deviation
$ln(C_{FM})$	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 2.2. Description of the discretization schemes considered in the sensitivity analysis, for both parametric and deterioration rate POMDP models.

Variable	Interval boundaries
Parametric model	
S_d	$0, \exp\left\{\ln(10^{-1}) : \frac{\ln(d_c) - \ln(10^{-1})}{\lvert S_d\rvert - 2} : \ln(d_c)\right\}, \infty$
S_K	$0, \exp\left\{\ln(10^{-5}) : \frac{\ln(1) - \ln(10^{-5})}{\lvert S_K\rvert - 2} : \ln(1)\right\}, \infty$
Deterioration rate model	
S_d	$0, \exp\left\{\ln(10^{-4}) : \frac{\ln(d_c) - \ln(10^{-4})}{\lvert S_d\rvert - 2} : \ln(d_c)\right\}, \infty$
S_τ	$0 : 1 : 30$

2.5.1 Discretization analysis

A discretization analysis is performed to select an appropriate state space for this application. As explained in Section 2.3, either a parametric model or a deterioration rate model can be used to track the deterioration. The transition models are calculated, for both DBN models, based on data collected from simulations of the fracture mechanics model in Eq. 2.21. The POMDPs associated with these models are graphically represented in Fig. 2.5. Note that the parameters C_{FM} and S_R are grouped together for the parametric model, resulting in a new parameter K. By combining two random variables into one, we alleviate computational efforts [7]. K thus corresponds to $C_{FM} S_R^m \pi^{m/2} n$.

The main purpose of a proper discretization is to allocate the relevant intervals so that a high accuracy is achieved, without hindering computational tractability. Although several simulations were run, the reported results are mainly related to the case in which two inspections are planned at years 18 and 25, resulting in a no-detection outcome. The inspection quality is quantified with a probability of detection curve $PoD(d) \sim Exp[\mu = 8]$.

A crude Monte Carlo Simulation (MCS), containing 10^7 samples, was run to estimate the cumulative failure probability $P_{F_{MCS}}$ (Eq. 2.5). The accuracy is quantified here as the squared difference between $P_{F_{MCS}}$ and the cumulative failure probability P_F retrieved by each discretized state space model. P_F was obtained by unrolling a DBN over time. Note that P_F can be calculated directly through a DBN, as the probability of being in the failure states of d. Both $P_{F_{MCS}}$ and P_F are normalized to $\bar{P}_F = (P_F - \mu_{P_{F-MCS}})/\sigma_{P_{F-MCS}}$, where $\mu_{P_{F-MCS}}$ and $\sigma_{P_{F-MCS}}$ are the mean and standard deviation of the failure probabilities computed by MCS, respectively. The error ξ is computed as the squared difference of $\bar{P}_{F_{MCS}}$ and $\bar{P}_F$ for each time step:

$$\xi = \sum_{t=0}^{N} \left[\bar{P}_{F_{MCS}}(t) - \bar{P}_F(t) \right]^2 \tag{2.22}$$

Table 2.2 lists the discretization intervals for both parametric and deterioration rate models. Since the discretization is arbitrary, the interval boundaries were selected by trial and error, according to the recommendations proposed in [7], i.e. a logarithmic transformation is applied to both S_d and S_k spaces. Different state spaces were also tested by varying the number of states for $|K|$ and $|d|$. Table 2.3 reports the error ξ for each considered state space. While the deterioration rate model of 930 overall states results in an error of magnitude less than 10^{-3}, the state space of the parametric model is increased up to 16,000 overall states to achieve an error of magnitude less than 10^{-3}. To illustrate the differences between the analyzed models, Fig. 2.6 shows the unnormalized error $|P_{F_{MCS}} - P_{F_{DBN}}|$ for each case. The error of the deterioration rate model is negligible before the first inspection update at 18 years, while the parametric model accumulates error throughout the whole analysis.

Table 2.3. Accuracy of the considered discretization schemes. The normalized error ξ and state spaces corresponding to each parameter are reported.

Model	$\lvert S_K \rvert$	$\lvert S_\tau \rvert$	$\lvert S_d \rvert$	$\lvert S \rvert$	ξ
Deterioration rate (DR_{d15})	-	31	15	465	$8.6 \cdot 10^{-3}$
Deterioration rate (DR_{d30})	-	31	30	930	$2.1 \cdot 10^{-4}$
Parametric ($PAR_{K50-d40}$)	50	-	40	2,000	$7.1 \cdot 10^{-2}$
Parametric ($PAR_{K50-d80}$)	50	-	80	4,000	$7.2 \cdot 10^{-3}$
Parametric ($PAR_{K50-d160}$)	50	-	160	8,000	$3.4 \cdot 10^{-3}$
Parametric ($PAR_{K100-d80}$)	100	-	80	8,000	$2.5 \cdot 10^{-3}$
Parametric ($PAR_{K100-d160}$)	100	-	160	16,000	$4.3 \cdot 10^{-4}$

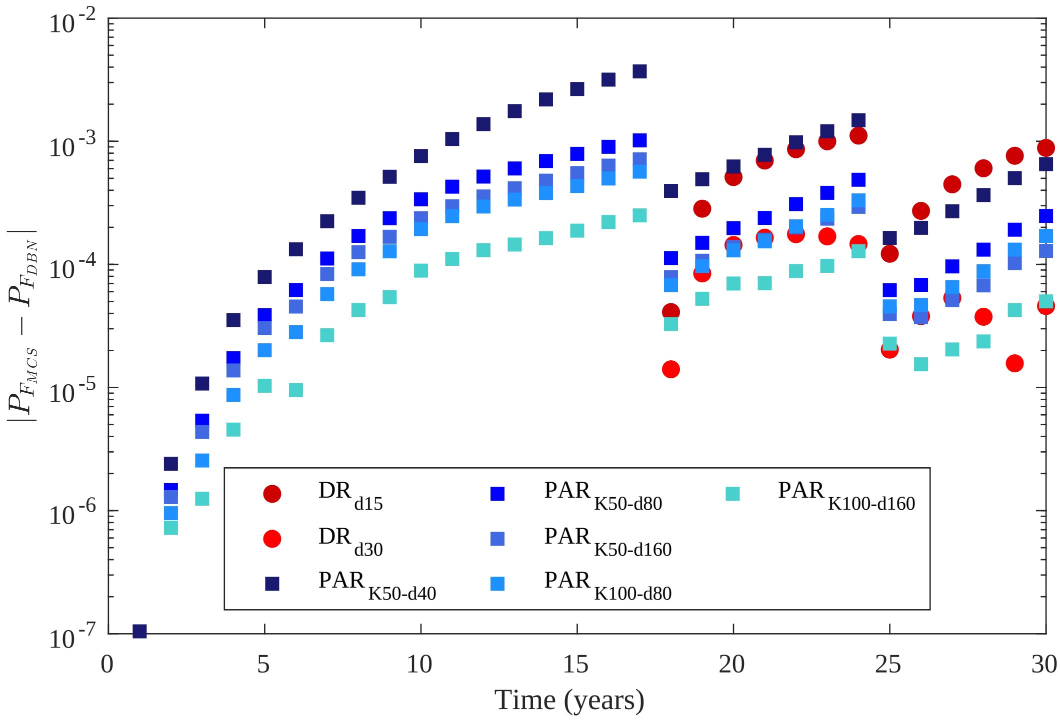


Fig. 2.6. Error $|P_{F_{MCS}} - P_{F_{DBN}}|$ between the continuous deterioration model and the considered discrete space models. The continuous model is computed by a Monte Carlo simulation of 10 million samples and is compared with discrete state-space DBN models. The circles in the graph represent the error from deterioration rate models and the squares represent the error from parametric models.

In general, the selection of the discretized model will depend on the available computational resources and required accuracy. For this application, the deterioration rate model with 930 states is utilized for the numerical experiments, due to its reduced state space as compared to the parametric models.

2.5.2 Case 1. Traditional I&M planning setting

The fatigue deterioration is modeled according to the time-invariant crack growth described at the beginning of Section 2.5. In this traditional setting, the decision maker is only allowed to control the deterioration by undertaking a perfect repair and is able to collect observations through one inspection technique type. The perfect repair returns the component to its initial condition d_0 and the quality of the inspection technique is quantified with a $PoD(d) \sim Exp[\mu = 8]$. This I&M decision-making problem is solved here by both POMDPs and heuristics. For the case of POMDPs, point-based solvers provide a theoretical guarantee to optimality, whereas RBI approaches can analytically compute the $\mathbf{E}[C_T]$ from a simplified decision tree, as explained in Section 2.2.

Alternatively, the computation of the $\mathbf{E}[C_T]$ can be performed in a simulation environment, in which the deterioration process is modeled by DBNs and the costs are evaluated according to the predefined heuristic policies, as shown in Eq. 2.11. To equally compare the policies generated by POMDP and heuristics, the total expected costs $\mathbf{E}[C_T]$ are computed both on an analytical basis and in a simulation environment.

Analytical comparison

Following the results of the discretization analysis, a finite horizon (FH) POMDP is derived from the deterioration rate model with 930 states ($|S_d| = 30$ and $|S_\tau| = 31$). Since the horizon spans over 30 years, the state space is augmented from 930 to 14,880 states, as explained in Section 2.4.2. Actions and observations are combined into three action-observation groups: (1) do-nothing/no-inspection, (2) do-nothing/inspection, and (3) perfect-repair/no-inspection. The fourth combination (repair/inspection) is not included as it will hardly be the optimal action at any time step. A total of three representative experiments are conducted, assigning different inspection, repair and failure costs to each of them. Each experiment is characterized by a different ratio between repair and inspection costs $R_{R/I}$, as well as the ratio between failure and repair costs $R_{F/R}$. Since these ratios are of relevance in this work, analyzing the problem from an optimization perspective, an explicit separation of economic, societal, and environmental consequences and their scaling to monetary units is not considered. The SARSOP point-based POMDP solver [31] is used for the computation of the optimal I&M policies. Additionally, the policies from FRTDP [32] and Perseus [75] point-based solvers are computed specifically for experiment $R_{R/I}50 - R_{F/R}20$. In this theoretical comparison, the expected costs are computed based on the lower bound alpha vectors, as explained in Section 2.4.2.

In contrast, the optimal RBI policies are determined based on the best identified heuristic decision rules. For this theoretical comparison, the decision tree is simplified to a single branch with two schemes considered here: equidistant inspections (EQ-INS) and annual failure probability ΔP_F threshold (THR-INS). For the maintenance actions, the component is perfectly repaired after a detection indication, behaving thereafter as if a crack was not detected at that inspection. The optimized heuristics for each experiment are listed in Table 2.4, e.g. an inspection every 4 years ($\Delta_{Ins} = 4$) is identified as the optimal equidistant inspection heuristic (EQ-INS) for Experiment $R_{R/I}20 - R_{F/R}100$.

The total expected cost $\mathbf{E}[C_T]$ resulting from finite horizon POMDPs and the best identified heuristics are listed in Table 2.4. Along with the $\mathbf{E}[C_T]$, the relative difference between each method and the finite horizon POMDP is also reported, and Table 2.4 demonstrates that finite horizon POMDP policies outperform heuristic-based policies.

Table 2.4. Analytical (AN) and simulation-based (SIM) comparison between POMDPs and optimized heuristic-based policies in a traditional setting. $\mathbf{E}[C_T]$ is the total expected cost and Δ%[POMDP FH] indicates the relative difference between each method and SARSOP finite horizon POMDP. Confidence intervals on the expected costs, assuming Gaussian estimators, are listed for the simulation-based cases.

Traditional setting	$\mathbf{E}[C_T]$ (95%$C.I$)	Δ%[POMDP FH]
Experiment $\mathbf{R_{R/I}20 - R_{F/R}100}$		
$C_i = 5$, $C_r = 10^2$, $C_f = 10^4$, $\gamma = 0.95$		
AN: POMDP FH. SARSOP - Lower bound	58.35	-
AN: Heur.* EQ-INS $\Delta_{Ins} = 4$	69.17	+18%
AN: Heur.* THR-INS $\Delta P_{F_{th}} = 3 \cdot 10^{-4}$	65.62	+12%
SIM: POMDP IH. SARSOP - 30 years**	60.23 (±0.76)	+3%
SIM: Heur. EQ-INS $\Delta_{Ins} = 4$	69.02 (±0.83)	+18%
SIM: Heur. THR-INS $\Delta P_{F_{th}} = 3 \cdot 10^{-4}$	64.81 (±0.75)	+11%
Experiment $\mathbf{R_{R/I}10 - R_{F/R}10}$		
$C_i = 1$, $C_r = 10$, $C_f = 10^2$, $\gamma = 0.95$		
AN: POMDP FH. SARSOP - Lower Bound	2.25	-
AN: Heur.* EQ-INS no inspections	2.25	+0%
AN: Heur.* THR-INS no inspections	2.25	+0%
SIM: POMDP IH. SARSOP - 30 years**	2.50 (±0.02)	+11%
SIM: Heur. EQ-INS no inspections	2.25 (±0.00)	+0%
SIM: Heur. THR-INS no inspections	2.25 (±0.00)	+0%
Experiment $\mathbf{R_{R/I}50 - R_{F/R}20}$		
$C_i = 1$, $C_r = 50$, $C_f = 10^3$, $\gamma = 0.95$		
AN: POMDP FH. SARSOP - Lower Bound	12.45	-
AN: POMDP FH. FRTDP - Lower Bound	12.45	+0%
AN: POMDP FH. PERSEUS - Lower Bound	12.96	+4%
AN: Heur.* EQ-INS $\Delta_{Ins} = 11$	17.06	+37%
AN: Heur.* THR-INS $\Delta P_{F_{th}} = 1 \cdot 10^{-3}$	16.69	+34%
SIM: POMDP IH (DR). SARSOP - 30 years**	12.99 (±0.24)	+4%
SIM: POMDP IH (PAR). SARSOP - 30 years**	13.08 (±0.23)	+5%
SIM: Heur. EQ-INS $\Delta_{Ins} = 11$	16.28 (±0.19)	+31%
SIM: Heur. THR-INS $\Delta P_{F_{th}} = 1.5 \cdot 10^{-3}$	16.43 (±0.20)	+32%
SIM: Heur. EQ-INS*** $\Delta_{Ins} = 5$	14.17 (±0.26)	+14%
SIM: Heur. THR-INS*** $\Delta P_{F_{th}} = 8 \cdot 10^{-4}$	13.29 (±0.23)	+7%

* The decision tree is simplified to one single branch, as explained in Section 2.2.1.

** Simulation of an infinite horizon POMDP policy over a horizon of 30 years.

*** Perfect repair actions are undertaken after two consecutive 'detection' observations.

(FH) Finite horizon; (IH) Infinite horizon.

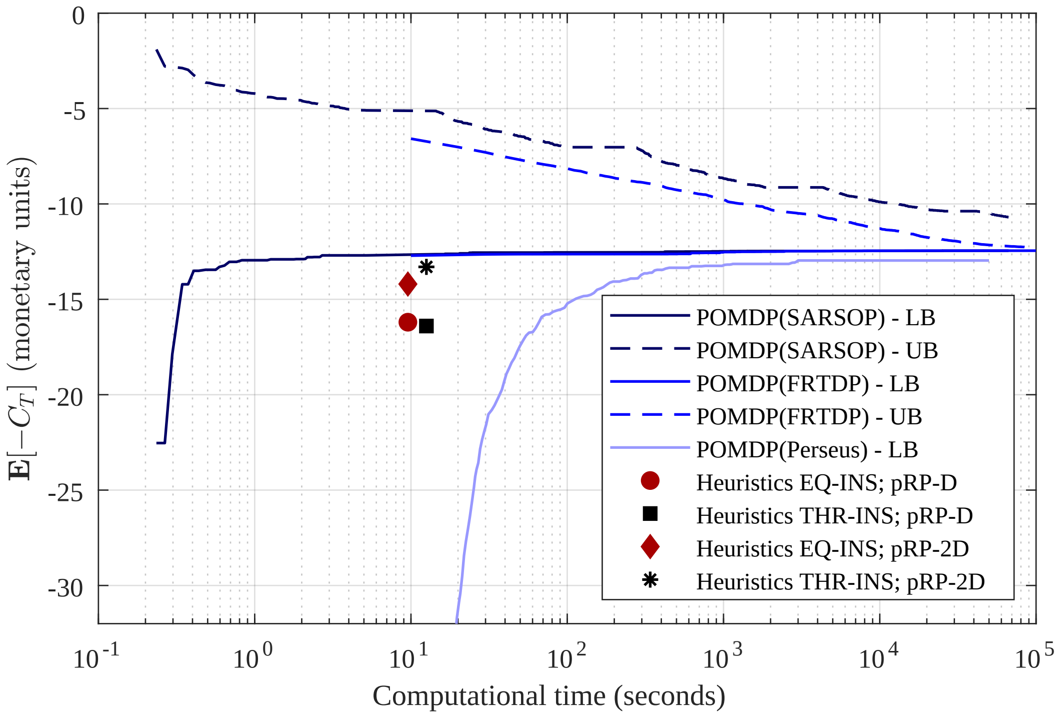


Fig. 2.7. Point-based POMDP solutions for Experiment $R_{R/I}50 - R_{F/R}20$. The expected total cost $\mathbf{E}[C_T]$ is represented over the computational time. Results of SARSOP, FRTDP and Perseus point-based POMDP solvers are plotted, with a continuous line for the low bound and a dashed line for the upper bound. Optimized heuristic methods are represented by markers; the equidistant inspection planning scheme in red, and the annual failure probability threshold in black. The markers also indicate whether the investigated heuristics plan performs repair after observing one detection outcome, $pRP - D$, or after the collection of two consecutive detection outcomes, $pRP - 2D$.

Even for this traditional I&M decision-making problem, POMDPs provide a significant cost reduction ranging from 11% in Experiment $R_{R/I}20 - R_{F/R}100$ to 37% reduction in Experiment $R_{R/I}50 - R_{F/R}20$. Experiment $R_{R/I}10 - R_{F/R}10$ is merely conducted to validate the comparative results by checking that all the methods provide the same results for the case in which repairs and inspections are very expensive relatively to the failure cost.

As pointed out in Section 2.4.2, point-based solvers are able to rapidly solve large state-space POMDPs. This is demonstrated in Fig. 2.7, where SARSOP outperforms heuristic-based schemes in less than one second of computational time. Note that POMDP policies are based on the lower bound, whereas the upper bound, when provided, is just an approximation, to optimally sample reachable belief points [24].

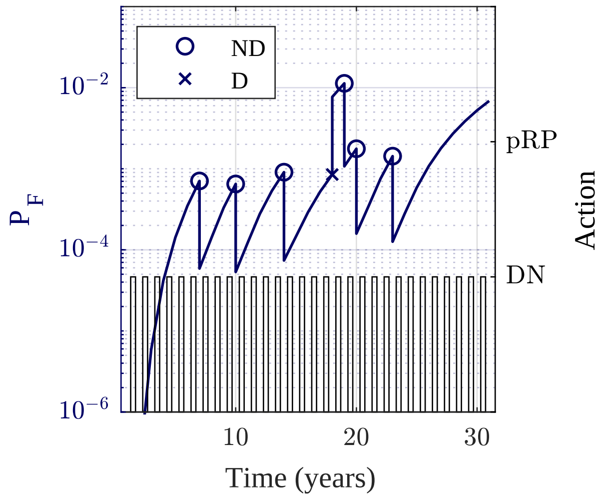
Comparison in a simulation environment

In this case, the total expected cost $\mathbf{E}[C_T]$ is evaluated in a simulation environment. Since the horizon can be controlled in a policy evaluation, infinite horizon POMDPs are also included in this comparison. The infinite horizon POMDP is directly derived from the deterioration rate model, and while the action-observation combinations remain the same as for the finite horizon POMDP, the belief space is now reduced to 930 states, offering a substantial reduction in computational cost, as explained before.
Note that even though policies generated by infinite horizon POMDPs can be evaluated over a finite horizon, the policies are truly optimal only in an infinite horizon setting.

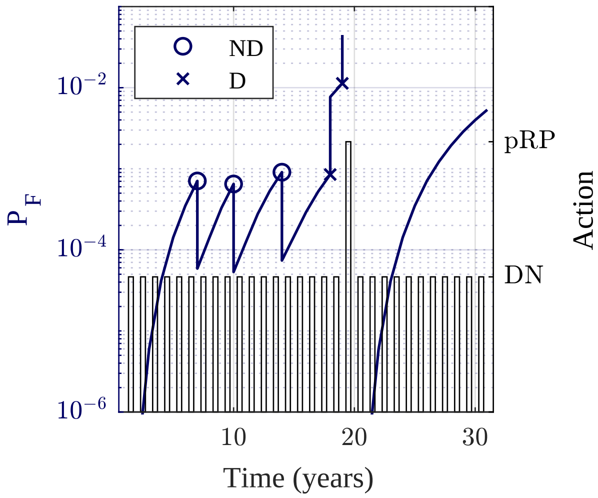
In this comparison, the best heuristic-based I&M policy is also identified by analyzing two inspection planning heuristics, as previously, either based on equidistant inspections (EQ-INS) or based on an annual failure probability threshold (THR-INS). However, in this simulation setting, the component naturally returns to its initial condition after a repair, instead of modeling its evolution as a no-detection event. This operation might add a significant computational expense for analytical computations, if the decision tree is explicitly modeled; however, it can be easily modeled in a simulation-based environment. The expected utility $\mathbf{E}[C_T]$ is estimated according to Eq. 2.11.

Table 2.4 lists the results of the comparison and given that the expected cost $\mathbf{E}[C_T]$ is estimated through simulations, the numerical confidence bounds are also reported, assuming a Gaussian estimator. All the methods are compared relatively to the finite horizon POMDP that again outperforms the heuristic-based policies. The reduced state-space infinite horizon POMDP policy results in only a slight increment to the total expected cost obtained by the finite horizon POMDP, in this finite horizon problem. The optimal policy for an infinite horizon in experiment $R_{R/I}20 - R_{F/R}100$ includes the possibility of maintenance actions, whereas the policy for a finite horizon prescribes only the action do-nothing/no-inspection. This explains the slight difference of expected costs for the infinite horizon POMDP. The infinite horizon POMDP for a parametric model of 16,000 states is also computed and listed in Table 2.4 for the experiment $R_{R/I}50 - R_{F/R}20$. As expected ,the $\mathbf{E}[C_T]$ for the parametric (PAR) model results in good agreement with the deterioration rate (DR) model and the small difference is attributed to the discretization quality.

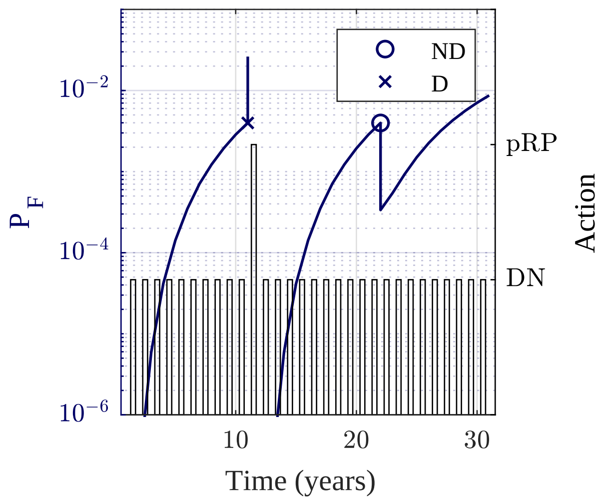
Finally, we showcase policy realizations to visualize the difference between POMDPs and heuristic-based policies over an episode, related to the experiment $R_{R/I}50 - R_{F/R}20$. Fig. 2.8a and Fig. 2.8b represent realizations of POMDP policies, whereas, Fig. 2.8c and Fig. 2.8d represent the realizations of heuristic-based policies.



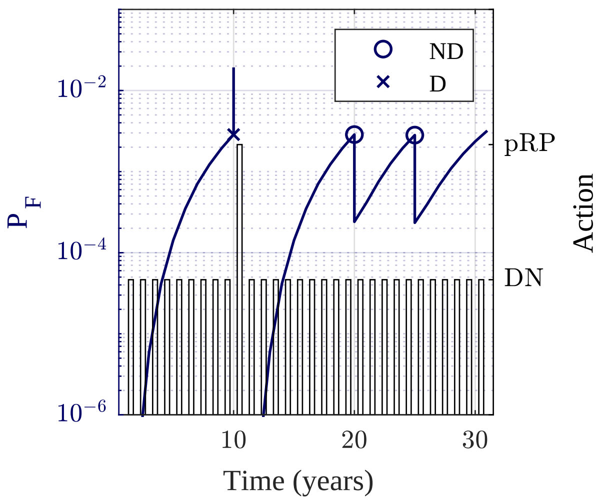
(a) POMDP policy realization



(b) POMDP policy realization



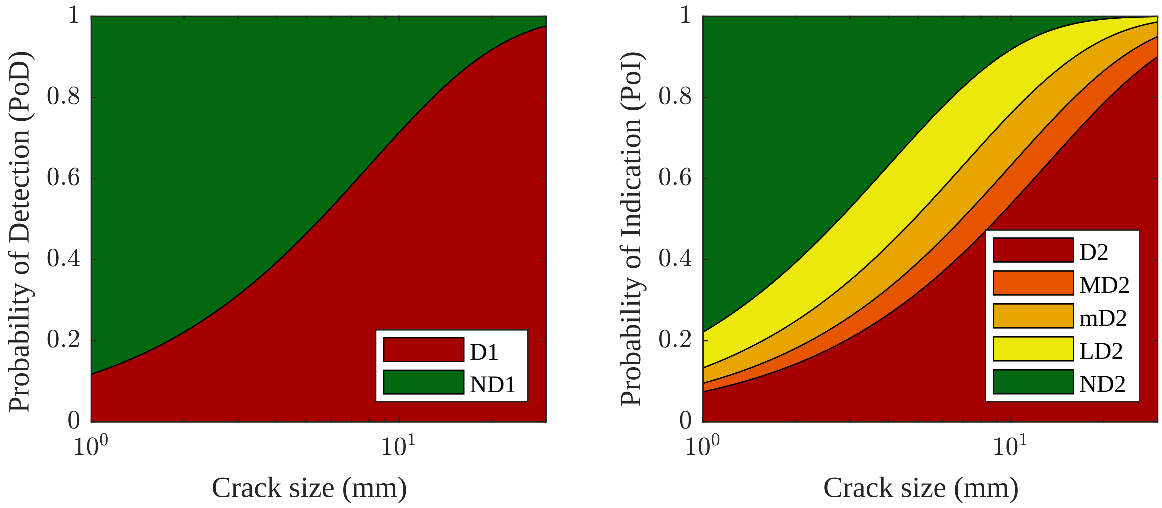
(c) Equidistant inspection policy realization



(d) ΔP_F threshold policy realization

Fig. 2.8. Experiment $R_{R/I}50 - R_{F/R}20$ policy realizations. The failure probability is plotted in blue and the prescribed maintenance actions are represented by black bars. A detection outcome is marked by a cross, whereas a no-detection outcome is marked by a circle.

While heuristic-based policies prescribe a repair action immediately after a detection, POMDP-based policies might also consider a second inspection after a detection outcome. If the second inspection results in a no-detection outcome, a repair action may not be prescribed; however, if the second inspection also results in detection, a perfect repair is planned. POMDP-based policies provide, therefore, more flexibility, in general, and can reveal interesting patterns, such that it might be worthy, in certain cases, to conduct a second inspection before prescribing an expensive repair action. As such, based on analyzed POMDP policy patterns, heuristic rules can be informed and defined anew.



(a) Inspection type 1 (2 indicators).

(b) Inspection type 2 (5 indicators).

Fig. 2.9. Quantification of the inspection uncertainty. The probability of retrieving each indicator is represented as a function of the crack size. For inspection type-1 , the observation model includes two indicators: "detection" D1 and "no-detection" ND1. For inspection type-2 , the observation model is composed of five indicators: "no-detection" ND2, "low damage" LD2, "minor damage" mD2, "major damage" MD2, and "extensive damage" D2.

As reported in Table 2.4, two additional heuristic rules are thus examined, where perfect repair actions are undertaken after two consecutive 'detection' observations. These modified heuristics yield results closer to those provided by POMDP policies, with POMDP policies surpassing now the two heuristic ones by 7% and 14%, respectively. While an experienced operator might have initially guessed these more sophisticated heuristic decision rules, based on the imperfect and cheap observation model specified in this setting, in more complex settings, e.g. an I&M planning scenario with inspections that provide more than two indications (as shown in Section 2.5.3), decision makers might guide their choices for the selection of more advanced heuristic rules through an investigation of the patterns exposed by POMDP policy realizations.

2.5.3 Case 2. Detailed I&M planning setting

While only a perfect repair and one inspection technique have been available for the traditional setting applications, two repair actions and two inspection techniques are now available in this more complex case. Fatigue deterioration in this setting can be controlled by either performing a perfect or a minor repair. The perfect repair returns the component to its initial condition and the minor repair transfers the component two deterioration rates back. The two inspection techniques considered are inpection 1 (I1) with only 2 indicators: detection (D) or no-detection (ND); and inspection 2 (I2) with 5 indicators: no-detection (ND), low damage (LD), minor damage (mD), major damage (MD) and extensive damage (D). The quality of each inspection technique is quantified through probability of indication (PoI) curves. Fig. 2.9a corresponds to the first inspection type with a $PoD(d) \sim Exp[\mu = 8]$. This inspection method is the same as the one used in the traditional I&M planning setting. The second inspection method includes, however, the following detection boundaries: $PoI(d) \sim Exp[\mu = 4]$; $PoI(d) \sim Exp[\mu = 7]$; $PoI(d) \sim Exp[\mu = 10]$; and $PoI(d) \sim Exp[\mu = 13]$. The probability of observing each indicator is represented in Fig. 2.9b as a function of the crack size.

Similar to the previous case, we solve a finite horizon POMDP with 14,880 states to identify the optimal policy. However, in this setting, actions and observations are combined into seven groups: (1) do-nothing/no-inspection (DN-NI); (2) do-nothing/inspection-1 (DN-I1); (3) do-nothing/inspection-2 (DN-I2); (4) minor-repair/no-inspection (mRP-NI); (5) minor-repair/inspection-1 (mRP-I1); (6) minor-repair/inspection-2 (mRP-I2); and (7) perfect-repair / no-inspection (pRP-NI), and analyses are conducted for a modified version of experiment $R_{R/I}50 - R_{F/R}20$. The individual costs for this example are listed in Table 2.5. Inspection type-2 costs twice the cost of inspection type-1, as it is more accurate and provides more information about the deterioration.

Table 2.5. Comparison between POMDP and optimized heuristic-based policies in a detailed setting. $\mathbf{E}[C_T]$ is the total expected cost and $\Delta\%$[POMDP FH] indicates the relative difference between each method and SARSOP finite horizon POMDP results. Confidence intervals on the expected costs, assuming Gaussian estimators, are also listed.

Detailed setting	$\mathbf{E}[C_T](95\%C.I)$	$\Delta\%$[POMDP FH]
$C_{i_1} = 1$, $C_{i_2} = 2$, $C_{mRP} = 10$		
$C_{pRP} = 50$, $C_f = 10^3$, $\gamma = 0.95$		
POMDP FH. SARSOP - Lower Bound	12.26	-
POMDP FH. FRTDP - Lower Bound	12.30	<1%
Heur. EQ-INS1 $\Delta_{Ins} = 11$; pRP-$D1$	16.23 (±0.19)	+32%
Heur. EQ-INS2 $\Delta_{Ins} = 11$; pRP-$D2$	18.08 (±0.31)	+47%
Heur. THR-INS1 $\Delta P_{F_{th}} = 1.5 \cdot 10^{-3}$; pRP-$D1$	16.40 (±0.20)	+33%
Heur. THR-INS2 $\Delta P_{F_{th}} = 1.1 \cdot 10^{-3}$; pRP-$D2$	15.55 (±0.21)	+26%
Heur. THR-INS2 $\Delta P_{F_{th}} = 5.0 \cdot 10^{-4}$; pRP-$P_{F_{th}}{}^{*}$	13.88 (±0.29)	+13%
Heur. THR-INS2 $\Delta P_{F_{th}} = 1.0 \cdot 10^{-3}$; pRP^{**}	13.66 (±0.24)	+11%

* pRP-$P_{F_{th}} = 2.2 \cdot 10^{-2}$

** pRP-$\mathbf{E}[d] > 4$

(FH) Finite horizon.

For this setting, heuristic inspection decision rules are prescribed considering again both equidistant inspections and annual failure probability ΔP_F threshold schemes. All heuristics are evaluated in a simulation environment, computing the expected cost $\mathbf{E}[C_T]$, as indicated in Eq. 2.11. Maintenance heuristic rules are accordingly defined considering the following two schemes:

- Observation-based maintenance rules: a maintenance action is undertaken after an observation. For example, a minor repair is undertaken if a minor damage is observed. The number of potential observation-based maintenance rules scales to $|A_R|^{|O|}$ pairs, where, $|O|$ and $|A_R|$ are the number of observations and maintenance actions, respectively. If we consider inspection type-2, the heuristic rules result in 3^5 combinations. Such combinatoric heuristic rules, together with failure probability thresholds or intervals for inspections, have been evaluated against POMDPs in [29]. Due to the large computational cost of evaluating all possible decision rules, we evaluated only a subset of these combinations here. The most competitive set of heuristic rules for this case are listed in Table 2.5, e.g. the optimized equidistant inspection type-1 heuristic (EQ-INS1) prescribes an inspection every 11 years ($\Delta_{Ins} = 11$), and a perfect repair after a detection observation (pRP-$D1$).

- Threshold-based maintenance rules: a maintenance action is undertaken when a specific threshold is reached after an observation. The threshold can be prescribed in terms of failure probability P_F or expected damage size, as proposed in [9]. We consider both cases here, i.e. a failure probability threshold $P_{F_{th}}$ and an expected damage size threshold, $\mathbf{E}[d]$. Threshold-based maintenance rules based on expected damage have also been evaluated against POMDPs in [28].

The expected costs $\mathbf{E}[C_T]$ resulting from both POMDP and heuristic-based policies are reported in Table 2.5. Additionally, we list the relative difference between each policy and a finite horizon POMDP policy solved by SARSOP. In this detailed setting, POMDP-based policies outperform again heuristic-based ones. In terms of POMDP-based policies, SARSOP and FRTDP achieve similar results. Results obtained from heuristic-based policies vary depending on their prescribed set of heuristics. For equidistant inspection planning, inspection type-1 is preferred rather than inspection type-2, because the inspections are fixed in time, and the additional information provided by inspection type-2 becomes too expensive. In contrast, inspection type-2 is the best scheme for annual failure probability threshold inspection planning. The threshold-based maintenance heuristics proved to be better than observation-based heuristics, yet threshold-based maintenance heuristics imply additional computational costs, as generally, more heuristic rules must be evaluated.

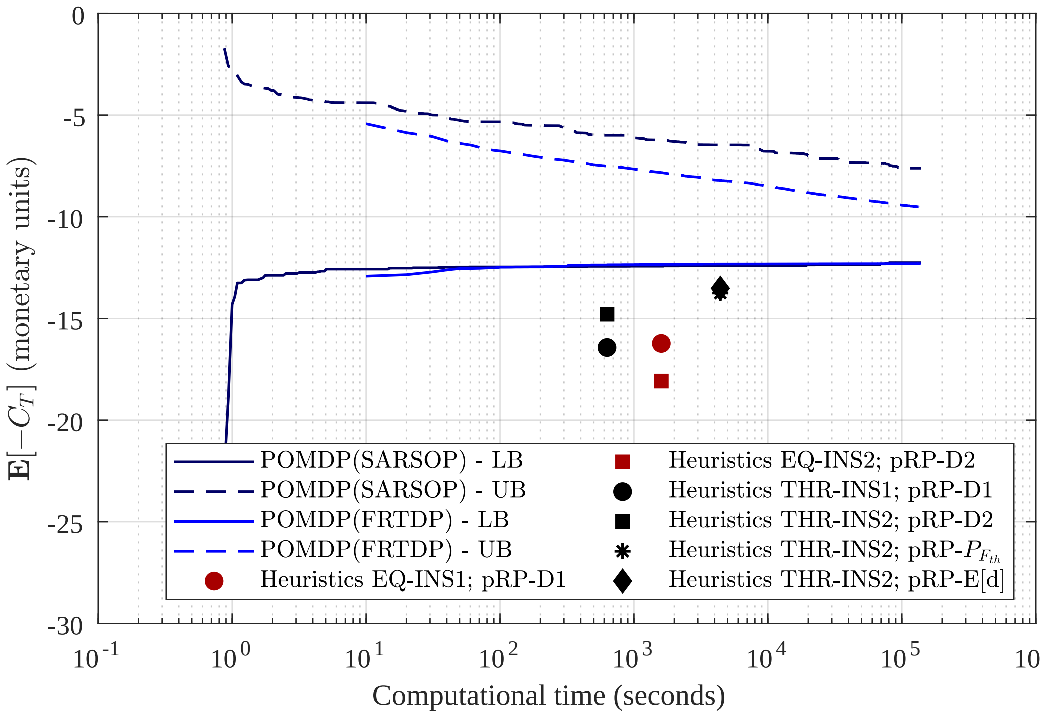


Fig. 2.10. Computational details of POMDP and simulation-based heuristic schemes in a detailed setting. The expected total costs $\mathbf{E}[C_T]$ are represented over the computational time. Results of SARSOP and FRTDP point-based POMDP solvers are plotted, with a continuous line for the low bound and a dashed line for the upper bound. Optimized heuristic policies results are reported by markers and are directly linked to the schemes shown in Table 2.5.

Fig. 2.10 illustrates the expected cost $\mathbf{E}[C_T]$ of each policy as a function of the computational time. We can see how the POMDP point-based solvers improve their low bounds in time, along with the computational cost incurred by evaluating the various heuristic rules.

To visualize the actions prescribed by each approach, Fig. 2.11 displays a frequency histogram of the actions taken over 10^4 policy realizations. The action do-nothing/no-inspecion (DN-NI) predominates over all other actions. While heuristic policies conduct either inspection type-1 (DN-I1) or inspection type-2 (DN-I2), the POMDP-based policy utilizes both inspection types. This is also true for the maintenance actions, in which heuristic policies prescribe only perfect repairs, whereas POMDP policies choose sometimes to undertake minor-repairs (mRP) as well.

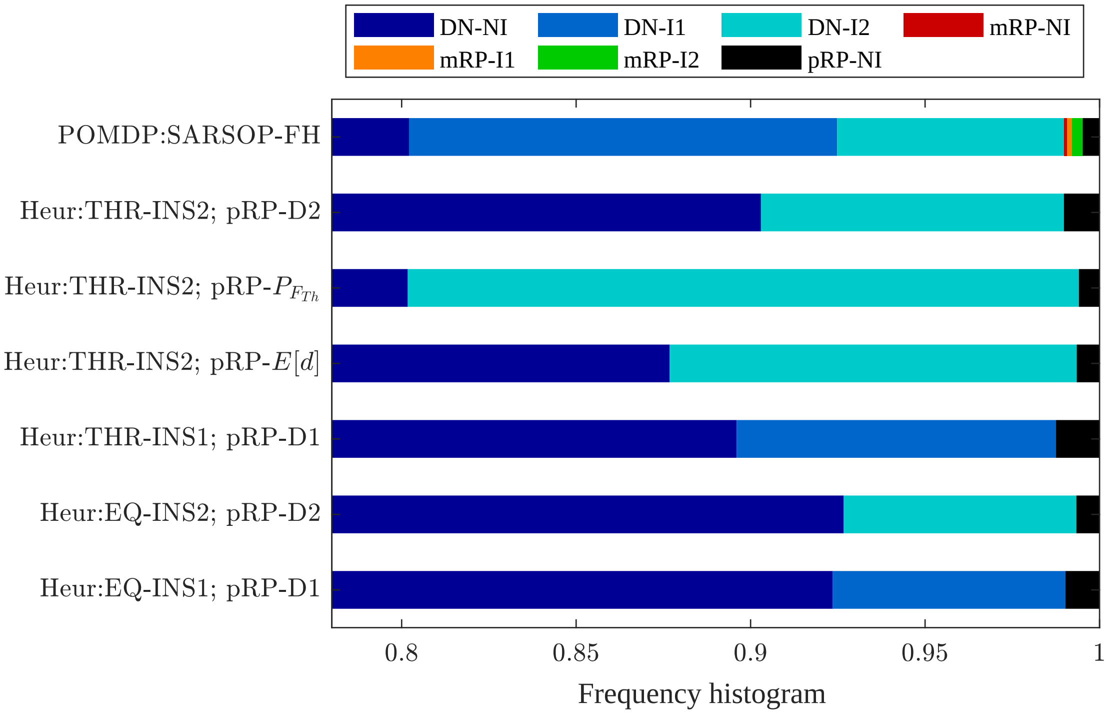


Fig. 2.11. Frequency histogram of the actions prescribed by each considered approach over 10^4 policy realizations. The policies presented here are linked to those listed in Table 2.5.

2.6 Discussion

The results of this investigation show that POMDPs are able to identify optimal I&M policies for deteriorating structures and offer substantially lower costs than heuristic-based policies, as is theoretically explained and justified, and as it has also been demonstrated through numerical examples in Sections 2.5.2 and 2.5.3. The policy optimization based on heuristic-based approaches may be constrained by the limited number of decision rules assessed, out of all possible decision rules. Avoiding these limitations, POMDPs prescribe actions as a function of the belief state, which is a sufficient statistic of the whole, dynamically updated, action-observation history. This implies that the actions are taken according to the whole history of actions and observations, rather than as a result of an immediate inspection outcome or pre-defined static policies.

As demonstrated in Section 2.5.3, POMDPs can be applied to detailed I&M decision settings, in which multiple actions and inspection methods are available. In terms of computational efficiency, state-of-the-art point-based solvers are able to solve high-dimensional state space POMDPs within a reasonable computational time. In particular, SARSOP point-based solver very quickly improves its policy at the beginning of the solution process and employs an approximate upper bound to gradually reach a converged solution. For both traditional and detailed settings, both SARSOP and FRTDP point-based solvers outperform heuristic-based policies after only few seconds of computational time.

For modeling the deterioration process, one can utilize either a parametric or a deterioration rate model, as explained in Section 2.2. A deterioration rate model generally results in a smaller state space than a parametric model, except for very long horizons. In this latter case, a parametric model might lead to a smaller state space, due to its stationary nature. In any case, a discretization analysis must be conducted to select the appropriate state model for the problem at hand. More efforts are worth being made in the future towards continuous state space POMDPs and optimal discretization schemes for discrete state spaces.

2.7 Concluding remarks

In this paper, we examine the effectiveness of Partially Observable Markov Decision Processes (POMDPs) to identify optimal Inspection and Maintenance (I&M) strategies for deteriorating structures, and we clarify that Dynamic Bayesian Networks (DBNs) can be combined with POMDPs, providing a joint framework for efficient inspection and maintenance planning. The formulation for deriving POMDPs in a structural reliability context is also presented, and two alternative DBN formulations for deterioration modeling are described, together with their POMDP implementations.

Modern Risk Based Inspection (RBI) planning methodologies are often supported by DBNs, and a pre-defined set of decision rules is evaluated. These policies can on occasions diverge significantly from globally optimal solutions, because of the limited domain space of searched policies that may not include the global optimum. In contrast, POMDP policies prescribe an action as a function of the belief state, which is a sufficient statistic of the whole action-observation history.

I&M policies generated by finite horizon POMDPs are compared with heuristic-based policies, for the case of a structural component subjected to fatigue deterioration. In the first example, the stochastic deterioration is modeled as a function of time-invariant parameters, with only one inspection type and one perfect repair available. Our numerical findings verify that POMDP-based policies can approximate the global solution better than heuristic-based policies, thus being more efficient even for typical RBI applications. The 14,880 states finite-horizon POMDP outperforms heuristic-based policies in less than a second of computational time. For the second numerical example, we consider an I&M decision-making problem in a more detailed setting, including two inspection methods and two repair actions. Whereas the outcome of the first inspection type is set up as a binary indicator, the second inspection technique indicates the damage level through five alarms.

With this application, we demonstrate the capabilities of POMDPs in efficiently handling complex decision problems, outperforming again heuristic-based polices.

The main limitation of the presented approaches, including POMDPs, is the increase of computational complexity for very large state and action spaces, such as the ones for a system of multiple components. Dynamic Bayesian networks with large state spaces are similarly constrained by the curse of dimensionality. To overcome this limitation, we suggest further research efforts toward the development of POMDP-based Deep Reinforcement Learning (DRL) methodologies. As demonstrated in [28, 29], a multi-agent actor-critic DRL approach is able to identify optimal strategies for multi-component systems with large state, action and observation spaces. In particular, POMDP-based actor-critic DRL methods approximate the policy and the value function with neural networks, alleviating therefore the curse of dimensionality through the deep networks parametrizations, and the curse of history through the reliance on dynamic programming MDP principles, the full advantages of which may be compromised if heuristic rules are instead considered.

Authorship contribution statement

Morato, P. G.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Papakonstantinou, K. G.:** Conceptualization, Methodology, Validation, Formal analysis, Writing - Review & Editing, Supervision. **Andriotis, C. P.:** Software, Methodology, Validation, Formal analysis, Writing - Review & Editing. **Nielsen, J. S.:** Writing - Review & Editing. **Rigo P.:** Supervision, Project administration.

Acknowledgements

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 and Dr. Andriotis would further like to acknowledge that this material is also based upon work supported by the U.S. National Science Foundation under Grant No. 1751941.

Chapter 3

Managing Offshore Wind Turbines through Markov Decision Processes and Dynamic Bayesian Networks

Paper Morato, P. G., Papakonstantinou, K. G., Andriotis, C. P. and Rigo P. (2021). Managing Offshore Wind Turbines through Markov Decision Processes and Dynamic Bayesian Networks. In *Proc. ICOSSAR 2021*, under internal review.

Abstract Efficient planning of Inspection and Maintenance (I&M) actions in civil and maritime environments is of paramount importance to balance management costs against failure risk, caused by deteriorating mechanisms. Determining I&M policies for such cases constitutes a complex sequential decision-making optimization problem under uncertainty. Addressing this complexity, POMDPs provide a principled mathematical methodology for stochastic optimal control, in which the optimal actions are prescribed as a function of the entire, dynamically updated, state probability distribution. As shown, by integrating dynamic Bayesian networks with Partially Observable Markov Decision Processes (POMDPs), advanced algorithmic schemes of probabilistic inference and decision optimization under uncertainty, respectively, can be combined into an efficient planning platform. To demonstrate the capabilities of the proposed approach, POMDP and heuristic-based I&M policies are compared, with emphasis on an offshore wind substructure subject to fatigue deterioration. Results verify that POMDP solutions offer substantially reduced costs compared to their counterparts, even in traditional problem settings.

3.1 Introduction

Civil and maritime infrastructures are exposed to deterioration mechanisms, such as fatigue or corrosion, thereby constituting a risk of structural failure. Deterioration models are, nonetheless, intrinsically uncertain, characterized by model and load uncertainties reaching coefficients of variation in the order of 25-30% [79]. In-service inspection and maintenance planning, i.e. collecting information through inspections and undertaking maintenance actions when needed, becomes therefore of paramount importance to optimally manage such systems throughout their lifetime. To this end, inspection and maintenance (I&M) planning targets the identification of a strategy able to optimally balance the risk of structural failure against inspection and maintenance efforts. Finding an optimal I&M policy demands, however, in most practical cases, the solution of a complex sequential decision-making problem under uncertainty.

Originally targeted to the management of oil and gas platforms, risk-based inspection planning approaches simplify the I&M decision problem by evaluating only a predefined subset of heuristic rules out of all possible policies, thus alleviating the computational complexity [61]. Modern risk-based I&M planning methods evaluate the set of prescribed heuristic rules in a simulation environment, conducting Bayesian inference via dynamic Bayesian networks [6, 9, 11]. Heuristic-based policies are however compromised by the limited number of explored and evaluated policies out of an immense policy space.

In contrast, Partially observable Markov decision processes (POMDPs) constitute a principled mathematical framework for sequential decision-making under uncertainty, in which the policy is defined as a function of a sufficient statistic, i.e. the dynamically updated history of actions and observations. Recent works on POMDPs for infrastructure management can be found in [21, 22, 80]. With the advent of point-based solvers in this class of applications [24], POMDP-based policies can be efficiently traced for medium-to-large deteriorating problems [22, 81].

We adopt in this paper the methodology proposed in our earlier work [81], integrating dynamic Bayesian networks (DBNs) into the underlying structure of partially observable Markov decision processes (POMDPs), and we apply it for optimally managing an offshore wind structural detail subject to fatigue deterioration. Formulation schemes are described for encoding non-stationary stochastic deterioration processes in models parameterized by the influencing random variables or in terms of the deterioration rate. Both parametric and deterioration rate POMDP models are built based on the fatigue deterioration mechanism experienced by an offshore structural detail. POMDP and heuristic-based policies are then computed and thoroughly compared for typical I&M and lifetime extension planning settings, and results verify that POMDP solutions offer substantially reduced costs in all the explored settings.

3.2 Joint DBN-POMDP framework

3.2.1 Stochastic deterioration processes via DBNs and heuristic decision rules

The evolution of the stochastic deterioration process experienced by a structural component can be quantified in terms of a group of influencing random variables or random processes. DBNs encode the relationship amongst the involved random variables through conditional structures, enabling efficient inference, e.g. updating the deterioration process based on inspection outcomes. In most cases, the involved random variables are continuous and must be properly discretized in order to guarantee exact inference [7].

A parametric DBN structure encodes the deterioration d conditional on a set of random variables $\boldsymbol{\theta}$. In this case, damage d_{t+1}, at time $t+1$, evolves conditional on the damage at the previous time, d_t, set of random variables $\boldsymbol{\theta}$ and observations $\boldsymbol{o_0}, ..., \boldsymbol{o_t}$:

$$P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) = \sum_{d_t} \sum_{\boldsymbol{\theta_t}} P(d_{t+1}, \boldsymbol{\theta_{t+1}} | d_t, \boldsymbol{\theta_t}) P(d_t, \boldsymbol{\theta_t} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{3.1}$$

After collecting an observation o_{t+1} with likelihood $P(\boldsymbol{o_{t+1}} | d_{t+1})$, the deterioration process, conditional on all observations up to time $t+1$, can be updated through Bayesian inference:

$$P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_{t+1}}) \propto P(\boldsymbol{o_{t+1}} | d_{t+1}) P(d_{t+1}, \boldsymbol{\theta_{t+1}} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{3.2}$$

The deterioration process can be alternatively encoded in a deterioration rate DBN, tracing the damage evolution d as a function of the deterioration rate τ. In this case, damage d_t at deterioration rate τ_t, conditional on observations $\boldsymbol{o_0}, ..., \boldsymbol{o_t}$, transitions in one time step as:

$$P(d_{t+1}, \tau_{t+1} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) = \sum_{d_t} \sum_{\tau_t} P(d_{t+1}, \tau_{t+1} | d_t, \tau_t) P(d_t, \tau_t | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{3.3}$$

Bayesian inference considering an observation o_{t+1}, with likelihood $P(\boldsymbol{o_{t+1}} | d_{t+1})$, can then be performed as:

$$P(d_{t+1}, \tau_{t+1} | \boldsymbol{o_0}, ..., \boldsymbol{o_{t+1}}) \propto P(\boldsymbol{o_{t+1}} | d_{t+1}) P(d_{t+1}, \tau_{t+1} | \boldsymbol{o_0}, ..., \boldsymbol{o_t}) \tag{3.4}$$

In a structural reliability context, the probability of a failure event $P_{F,t}$, at time t, corresponds to the probability of being in a damage state $P(d_{F,t})$. Additionally, an annual risk performance measure can be computed as the failure probability between two successive years, i.e. $\Delta P_{F,t} = P_{F,t+1} - P_{F,t}$.

The risk of structural failure can be controlled through an I&M policy regulated by a set of predefined heuristic decision rules, e.g. equidistant inspections or planned maintenance after an indication event. DBNs models, either parametric, deterioration rate, or others, can be employed, in a simulation environment, to identify the most optimal heuristic from all the set of evaluated decision rules. The total discounted reward $V_{T_i}^{(h)}$, resulting from a set of heuristic decision rules h, can be evaluated for each simulation as the sum of inspection C_i, repair C_r, decommissioning C_d, and failure C_f costs, discounted by the factor γ:

$$V_{T_i}^{(h)} = \sum_{t=t_0}^{t_N} \gamma^t \Big[C_i(t) + C_r(t) + C_d(t) + \Delta P_F(t) C_f \Big] \tag{3.5}$$

The total expected reward $V_T^{(h)}$ can then be computed through a Monte Carlo simulation of n_{ep} episodes (policy realizations):

$$V_T^{(h)} = \frac{\sum_{i=1}^{n_{ep}} \Big[V_{T_i}^{(h)} \Big]}{n_{ep}} \tag{3.6}$$

3.2.2 Optimal I&M planning through partially observable MDPs

Dynamic Bayesian networks (DBNs) can be integrated into the underlying structure of Partially observable Markov decision processes (POMDPs) for optimal inspection and maintenance (I&M) planning, as proposed in [81]. A POMDP is a 7-tuple $\langle S, A, O, T, Z, R, \gamma \rangle$ controlled stochastic process in which the decision maker (intelligent agent) interacts in a stochastic environment. For a more complete overview of POMDP theoretical foundations and detailed formulations, the reader is directed to [21, 22].

The state space S of a POMDP based on a parametric or deterioration rate DBN model is defined as the joint space of $d \times \tau$ or $d \times \boldsymbol{\theta}$, respectively. Figs. 3.1 and 3.2 represent the dynamic decision network corresponding to POMDPs based on parametric and deterioration rate DBN models. The stochastic condition of the deterioration process is tracked by its belief state $\mathbf{b}(s) \equiv P(s)$ or probability distribution over states, and the POMDP dynamics consists, therefore, on an agent taking an action a_t, at time step t, transferring the state $s_t \in S$ to state $s_{t+1} \in S$, according to the transition model $T \equiv P(s_{t+1}|s_t)$. If a maintenance action is not planned, the deterioration process evolves naturally; in this case, the action do-nothing $a_{DN} \in A$ is linked with a transition model T_{DN} defined as $P(d_{t+1}, \boldsymbol{\theta}_{t+1}|d_t, \boldsymbol{\theta}_t)$ or $P(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$, and equivalent to the transition model formulated in Eqs. 3.1 and 3.3.

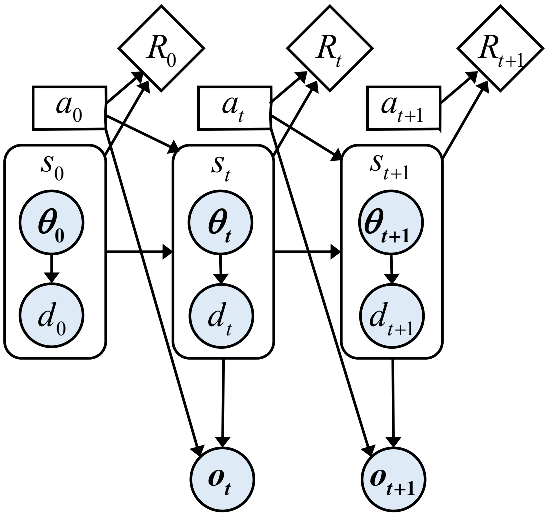


Fig. 3.1. Dynamic decision network of a POMDP built based on a parametric DBN model [81].

A perfect repair maintenance action $a_{PR} \in A$ transfers, instead, the belief $\mathbf{b_t}$ at time step t to its initial belief state $\mathbf{b_0}$:

$$\mathbf{P}(s'|s, a_{PR}) = \begin{pmatrix} b_0(s_0) & b_0(s_1) & \cdots & b_0(s_{|S|}) \\ b_0(s_0) & b_0(s_1) & \cdots & b_0(s_{|S|}) \\ \vdots & \vdots & \ddots & \vdots \\ b_0(s_0) & b_0(s_1) & \cdots & b_0(s_{|S|}) \end{pmatrix} \tag{3.7}$$

The quality of an inspection technique can be quantified through an observation model Z, defined as the probability of collecting an observation $o \in O$ at state $s \in S$. If inspections provides binary indication outcomes, i.e. either observing detection o_D or no-detection o_{ND}, the observation model Z_I can be often deduced as $P(o|s) = PoD(s)$ from Probability of Detection curves, PoD, corresponding to the inspection type. If no inspection is conducted, the observation model Z_{NI} assumes that observation $o_0 \in O$ is collected independently of the state $P(o_0|s_{t+1}) = 1$, thus leaving the belief state unaffected.

The total discounted reward, or sum of discounted rewards, is denoted in POMDP terminology as value function V_T. In a partially observable environment, the reward collected after taking an action a at belief state $\mathbf{b}$ is the average of rewards associated to action a and states $s \in S$:

$$R(\mathbf{b}, a) = \sum_{s \in S} b(s) R(s, a) \tag{3.8}$$

In an I&M framework, both action and inspection actions should be determined and can be combined into maintenance-inspection decision groups. For instance, two maintenance actions: do-nothing (DN) and repair (PR), combined with two inspection decisions: no-

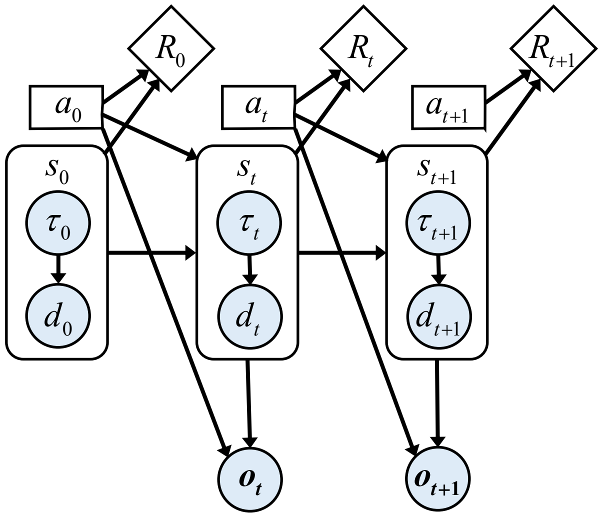


Fig. 3.2. Dynamic decision network of a POMDP built based on a deterioration rate DBN model [81].

inspection (NI) and visual inspection (VI), result in four action groups: DN-NI, DN-VI, PR-NI and PR-VI. Costs are then assigned to each of this combinations. Considering the do-nothing & no-inspection action (DN-NI), the reward $R(s, a_{DN-NI})$ corresponds simply to the risk of structural failure, assigning a failure cost C_f to the failure states d_F (Section 3.2.1). If the action also features inspections, then, an inspection cost C_i is added to each state, along with failure risk $R(s, a_{DN-NI}) + C_i$. Similarly, a repair cost C_r is included, for all states $s \in S$, if a repair action a_{PR} is undertaken. Note that costs can be considered as negative rewards.

For most practical applications, POMDPs state space is high-dimensional and might be computationally intractable if solved by exact value iteration or grid-based approaches. State-of-the-art point-based POMDP solvers are, however, capable of solving scaling to spaces of realistic dimensions, as demonstrated in [24]. Point-based solvers restrict the computation of Bellman backups to only a subset of reachable belief points, thus significantly improving computational efficiency. The value function $V_T(\mathbf{b})$ is parameterized by a set of hyperplanes (α-vectors), each of them associated with an action a; and the optimal policy π^* corresponds to the α-vector that maximizes the value function $V_T(\mathbf{b})$:

$$V_T(\mathbf{b}) = \max_{\alpha \in \Gamma} \sum_{s \in S} \alpha(s) b(s) \tag{3.9}$$

State-of-the-art point-based solvers are implemented for solving infinite horizon POMDPs, yet in many cases, the decision maker deals with finite horizon policies, e.g. 20 years lifetime. The state state space can be augmented, following the approach in [21] to transform previously constructed infinite horizon POMDPs into finite horizon POMDPs.

3.3 Fatigue deterioration environment

A monopile foundation, dominant in most installed offshore wind turbines, is an assembly of rolled plates welded transversely and forming a hollow steel pipe. A transverse butt weld is therefore deemed to be a representative structural detail in this case. The fatigue deterioration of the joint is modeled, following DNV-GL design standards [82], by a cumulative fatigue damage law. A limit state $g_{SN}(t)$ is then formulated based on cumulative damage Miner's rule, over time t:

$$g_{SN}(t) = \Delta - vt\left[\frac{q^{m_1}}{C_{1,SN}}\gamma_1\left\{1+\frac{m_1}{h}; \left(\frac{S_1}{q}\right)^h\right\} + \frac{q^{m_2}}{C_{2,SN}}\gamma_2\left\{1+\frac{m_2}{h}; \left(\frac{S_1}{q}\right)^h\right\}\right] \tag{3.10}$$

where $C_{1,SN}$, $C_{2,SN}$, m_1, and m_2 are material parameters corresponding to a 'D' category bi-linear SN curve; the expected stress range is parameterized by Weibull factors q and h; v represents the cycle rate; and Δ corresponds to the fatigue limit. Note that γ_1 and γ_2 stand for lower and upper incomplete gamma functions, respectively. Assuming the structure is designed to the limit, the loading scale factor q is back calculated considering a fatigue design factor of one for the I&M planning setting, and a fatigue design factor of two for the lifetime extension planning scenario. Table 3.1 lists all relevant parameters.

Since inspections cannot reveal the accumulated fatigue damage computed through Miner's rule, fracture mechanics models are normally utilized instead for in-service inspection and maintenance planning, as in that case, the crack size belief state can be updated based on collected crack observations. In this sense, a probabilistic fracture mechanics model is calibrated with the objective of achieving the structural reliability computed previously by the cumulative fatigue damage law (Eq. 3.10). Let us model the crack growth with a Paris' law model, originally introduced in [63]:

$$d_{t+1} = \left[d_t^{\frac{2-m}{2}} + \frac{2-m}{2}C_{FM}(Y\pi^{0.5}S_e)^m n)\right]^{\frac{2}{2-m}} \tag{3.11}$$

where the crack depth is modeled by d, with crack growth parameters C_{FM}, and m, n cycles per time step, geometric factor Y, and the same loading as for the damage cumulative law, described by the expected stress range $S_e = q\Gamma(1+1/h)$ through the parameters q and h. Table 3.1 lists all relevant parameters and the fatigue limit state is formulated as $g_{FM}(t) = d_c - d(t)$. Assuming a through-thickness failure, the critical crack size d_c corresponds to the plate thickness.

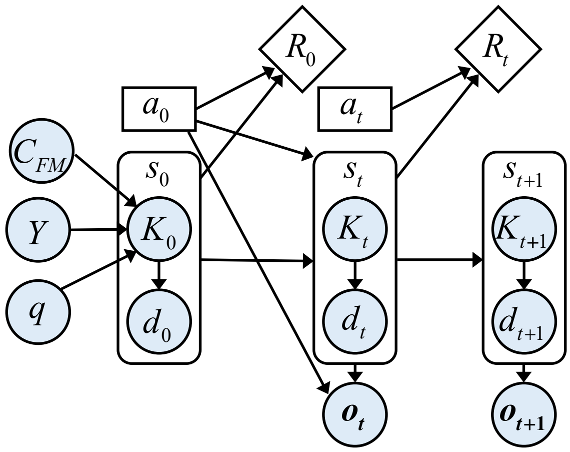


Fig. 3.3. Parametric DBN-POMDP dynamic decision network designed for the numerical experiments.

Table 3.1. Random variables and deterministic parameters for modeling the fatigue deterioration.

Parameter	Distribution	Mean	Std
Miner's cumulative damage model			
$C_{1,SN}$*	Normal	12.564	0.2
$C_{2,SN}$*	Normal	16.006	0.2
q** (MPa)	Trunc. Normal	10.209	2.55
q***	Trunc. Normal	8.834	2.21
Δ	Lognormal	1	0.3
h	Deterministic	0.8	-
v (cycles/s)	Deterministic	0.16	-
m_1	Deterministic	3	-
m_2	Deterministic	5	-
Fracture mechanics model			
lnC_{FM}**	Normal	-26.432	0.126
lnC_{FM}***	Normal	-26.501	0.131
d_0 (mm)	Exponential	0.11	-
Y	Lognormal	1	0.1
d_c (mm)	Deterministic	20	-
m (mm)	Deterministic	3	-

*Fully correlated.

**Inspection and maintenance planning application.

***Lifetime extension planning application.

Table 3.2. Description of the discretization scheme applied to DBN-POMDP deterioration rate and parametric models.

Variable	Interval boundaries
Deterioration rate model	
d	$[0, d_0 : (d_c - d_0)/(\lvert d\rvert - 2) : d_c, \infty]$
τ	$[0 : 1 : 20]$
$\tau*$	$[0 : 1 : 60]$
Parametric model	
d	$0, \exp\left\{\ln(10^{-2}) : \dfrac{\ln(d_c) - \ln(10^{-2})}{\lvert d\rvert - 2} : \ln(d_c)\right\}, \infty$
K	$0, \exp\left\{\ln(10^{-4}) : \dfrac{\ln(2) - \ln(10^{-4})}{\lvert K\rvert - 2} : \ln(2)\right\}, \infty$

*Lifetime extension planning setting.

We then translate the proposed probabilistic fracture mechanics model into both a deterioration rate and a parametric dynamic Bayesian network. The state transition models $p(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$ and $p(d_{t+1}, K_{t+1}|d_t, K_t)$ are constructed through sequential Monte Carlo simulations, with basis on Eq. 3.11, and discretized according to the scheme shown in Table 3.2. The observation quality $p(o|d)$ is modeled depending on the inspection type and will be explained on the respective case studies. An accurate enough discretization is achieved by including 60 crack states $|d|$, and resulting in a root mean square error of $2.4 \cdot 10^{-3}$ when comparing the reliability index with a Monte Carlo simulation featuring two eddy current inspections at years 8 and 16.

The developed DBN structures serve as the backbone for the evaluation of all the heuristic decision rules explored in the numerical investigations, and can be directly integrated into the underlying structure of the POMDP models, as described in Section 3.2.1. The POMDP state space for the deterioration rate model contains the joint distribution of d and τ, as illustrated in Fig. 3.2, summing up to a total of 1260 states. For the finite horizon I&M planning setting examined in Section 3.4.1, the state space is augmented to 13,860 states, due to the fact that time needs to be also included in the state vector.

The joint distribution of d and K forms the parametric POMDP model state space, as illustrated in Fig. 3.3, summing up in this case up to 4000 states. The time-invariant parameters C_{FM}, Y and q are combined into the chance node $K \equiv C_{FM}(Y\pi^{0.5}S_e)^m n$, in order to alleviate computational complexity. When translated into a finite horizon model applicable to the I&M planning setting, the parametric model augments to 156,000 states.

3.4 Numerical experiments

3.4.1 Inspection and maintenance planning of offshore wind structural components

In this first example, we explore a typical risk-based inspection planning setting with an assumed 20-year finite horizon, in which the inspection quality is modeled by probability of detection curves and with the possibility of planning perfect repair maintenance actions.

In this scenario, the transition model assigned to the action do-nothing a_{DN} is modeled according to the fatigue deterioration rate introduced in Section 3.3, and the perfect repair transition model a_{PR} transfers the current belief state $\mathbf{b_t}$ to its initial belief $\mathbf{b_0}$, as stated in Eq. 3.7.

While only one inspection type is available in most traditional inspection and maintenance planning applications, three inspection techniques are possible here, namely, eddy current, ultrasonic testing and visual inspection. Table 3.3 lists the parameters corresponding to each inspection technique, following the probability of detection (PoD) formulation proposed by [79]:

$$PoD(a) = 1 - \frac{1}{1 + (a/X_0)^b} \tag{3.12}$$

Based on the proposed transition and observation models, we construct a finite horizon deterioration rate POMDP by combining the following action and observation decisions: do-nothing & no-inspection (DN-NI), do-nothing & eddy current inspection (DN-EC), do-nothing & ultrasonic inspection (DN-UT), do-nothing & visual inspection (DN-VI), and perfect-repair & no-inspection (PR). The costs for this case are listed in Table 3.4. Note that perfect-repair is not paired with any inspection type since an observation is generally expected to be suboptimal after the component is fully repaired.

The finite horizon POMDP model is then computed via SARSOP and FRTDP point-based solvers. Furthermore, policies regulated by predefined heuristics are also evaluated. Heuristics include planning of (i) equidistant inspections (EQ-INS) or inspections upon exceedance of an annual failure probability threshold (THR-INS); and (ii) repairs automatically scheduled upon crack detection. Results from both POMDP and heuristic-based policies are reported in Table 3.5.

Table 3.3. Inspection quality.

Inspection technique	X_0	b
Eddy current (EC)	1.16	0.9
Ultrasonic testing (UT)	0.41	0.642
Visual inspection (VI)	83.03	1.079

Table 3.4. Definition of the cost model.

Failure	-1000 (money units)
Eddy current inspection	-1 (money units)
Ultrasonic inspection*	-1.5 (money units)
Visual inspection*	-0.5 (money units)
Perfect repair*	-100 (money units)
Production**	+5 (money units)
Replacement**	-100 (money units)
Decommissioning**	-20 (money units)
Discount factor	0.95 (-)

*Inspection and maintenance planning setting.
**Lifetime extension planning setting.

3.4.2 Lifetime extension planning setting

In this second example, we consider a lifetime extension planning setting, in which the decision maker opts for either replacing, decommissioning, or extending the lifetime of the structure. Suppose an offshore wind turbine on operation for 16 years without planned inspections or repairs up to that point. The initial belief state corresponds to the state of the structure, $\mathbf{b_{16}} = \mathrm{T}^{16}\,\mathbf{b_0}$, at year 16 .

Both deterioration rate and parametric infinite horizon POMDP models are laid out and solved through point-based solvers, by combining the following actions and observations decisions: do-nothing & no-inspection (DN-NI), do-nothing & eddy current inspection (DN-I), replacement & no-inspection (REP), and decommissioning & no-inspection (DEC). The do-nothing action, both including and excluding inspections, is modeled equally as for the I&M planning setting, and a replacement is assumed as a perfect repair. A decommissioning action, however, transfers the current belief state $\mathbf{b_t}$ to an absorbing state s_{dec}, in which no further rewards are collected. Table 3.4 lists all the utilities considered for this experiment. Note that a positive production reward is collected every time the structure is operative.

Table 3.5 reports the results for both POMDP and and heuristic-based policies. Heuristic rules consist, in this setting, in planning equidistant inspections (EQ-INS) or after reaching an annual failure probability threshold (THR-INS); and either a replacement (PR) or a decommissioning (DEC) action is ordered after a detection inspection is observed.

3.5 Results and discussion

POMDP-based policies outperform heuristic-based policies in all the explored settings, resulting in a total expected reward benefit ranging from 32% to 105%. Table 3.5 reports the results corresponding to both I&M planning and lifetime extension planning investigations. For each policy, either POMDP or heuristic-based, Table 3.5 lists expected total rewards $\mathbf{E}[R]$ along with the 95% confidence intervals (95% C.I.), and the relative difference in expected total rewards between each policy and SARSOP (%SARSOP). In terms of POMDP-based policies, the difference between SARSOP and FRTDP point-based solvers is less than 1% in all the experiments. While SARSOP solver quickly reduces the lower bound within seconds of computational time, FRTDP solver is able to reduce faster the upper bound, leading to convergence for the finite horizon I&M planning setting. Fig. 3.4 shows the evolution of expected total rewards for each solver over computational time.

Inspections planned before surpassing a predefined annual failure probability threshold (THR-INS) tend to reach better cost or rewards than planning inspection at equidistant intervals (EQ-INS). Moreover, decision rules featuring eddy current (EC) inspections result more optimal, under the proposed cost model, than those employing ultrasonic testing (UT) and visual inspections (VI). Heuristic decision rule evaluations also indicate that undertaking a decommissioning action (INS/DEC), after observing a crack indication, results more optimal than replacing the structure (INS/REP).

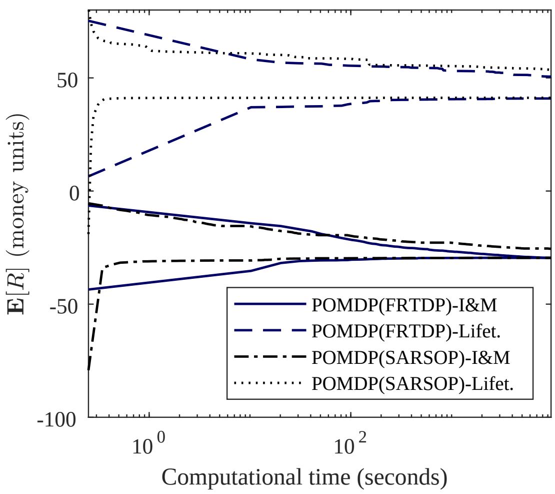


Fig. 3.4. Evolution of expected total rewards over computational time for each POMDP point-based solver.

Table 3.5. Comparison between POMDP and heuristic-based policies in I&M and lifetime extension settings.

Policy	$\mathbf{E}[R]$ (95% C.I.)	%SARSOP
I&M planning: 20 years finite horizon		
POMDP-SARSOP	-29.53	
POMDP-FRTDP	-29.53	<1%
Heur. EQ-INS (EC)	-39.62 (0.47)	-34.2%
Heur. THR-INS (EC)	-38.97 (0.35)	-32.0%
Heur. THR-INS (UT)	-48.63 (0.45)	-64.7%
Heur. THR-INS (VI)	-71.94 (0.18)	-143.6%
Lifetime extension planning: Infinite horizon		
POMDP-SARSOP*	41.20	
POMDP-FRTDP*	41.04	<1%
POMDP-SARSOP**	41.11	<1%
POMDP-FRTDP**	40.51	<2%
Heur. EQ-INS/DEC	16.16 (0.37)	-60.8%
Heur. EQ-INS/REP	-2.15 (0.17)	-105.2%
Heur. THR-INS/DEC	15.96 (0.34)	-61.3%
Heur. THR-INS/REP	1.16 (0.60)	-97.2%

*Deterioration rate POMDP (Fig. 3.2).

**Parametric POMDP (Fig. 3.3).

One can deduce that the optimality of heuristic-based policies will thus be influenced by the ability of exploring the appropriate space of decision rules. Selecting optimal heuristics is case dependent and can be achieved by experience or by probing a large set of decision rules. POMDP-based policies, on the other hand, offer a mapping from the current belief state (dynamically encoding the entire prior history of actions and observations) to the optimal action. The sequence of optimal actions might be non-trivial for certain scenarios. Consider, for instance, a POMDP policy realization in the lifetime extension application, dictating a decommissioning action after three successive crack detection indications, as illustrated in Fig. 3.5. The tested indication-based heuristic policy, also in the lifetime extension setting, assigns a decommissioning action after observing one detection indication, as shown in Fig. 3.6. The learned POMDP policy acknowledges that, under the explored application, the observation model might not be very accurate and more information shall be collected before ordering a decommissioning action. If a detection is followed by a no-detection indication, a do-nothing action is instead preferred.

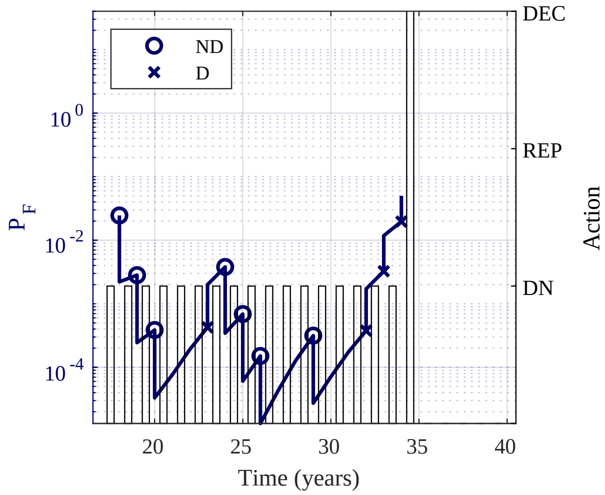


Fig. 3.5. Realization of a POMDP policy in the lifetime extension planning application.

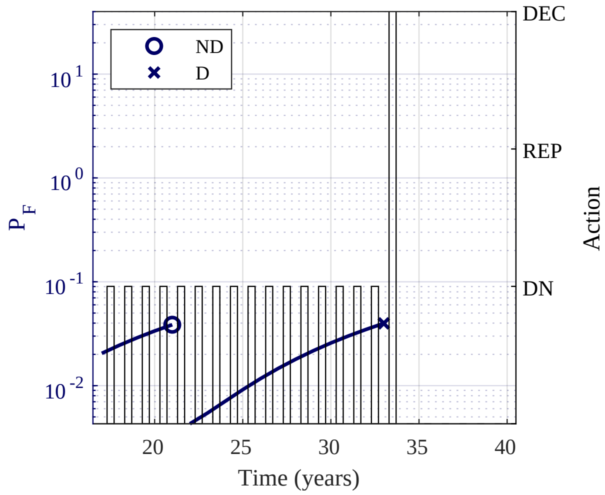


Fig. 3.6. Realization of a heuristic policy in the lifetime extension planning application.

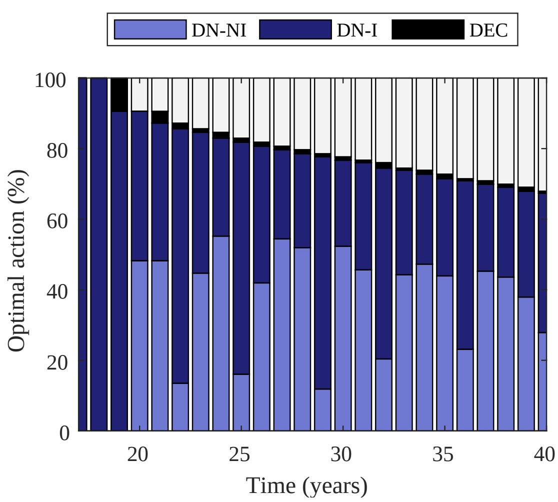


Fig. 3.7. Histogram of optimal actions assigned after 10^5 policy evaluations in the lifetime extension planning setting.

As explained in Section 3.2.2, point-based POMDP policies are parameterized by a set of hyper-planes (α-vectors), each of them associated to one action. A frequency histogram induced by policy evaluations offers to the decision maker a summary of the actions taken over the investigated horizon. Fig. 3.7 represents, for the lifetime extension application, the histogram of actions defined by the POMDP policy, collected through 10^5 policy realizations.

3.6 Conclusions

This paper studies the efficiency of integrating Dynamic Bayesian Networks (DBNs) and Partially Observable Markov Decision Processes (POMDPs) in a joint algorithmic context for optimal Inspection and Maintenance (I&M) planning. Time-invariant parameters and finite horizon settings can be implemented within this framework by simply augmenting the POMDP state space, generating high-dimensional sparse matrices, which can be efficiently solved by state-of-the-art point-based POMDP solvers.

The application of the methodology to the case of an offshore wind structural detail subject to non-stationary fatigue deterioration process verifies the computational efficiency of the proposed approach. The results show that POMDP-based policies outperform traditional heuristic-based policies in all the studied settings. Further efforts are envisaged towards encoding DBNs into factored-POMDP structures or approximating optimal POMDP policies through deep reinforcement learning approaches with or without stochastic constraints [28, 29] for settings featuring high-dimensional state, action, and observation spaces.

Authorship contribution statement

Morato, P. G.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Papakonstantinou, K. G.:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing - Review & Editing, Supervision. **Andriotis, C. P.:** Methodology, Software, Validation, Formal analysis, Writing - Review & Editing. **Rigo P.:** Supervision, Project administration.

Acknowledgements

This research is funded by the National Fund for Scientific Research in Belgium F.R.I.A. - F.N.R.S.

Chapter 4

POMDP-based Maintenance Optimization of Offshore Wind Substructures including Monitoring

---

Paper Morato, P. G., Nielsen, J. S., Mai, A. Q. and Rigo P. (2019). POMDP-based Maintenance Optimization of Offshore Wind Substructures including Monitoring. In *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP 2019)*, 270-277.

Abstract Sequential decision making under uncertainty is a complex task limited normally by computational requirements. A novel methodology is proposed in this paper to identify the optimal maintenance strategy of a structural component by using a point-based Partially Observable Markov Decision Process (POMDP). The framework integrates a dynamic Bayesian network to track the deterioration over time with a POMDP model for the generation of a dynamic policy. The methodology is applied to an example quantifying whether a monitoring scheme is cost effective. A decision problem comprised of 200 damage states is solved accurately within 60 seconds of computational time.

4.1 Introduction

Offshore wind energy is a sustainable solution for energy generation. Further from shore, higher and steadier wind speeds can be harnessed and the visual impact is reduced as compared with onshore wind. However, offshore wind substructures are subjected to a harsh deterioration due to the combined action of fatigue and corrosion.

Besides, maintenance operations are complex and expensive. It is therefore of utmost importance to provide decision support to a decision maker (operator) who is taking the decisions under uncertainty. The maintenance strategy can be optimized by following a risk-based approach where an optimal balance is achieved between the maintenance efforts and the large consequences associated with a structural failure.

In addition to inspections, Structural Health Monitoring (SHM) can be employed to gather more information about the state of the structures. SHM techniques have improved considerable as more accurate and reliable sensors are available. Nevertheless, there is a cost associated with a SHM scheme due to the installation and operation of the system and a risk of increased costs, if too many inspections are initiated on the basis of false alarms. Then, the decision maker must face the decision whether to utilize and install a SHM scheme or not. This decision can be optimally chosen by quantifying the value of the information.

The concept of the Value of Information (VoI) was introduced by [1], providing a theoretical framework to quantify the value of information within the Bayesian decision analysis. Based on this framework, a great number of research efforts have been devoted recently to quantify the value of monitoring for civil infrastructures, such as bridges or hydraulic structures. The reader is directed to [83] for a more exhaustive illustration on the VoI framework for sequential decision problems.

The main limitation of these methodologies strives on the assumptions and simplifications imposed due to the computational requirements involved in the solution of complex decision problems. For instance, the applications consider small state spaces or stationary decision rules such as "preset interference threshold" are imposed.

This work presents a methodology to quantify the value of monitoring by employing a point-based "Partially Observable Markov Decision Process" (POMDP). Since a POMDP point-based solver samples only a subset of the belief space, this methodology can be employed to generate dynamic maintenance policies, even when complex sequential decision problems are involved [21, 84].

4.2 POMDP-based methodology

A novel methodology is presented hereby to quantify the value of monitoring. The expected maintenance costs are estimated separately for the case when only inspections are included, and for the case when a monitoring system is also included. Thereafter, the Value of Information (VoI) can be computed as the difference.

Concerning the inspection planning, the influence diagram (Fig.4.1) displays how this sequential decision problem is approached. The damage evolving over time is represented by the chance node D_t and it is possible to choose an inspection method (including no-inspection) by means of the decision node I_t.

The chance node Z_t indicates the quality of the inspection method. Additionally, the node E_t tracks the probability of being in the last damage state, or in other words, the failure probability. The utility nodes C_{F_t} and C_{I_t} assign a cost of failure and a cost of inspection, respectively.

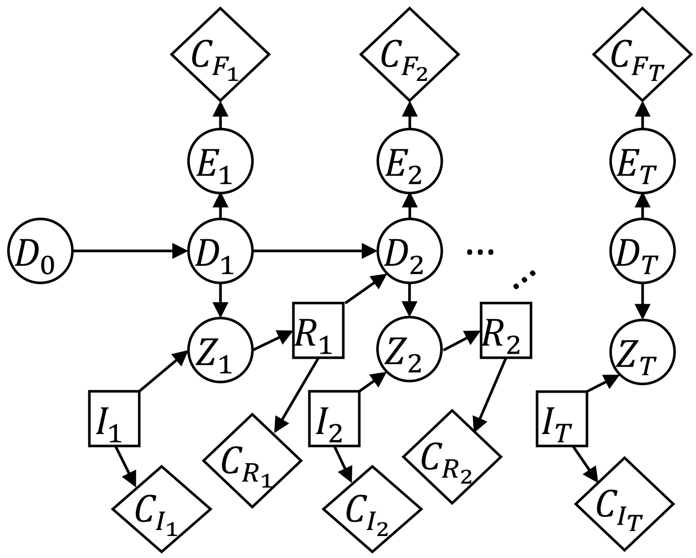


Fig. 4.1. Influence diagram corresponding to the inspection and maintenance planning decision-making problem.

Ultimately, the chance node R_t represents the decision of whether to perform a repair or not. If it is decided to make a repair, then the damage state will be transferred to a healthier state and it will have an associated cost of repair C_{R_t}.

4.2.1 Building the model

A POMDP model is built in order to solve the maintenance decision problem. The outcome of the POMDP model is a policy which informs the optimal decision depending on the current belief state (probability distribution for the node D). This decision is comprised within the context of this framework by a combination of an action and an observation. Examples of actions are "do-nothing" or "repair" and it is possible to gather observations by "inspection", "monitoring", including also the case of "no-inspection". A decision could be for instance "do-nothing / inspection" or "do-nothing / no-inspection".

The input of a POMDP simulation includes therefore: (1) the transition probability $[T]$ from one damage state to another depending on the action chosen, (2) the observation likelihood [O] depending on the inspection type selected, (3) the rewards associated with the taken "action/observation" decision (including a discount factor), and (4) the initial damage state D_0.

Transition probabilities

The transition matrices are defined according to the associated action. The transition probabilities for the action "do-nothing" can be obtained as the conditional probabilities corresponding to the "Dynamic Bayesian Network" (DBN) shown in Fig. 4.1, where only the damage nodes are kept. Thus, T_{DN} is equivalent to the conditional probabilities of damage at time step "t" given the damage at the previous time step "$t-1$" (Eq. 4.1). If no time-invariant uncertainties are involved, the conditional probabilities can be obtained by Monte Carlo simulations.

$$T_{DN} = P(D_t|D_{t-1}) \tag{4.1}$$

For the case of the repair action, the transition matrix simply transfers the component to a healthier state, depending on the repair quality.

Observation probabilities

If an inspection or monitoring is performed, then an observation is gathered. The observation matrix conveys the quality of this information (likelihood). The observation matrix is constructed depending on the selected observation: 1)"No-inspection": No information is gained, thus, the belief state must remain unaltered; 2)"Inspection": The observation matrix is directly computed from a "Probability of Detection" curve; 3)"Monitoring": The observation matrix is obtained in a similar manner as for the inspection case.

Decisions and associated rewards

The rewards depend on the decision taken and this will be greatly influenced by the nature of the problem. For instance, the following approach can be taken: 1)"Do-nothing / No-inspection" ($DN-\bar{I}$): Only the failure cost is considered; 2)"Do-nothing / Inspection" ($DN-I$): The inspection cost is included along with the failure cost; 3)"Repair / No-inspection" ($R-\bar{I}$): Here the repair cost is considered; 4)"Do-nothing / Monitoring" ($DN-M$): As the value of information will be calculated a posteriori, monitoring costs are not included in the POMDP model.

If desired, more decisions can be added into the model, yet with an additional computational cost. Finally, a discount factor must be included $\gamma \in (0-1)$ to quantify the present value of money over time. This discount factor becomes necessary if an infinite horizon POMDP is employed.

4.2.2 POMDP Simulation

Once the POMDP input has been prepared by including: transitions, observations, rewards and the initial state; then, a point-based solver is employed to generate a "POMDP" maintenance policy. This framework allows the computation of large state POMDP spaces due to the fact that point-based solvers are able to efficiently compute large belief states within a reasonable computational time [21, 84]. In the application presented in this paper

(Section 3), the solver "SARSOP" [31] is selected; nevertheless, other POMDP solvers such as "PERSEUS" or "HSVI" can be used instead.

The approach followed by this methodology leads to the creation of an infinite horizon POMDP, where the obtained policy is applicable for any time step. If a finite horizon POMDP is preferred; then, time must be encoded within the transition matrices at the cost of significantly increasing the belief space and computational time [21, 84].

4.2.3 Post-processing

After the simulation is conducted, a policy is obtained as a result of the POMDP model. Additionally, the POMDP solver provides for each computational time step: (1) expected costs (delimited by upper and lower bounds), (2) number of beliefs and α-vectors and (3) number of backups. The expected costs provides the main outcome for the decision problem, whereas the other parameters can be checked to understand more details about the generated policy and the complexity of the problem.

Furthermore, the obtained policy is comprised of a set of α-vectors (Γ), each of them associated to a decision. The optimal decision is the one which corresponds to the α-vector that maximizes the value function $V(b)$ as shown in (Eq. 4.2). Hence, the decision is chosen only based on the current belief state (b).

$$V(b) = \max_{\alpha \in \Gamma}(\alpha \cdot b) \tag{4.2}$$

Additionally, the influence diagram displayed in Fig. 4.1 can be used in combination with the generated policy to choose the optimal decision for a particular scenario. The inspection decision node I_t is then instantiated with the optimal decision by applying Eq. 4.2.

4.2.4 Quantifying the value of monitoring

If a monitoring system is installed, the uncertainties are reduced because an observation is continuously gathered (every time step). This will have an effect on the maintenance strategy as normally less inspections might be necessary. Therefore, the benefit of installing a maintenance scheme is quantified as the difference between: (1) the achieved reduction of expected costs (as additional information is provided by monitoring), and (2) the cost of the monitoring system. In other words, the objective is the quantification of the Value of Information (VoI) or in this case the value of monitoring.

The VoI is calculated as the difference between the expected costs if monitoring is not conducted ($\mathbb{E}(C_0)$) and the expected cost if monitoring is conducted ($\mathbb{E}(C_1)$).

However, the cost of the monitoring system C_M is neglected for this calculation:

$$VoI = \mathbb{E}(C_0) - \mathbb{E}(C_1) \tag{4.3}$$

Additionally, it is useful to introduce the concept of Net Value of Information (NVoI) which also includes the cost of monitoring:

$$NVoI = VoI - C_M = \mathbb{E}(C_0) - \mathbb{E}(C_1) - C_M \tag{4.4}$$

The NVoI is very helpful for the decision maker because it is used to decide whether monitoring should be performed or not: if the NVoI is positive, then monitoring provides an added value; if the NVoI is negative, then the monitoring system is more expensive than the benefit gained by its installation.

4.3 Application

The value of monitoring is now quantified for a maintenance decision problem partially based on the Example presented by [71]. However, here the decision maker must decide whether to install a monitoring system or not.

4.3.1 Model

The fatigue deterioration of a structural component is here modelled by a probabilistic fracture mechanic model based on the Paris' law (Eq. 4.5). Both the initial crack size and stress range are considered as random variables. The damage size or crack size is computed for each time step with the expression developed by [63]:

$$a_t = \left[\left(1 - \frac{m}{2}\right) C \Delta S^m \pi^{m/2} \Delta n + a_{t-1}^{1-m/2} \right]^{(1-m/2)^{-1}} \tag{4.5}$$

Where a_t is the damage (crack size) at the time step "t", a_{t-1} is the damage at the previous time step "$t-1$", C and m are material parameters which condition the crack propagation, ΔS is the stress range and Δn is the number of cycles per time step.

Thus, given an initial crack size ($a_{t=0}$), the crack size distribution can be computed for the following time steps. In this example, the time step is considered to be one month. The values of the parameters are listed in Table 4.1. The component fails once the crack has reached the critical crack size a_c, which is considered here to be 9 mm.

Table 4.1. Parameters for the fracture mechanics model

Parameter	Distribution	Mean	StDev
a_0	Exponential	0.2	-
a_c	Deterministic	9	-
$ln(C)$	Deterministic	-33.5	-
m	Deterministic	3.5	-
ΔS	Normal	60	10
Δn	Deterministic	10^6	-

The limit state is then formulated in Eq. (4.6) where the failure probability is computed as the probability of the limit state being negative.

$$g_{FM}(t) = a_c - a(t) \tag{4.6}$$

In principle, the failure probability can be estimated by using a crude Monte Carlo simulation. Nevertheless, a Dynamic Bayesian Network (DBN) is here proposed as the basis to both define the transition probabilities for the actions where no maintenance actions are involved and for evaluating the obtained policy.

Building the DBN model

A discretization scheme is used to convert the deterioration model from a continuous space to a discrete space so as to facilitate the inference of the DBN:

$$a \in (0, exp[ln(10^{-5}) : \frac{ln(9) - ln(10^{-5})}{states - 2} : ln(9)], \infty) \tag{4.7}$$

The DBN is derived from the influence diagram presented in Fig. 4.1 to track the damage. The crack size is represented by the nodes a_t. If an inspection is performed, a node Z_t is incorporated into the network as shown in Fig. 4.2. Additionally, the node E_t collects the failure probability which it is equivalent to the probability of being in the last damage state.

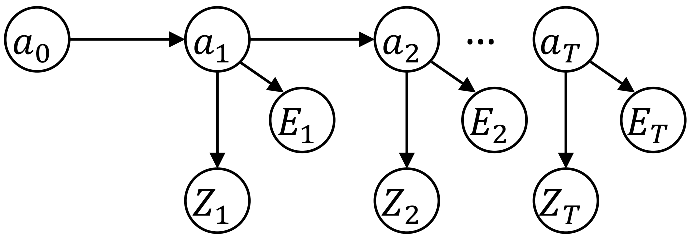


Fig. 4.2. Graphical representation of the DBN model.

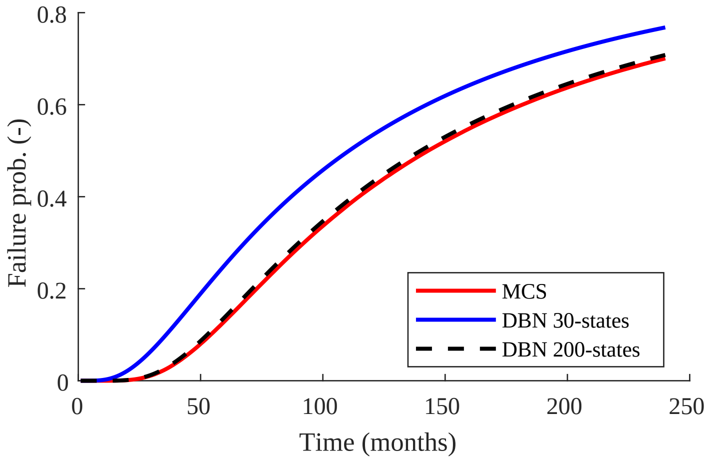


Fig. 4.3. DBN discretization accuracy.

Due to the fact that point-based POMDP solvers are able to solve large belief spaces in a reasonable computational time, the number of states for the damage size is not limited here. It is chosen in such a way that the computed failure probability by the DBN model is similar as the result obtained by a crude Monte Carlo Simulation (MCS).

As it can be seen in Fig. 4.3, a discretization with 200 states provides enough accuracy for the DBN as the failure probability is in good agreement with the result from the crude MCS.

POMDP model including inspections

A POMDP model is built by defining the transition probabilities, observation or emission probabilities, rewards and the initial state. For this case, three possible decisions are included:

- Do-nothing / No-inspection ($DN - \bar{I}$): comprised of the transition matrix "do-nothing" T_{DN} and the observation matrix "no-inspection" $O_{\bar{I}}$.
- Do-nothing / Inspection ($DN - I$): composed of the transition matrix "do-nothing" T_{DN} and the observation matrix "Inspection" O_I.
- Repair / No-inspection ($R - \bar{I}$): comprised of the transition matrix "repair" T_R and the observation matrix "No-inspection" $O_{\bar{I}}$.

The "do-nothing" transition matrix T_{DN} is easily defined by utilizing the conditional probabilities used for the development of the DBN, as shown in Eq. 4.8. For the transition corresponding to the repair action T_{RP}, the damage is transferred to a healthier state (initial damage size), independently of the current damage state.

$$T_{DN} = \begin{bmatrix} p(a^1_{t+1}|a^1_t) & 0 & \dots & 0 \\ p(a^2_{t+1}|a^1_t) & p(a^2_{t+1}|a^2_t) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p(a^n_{t+1}|a^1_t) & p(a^n_{t+1}|a^2_t) & \dots & p(a^n_{t+1}|a^n_t) \end{bmatrix} \tag{4.8}$$

Inspection outcomes are normally defined by "Probability of Detection" curves (PoDs) within the context of traditional risk-based inspection methods. PoDs determine the measurement uncertainty by assigning the probability of detection given the damage size. This is can be translated to the DBN because PoDs are equivalent to the conditional probabilities corresponding to the inspection node Z_t given the damage a_t. Furthermore, in this example, an inspection can lead to six different outcomes, each of them depending on the damage size. Table 4.2 states the probability of obtaining each outcome as lognormal distributions, defined similarly as [71].

The measurement uncertainties are therefore employed to define the observation matrix for the case where an inspection is performed O_I:

$$O_I = \begin{bmatrix} p(ins_1|a^1) & p(ins_2|a^1) & \dots & p(ins_m|a^1) \\ p(ins_1|a^2) & p(ins_2|a^2) & \dots & p(ins_m|a^2) \\ \vdots & \vdots & \vdots & \vdots \\ p(ins_1|a^n) & p(ins_2|a^n) & \dots & p(ins_m|a^n) \end{bmatrix} \tag{4.9}$$

If the component is not observed (inspection is not planned $O_{\bar{I}}$); then, the observation matrix is defined as shown in Eq. 4.10. By using this observation matrix, the belief state prevails invariable. Since the belief state remains unaltered, it is equivalent to the case where no information is obtained.

Table 4.2. Inspection: measurement uncertainty

State	Description	Mean	COV
1	no detection	-	-
2	mild damage	2.0	1.0
3	some damage	4.0	0.8
4	significant damage	6.0	0.6
5	severe damage	8.0	0.4
6	failure	9.0	0.0

$$O_{\bar{I}} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 1 & 0 & \dots & 0 \end{bmatrix} \tag{4.10}$$

Once, the transition and observation probabilities are stated, the next step is to define the rewards. In this example, an inspection is associated with a cost of 1 money unit, a repair costs 50 money units and if the failed state is reached, a penalization of 500 money units must be paid as listed in Table 4.3. A exact definition of the costs is not crucial for a risk-based analysis whereas the relative difference between the cost associated to each decision (action/observation) is very important because it conveys the preference of the decision maker. Finally, the discount factor is defined as $\gamma = 0.95$.

Table 4.3. Rewards

	State 1	...	Failed state
Do-nothing	0	0	500
Repair	50	50	50
Inspection	1	1	1

POMDP model including inspections and monitoring

A POMDP is now built for the case when both inspections and monitoring are incorporated. The transition probabilities, rewards and initial state are defined in the same manner as for the case when only inspections were included. However, the decision "Do-nothing / No-inspection" ($DN - \bar{I}$) is here replaced by the decision "Do-nothing / Monitoring" ($DN - M$) as the structure is monitored continuously. Thus, three decisions are now possible: (1) "Do-nothing / Monitoring" ($DN-M$), (2) "Do-nothing / Inspection" ($DN-I$) and (3) "Repair / No-inspection" ($R - \bar{I}$).

Table 4.4. Monitoring: measurement uncertainty

State	Description	Mean	COV
1	no alarm	-	-
2	low alarm	2.0	1.0
3	high alarm	5.0	1.0
4	failure	9.0	0.0

The observation probabilities for the case when monitoring is performed (O_M) correspond to the conditional probabilities of obtaining each monitoring outcome given the damage size ($mon|a$). Hence, the observation matrix is defined according to Eq. 4.11. The probability of obtaining each outcome is modelled by a lognormal distribution and it is presented in Table 4.4.

$$O_M = \begin{bmatrix} p(mon_1|a^1) & p(mon_2|a^1) & \dots & p(mon_m|a^1) \\ p(mon_1|a^2) & p(mon_2|a^2) & \dots & p(mon_m|a^2) \\ \vdots & \vdots & \vdots & \vdots \\ p(mon_1|a^n) & p(mon_2|a^2) & \dots & p(mon_m|a^n) \end{bmatrix} \tag{4.11}$$

4.3.2 Results

Both POMDP models (only inspections / monitoring and inspections) are simulated with the point-based solver "SARSOP". Firstly, the expected costs resulting from each POMDP model are presented; secondly, the value of information is computed, and finally, the application of the POMDP policy for a particular case is conducted.

Expected costs and policies

The total expected costs are presented in Fig. 4.4. As expected, the total costs are higher for the case where only inspections are included. It is interesting to notice that the POMDP solution provides an upper and lower boundary for the total expected costs.

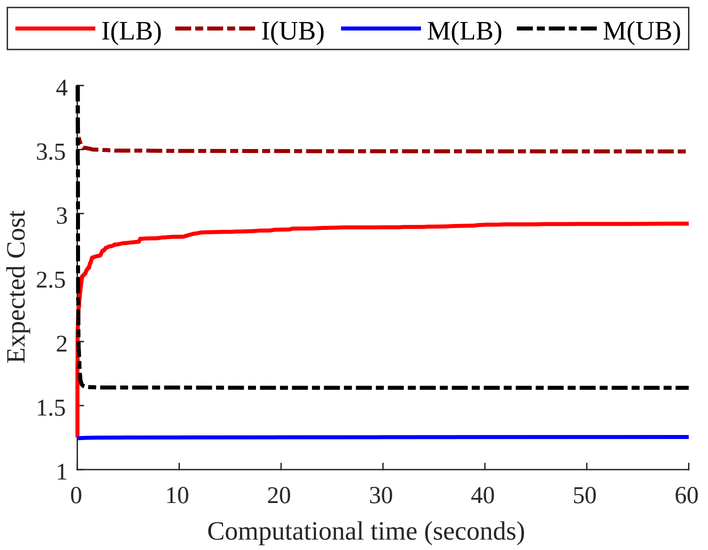


Fig. 4.4. POMDP results: Expected costs.

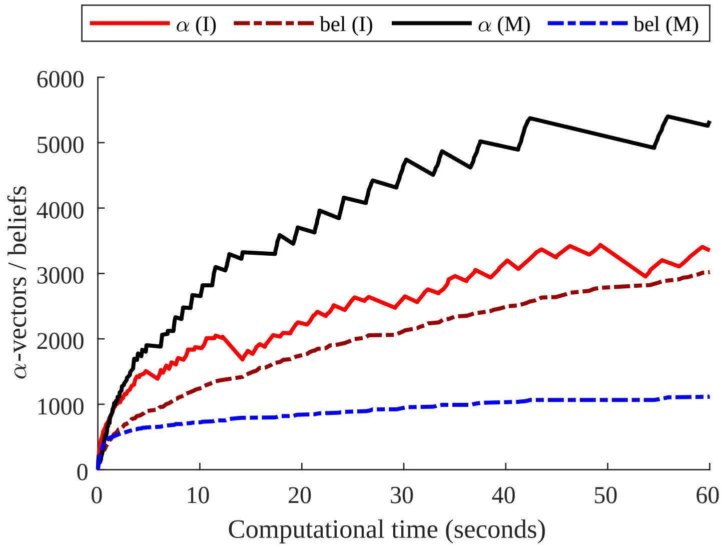


Fig. 4.5. POMDP results: α vectors and beliefs.

Due to the nature of the algorithm, new belief states are sampled and evaluated over time as shown if Fig. 4.5. The precision can be improved if the simulation is run for a longer time (new beliefs and α vectors will be generated); however, the accuracy is considered acceptable for this example within 60 seconds of CPU time, with an Intel Core I9 7900X @3.0 GHz and RAM 64GB.

Quantification of the value of information

It is possible at this point to provide decision support under uncertainty by quantifying the value of monitoring. Fig. 4.6 represents the upper and lower boundaries of the Net Value of Information (NVoI) for a given monitoring cost. The monitoring system is considered as economically feasible if the NVoI is positive and it is infeasible if the NVoI is negative.

The result suggests that it is cost-effective to install the monitoring system if its cost is lower than approximately 1.3 money units. Since the expected costs are delimited by upper and lower bounds, the NVoI is also delimited by bounds. A better accuracy can be achieved by increasing the simulation time, however, the precision is considered acceptable for this example.

Application of the POMDP policy for a particular case

The generated policy by the POMDP where only inspections are included is utilized now to select the optimal decisions for three particular cases. The result is depicted in Fig. 4.7. For the case 1, it is assumed that the inspection outcome is always "No-detected"; for the case 2, the inspection outcome is "No-detected" up to the year 14, after, the outcome is

"Mild damage"; and for the case 3, the inspection outcome is "Severe damage" after year 14.

Although the decision "do-nothing" dominates for the case 1, "inspections" are also scheduled. Regarding the case 2, a repair is performed between the years 14 and 15. This repair is undertaken after successive inspections where the assumed outcome is "mild damage". However, the repair is selected only after one inspection for the case 3.

This result is reasonable because a repair is planned if a severe damage is found, whereas several inspections are necessary before the repair, if the outcome is mild damage. It is demonstrated by this example how the generated policy can be used in a dynamic fashion, providing support for complex sequential decision making, where different inspection outcomes can be expected.

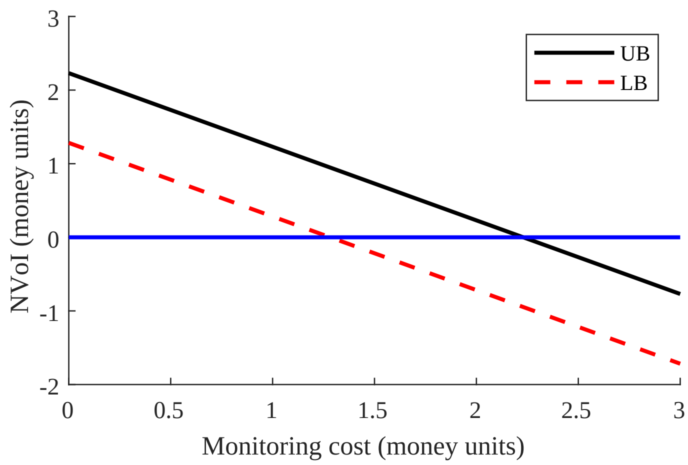


Fig. 4.6. Quantification of the net value of information.

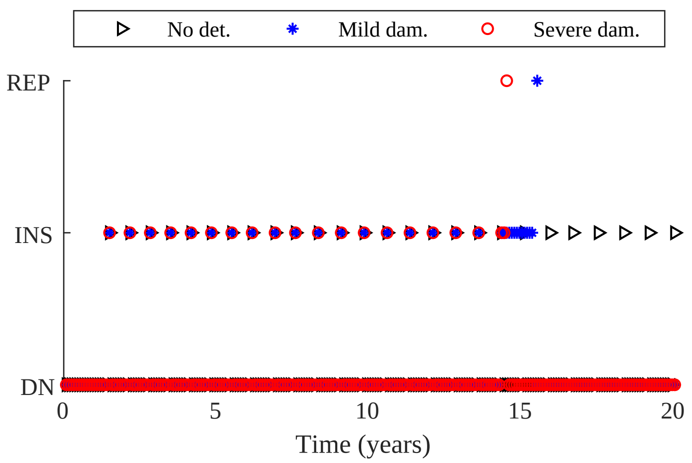


Fig. 4.7. POMDP policy realization.

4.4 Conclusions

Structural Health Monitoring (SHM) provides information about the state of the structural components leading to a reduction of uncertainty. As the uncertainties are reduced, better decisions can be taken. Within the context of maintenance planning, if information is gathered by monitoring, less inspections might be necessary, becoming especially important for the case of offshore wind structures, where inspections are complex and expensive.

However, there is a cost associated to the installation of a monitoring scheme and it must be decided whether it is cost-effective or not. A methodology is proposed in this paper to quantify the benefit or Value of Information (VoI) achieved by monitoring.

The methodology is applied to identify the optimal maintenance policy for a structural element subjected to fatigue deterioration. The policy is generated within a reasonable computational time for this complex case, where the damage size is discretized into 200 states and different outcomes of inspections and monitoring are possible.

In the future, efforts should be made to enable the use of time-invariant uncertainties within the deterioration modelling. Besides, the maintenance will be more optimal if it is performed at the system level incorporating the correlations or dependencies amongst the involved random variables. The development of hierarchical models is encouraged.

Authorship contribution statement

Morato, P. G.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.
Nielsen, J. S.: Software, Validation, Supervision, Writing - Review & Editing.
Mai, A. Q.: Software. **Rigo P.:** Supervision, Project administration.

4.5 Acknowledgements

This research is funded by the National Fund for Scientific Research in Belgium F.R.I.A. - F.N.R.S.

Chapter 5

Optimal Management of Deteriorating Structures considering System Effects: a Deep Reinforcement Learning Approach

Paper Morato, P. G., Andriotis, C. P., Papakonstantinou, K. G. and Rigo P. (2021). Optimal Management of Deteriorating Structures considering System Effects: a Deep Reinforcement Learning Approach. *Reliability Engineering and System Safety*, under internal review.

Abstract In the context of modern environmental and societal concerns, there exists an emerging 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) interventions. Most available methods simplify the I&M decision problem to the component level, often assuming independence among the components, due to the computational complexity associated with a global optimization. In this paper, we propose an efficient algorithmic platform for decision-making under uncertainty of engineering systems, providing optimal management strategies 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 also handle environments under equal or unequal deterioration dependence among components, through Gaussian hierarchical structures, decoupling the originally joint system space to component networks conditional on common source random variables. In terms of policy optimization, we adopt a deep decentralized multi-agent actor-critic (DDMAC) scheme, 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 reveal not only that DDMAC policies offer substantial benefits when compared to state-of-the-art heuristics, but the inherent consideration of the 'system-effects' by DDMAC strategies is also reflected in the explored settings.

5.1 Introduction

Managing engineering infrastructures, by controlling the risk of adverse events and optimally allocating inspection and repair actions, is crucial for safeguarding societal progress, improving the quality of life at the community level and maximizing economical returns from an individual perspective [52, 85]. The research efforts devoted to the development of risk-based inspection and maintenance planning methods have increased considerable during the last decade [86, 87, 88]. Modern societal concerns on sustainability, and on the preservation of the environment, along with the possibility of collecting operational data, demands for more rational management methodologies able to optimally dictate both repair and inspection actions throughout the infrastructure lifetime [50].

Most available inspection and maintenance (I&M) planning methods assume independence among the constitutive components, due to the computational complexity associated to solving such a decision-making problem under uncertainty [62]. Besides, failure-rate deterioration models built from collected statistics are not suitable for deteriorating structures due to the scarce data available on structural failures and because each structure is different on its own. At the component level, existing risk-based I&M methods can be classified according to their capabilities in modeling physically-based deterioration processes, e.g. simulated fatigue or corrosion deterioration [11, 66, 89], and depending on the policy optimization, 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 concentrate on the optimization of predefined static decision rules, planning inspections at equidistant intervals or when a prescribed failure probability threshold is surpassed, and ordering maintenance interventions if a certain indication is observed, e.g. crack detection [9, 62, 90, 91]. While these approaches provide reasonable policies in some specific scenarios, the optimality of the policies depend greatly on the designer's experience when defining the heuristic combinations for the policy search, disregarding, in any case, unexplored policies within the vast available policy space, which could in turn result more optimal that the originally predefined heuristics [81]. In other existing methods, while inspection planning decision rules are defined a priori, the maintenance policy is adaptive, often represented by an influence diagram [12, 20]. In this case, a large number of simulations is required to achieve a low variability in the expected life-cycle costs obtained by the resulting policies, and the inspection planning is in any case formulated based on a static optimization approach, constrained by heuristic combinations.

Early works on the application of Markov decision processes for managing deteriorating engineering processes include [21, 22, 23]. Justified and relying on the principled mathematical properties offered by Markov decision processes, either under fully or partial observability, additional methods have been proposed, e.g. in [25, 26, 48]. Recently, a POMDP-based approach proposed in [81] demonstrated that Partially Observable Markov Decision Processes (POMDPs) based policies outperform heuristic-based policies, substantiated with physically-based high-dimensional numerical examples featuring fatigue deterioration processes. POMDP policies are defined as a function of the belief state, a sufficient statistic describing the probability distribution over states, storing the dynamically updated history of actions and observations.

In all the aforementioned I&M methods, the decision-making problem is formulated at the component level. Disregarding the interrelations among the constitutive elements, even if allowing a simplification of the decision-making problem, may result in sub-optimal (and even) non-conservative policies for some cases. The need for I&M methods capable of identifying policies at the 'system level' is a recurrent claim within the risk research community. Early works approached at the system level include [13, 14, 15, 16]. In [17], the fatigue details were classified according to the fatigue design factor, establishing a simplified approach for identifying the policies at a system level. More recent approaches [6, 33] proposed a static optimization I&M planning relying on dynamic Bayesian networks to efficiently model deterioration, cost and reliability dependence among the structural elements. In this method, the policy is obtained by optimizing static heuristic decision rules: equidistant inspections, number of inspected components, selection of inspected components based on a proxy (failure probability) and repairs are planned after a detection indication is observed. As for other static policy optimization methods, the policies are constrained to the set of predefined heuristic rules, out of the immense possible available policies, even more numerous in structural system settings.

Addressing the complexities of managing large engineering systems, a deep reinforcement learning method has been introduced in [28], motivated by the success of deep reinforcement learning algorithms in complex game environments, e.g. in [34, 35, 36]. In particular, a multi-agent actor-critic is developed in [28], 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 I&M decision-making problem under constraints, e.g. imposed risk thresholds or budget limitations.

In this work, we propose an efficient algorithmic platform for decision-making under uncertainty of engineering systems, generating management strategies at the system level. The decision-making problem is formulated as a factored POMDP, whose dynamics are encoded as Bayesian network conditional structures, resulting in lower dimensional transition models than those relying on flat POMDPs. Within the proposed methodology, environments under deterioration dependence among components are formulated through Gaussian hierarchical structures, decoupling the originally joint system space to component networks conditional on common source random variables. This decomposition derives in a linear growth of computational complexity with increasing number of components that otherwise would increase exponentially in the joint system space. Furthermore, the Gaussian hierarchical model is generalized to enable the treatment of unequal deterioration correlation scenarios and dependence alterations after a maintenance action is taken. As the transitional model should consider, in this case, the common source random variables, we list the algorithmic steps for updating the belief state under deterioration correlation.

In terms of decision-making optimization, we adopt a deep decentralized multi-agent actor-critic (DDMAC) scheme, in which the policies are approximated by actor neural networks, at a component level, guided via value function estimates approximated by a critic network, at the system level. As DDMAC adjusts the weights of the actor networks according to noisy rewards collected at the system level, DDMAC policies intrinsically consider system-effects. Through numerical experiments, we demonstrate the efficacy of the proposed method for I&M planning of structural systems exposed to fatigue deterioration. 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, featuring a steel frame structural system, the focus is on examining the inherent allocation of maintenance interventions by DDMAC policies according to the element importance to the global structural reliability. In all the explored experiments, DDMAC policies are compared against state-of-the-art heuristic policies.

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 deterioration dependence. The integration of the simulator, defined as a factored POMDP, with DDMAC is presented in Section 4. The numerical experiments are introduced and discussed in Section 5, concluding with some final remarks in Section 6.

5.2 I&M decision problem formulated as a factored POMDP

5.2.1 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 $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{Z}, \mathcal{R}, \gamma \rangle$. In a POMDP, a decision maker (henceforth denoted as agent) interacts with a stochastic environment, described by the state space $s \in \mathcal{S}$, by taking actions $a \in \mathcal{A}$ over a finite or infinite horizon t_N. The dynamics correspond to those in a Markov Decision Process (MDP): at each time step t, an agent takes an action $a_t \in \mathcal{A}$, and the environment evolves from state $s_t \in \mathcal{S}$ to state $s_{t+1} \in \mathcal{S}$, according to the transition model $\mathcal{T} \coloneqq p(s_{t+1}|s_t, a_t)$. In an MDP, the agent receives a reward based on the cost model $\mathcal{R} \coloneqq r_t(s_t, a_t, s_{t+1})$ discounted by the factor γ, and the objective is to find the policy π^* that induces the optimal value function $V^*(s)$:

$$V^*(s) = \max_{a \in \mathcal{A}} \left\{ r(s,a) + \gamma \sum_{s_{t+1} \in \mathcal{S}} p(s_{t+1}|s_t, a) V^*(s_{t+1}) \right\} \tag{5.1}$$

In a POMDP, however, states $s \in \mathcal{S}$ are not directly observed, and instead, observations $o \in \mathcal{O}$ can be collected according to the observation model $\mathcal{Z} \coloneqq p(o_{t+1}|s_{t+1}, a_t)$. Note that the observation model is the likelihood of collecting an observation $o \in O$ after taking an action a_t and transitioned to state s_{t+1}. In an I&M context, the observation model can be directly modeled from probability of detection curves (PoD) or according to inspection noise [81]. A POMDP policy is defined as a function of the dynamically updated history of actions and observations, through the sufficient statistic denoted as belief state $b(s)$. The optimal policy π^* corresponds therefore to the value function [31] that satisfies:

$$V^*(\mathbf{b}) = \max_{a \in A} \left\{ \sum_{s \in \mathcal{S}} r(s,a) b(s) + \gamma \sum_{o \in \mathcal{O}} p(o|b,a) V^*(\mathbf{b}_{t+1}) \right\} \tag{5.2}$$

Assuming a Markovian environment, reasonable in most practical applications and state augmentation techniques [21] can be applied otherwise, an I&M planning decision problem can be therefore formulated as a POMDP.

The identification 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 efforts:

$$\mathbb{E}[r_{tot}] = \sum_{t=0}^{t_N} \left[\gamma^t (r_{ins,t} + r_{rep,t} + r_{F,t})\right] \tag{5.3}$$

where r_{ins}, r_{rep} and r_f stand for inspection, repair and failure costs, respectively, defined as negative rewards. In terms of utility, 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, including environmental consequences specified in monetary units. The definition of the failure risk at the system level will be further elaborated in Section 5.3.

Existing I&M planning applications often model the deterioration evolution d, at the component level, conditional on a set of random variables $\boldsymbol{\theta_d}$ [6, 7, 33] or as a function of the deterioration rate τ [22, 28]. Both formulations are equivalent for modeling deterioration processes, as already demonstrated in [81], and their applicability is case dependent. When observations are collected, through inspections or monitoring, Bayesian updating can then be conducted. Existing algorithms allow exact Bayesian inference if the problem is formulated in a discrete state space [68], as the computation of Bayes' normalization constant is a complex task in continuous state spaces. 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 [7, 81], especially when the problem deals with rare events, e.g. failure events. In general, an efficient discretization aims at minimizing the computational expense while preserving the required level of accuracy.

In a POMDP, the states cannot be directly observed, and the decision maker reasons, instead, under partial observability, only informed by a belief state $\mathbf{b}(s)$, which is formally defined as the probability over states. The belief state reflects the condition of the system conditional on all the taken actions and collected observations. The belief state $\mathbf{b}(s)$ is a sufficient statistic that, for the case of the previously mentioned I&M deterioration models, corresponds to the joint distributions $p(d, \tau)$ or $p(d, \boldsymbol{\theta_d})$, respectively. At each time step, the belief is dynamically updated, with basis on Bayes rule, depending on the action a_t taken and collected observation o_t 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 estimated 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 [68]. At the system level, the belief of each component can be updated by implementing the steps listed in Algorithm 1.

Algorithm 1 Belief update for a system with N_c components

function UPDATEBELIEF($\mathbf{b}_t, \mathbf{a}_t, \mathbf{o}_t$)
  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
    $b(s_{t+1}) \leftarrow b(s_{t+1})/p(o_{t+1}|\mathbf{b}, a_t)$ ▷ normalization step
  end for
end function

While state-of-the-art POMDP solvers often require the modelling of the POMDPs in a flat structure, which can be encoded by augmenting the state space [21] if the process is described by multiple random variables, POMDPs can also be formulated in a factored structure, exploiting the dependence structure among random variables and alleviating thus the required computational effort. We specify here the transition and observation models with basis on conditional structures described by dynamic Bayesian networks (DBNs), and while the belief state $\mathbf{b}(s)$ remains the same than for flat POMDPs, the transition and observation models are constructed by taking advantage of the involved dependencies. For instance, the deterioration rate model is constructed as $p(d_{t+1}|d_t, \tau_{t+1})\, p(\tau_{t+1}|\tau_t)$ instead of $p(d_{t+1}, \tau_{t+1}|d_t, \tau_t)$. The incorporation of conditional structures allows a reduction of the transition model dimensionality from $|\mathcal{S}_d|^2|\mathcal{S}_\tau|^2$ to $|\mathcal{S}_d|^2|\mathcal{S}_\tau| + |\mathcal{S}_\tau|^2$, and might achieve significant computational benefits when multiple random variables are involved. This formulation can be easily applied to simulate the deterioration environment, as explained in Section 5.4, due to the flexibility naturally offered by the proposed deep reinforcement learning approach.

5.3 System-effects in I&M planning

5.3.1 Deterioration dependence in a hierarchical Gaussian structure

Existing methods model the deterioration correlation among components either via random fields or through common influencing factors. Whereas the former are 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 simplifications of physical phenomena, e.g. similar manufacturing techniques or similar loading, lead to shared model uncertainties among the components. In a hierarchical structure, the deterioration of each component is defined conditional on a set of common influencing factors, shared among all the components and represented at the highest level of the hierarchy. In theory, the

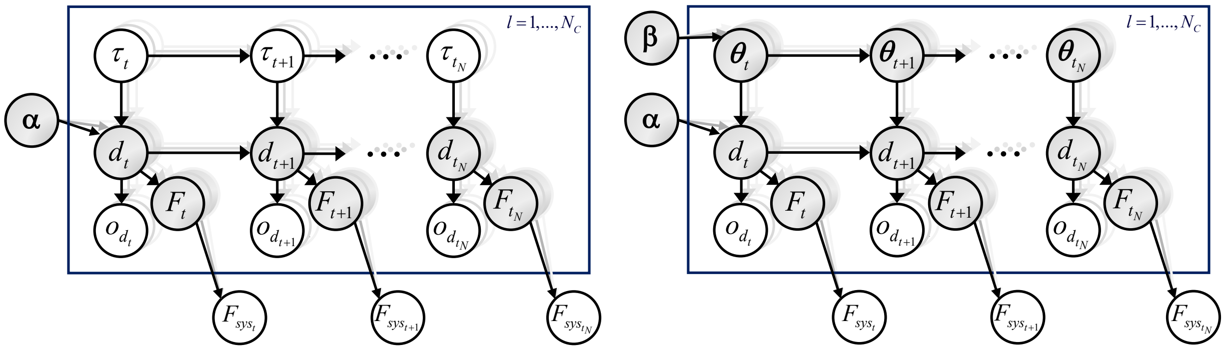


Fig. 5.1. Graphical representation of dynamic Bayesian networks for modeling deterioration processes. At the component level, the damage d evolves over time t as a function of the deterioration rate τ (left) or conditional on a set of parameters $\boldsymbol{\theta_d}$ (right). While d and $\boldsymbol{\theta_d}$ are hidden states, partially observed through the observations o_d, τ is fully observable. The deterioration dependence among components is encoded by the hyperparameter, or set of hyperparameters $\boldsymbol{\alpha_d}, \boldsymbol{\beta_d}$. The binary failure state (survival or failure) of the system F_{sys} depends on the components failure state F.

state space of a system under deterioration dependence can be model directly as the joint space of all the parameters involved in the deterioration process of the system. In that case, the discretized state space would grow exponentially with the number of components included, into a $|S|^{N_C}$ dimensional space. To overcome the increase in dimensionality, we adopt the hierarchical Gaussian structure previously proposed in [33], in which the belief state of each component is encoded conditional on a hyperparameter (common influencing factor) α or set of hyperparameters $\boldsymbol{\alpha}$. The central idea behind this hierarchical structure is that component beliefs for a given hyperparameter $\mathbf{b}(s|\alpha)$ are independent, enabling a decoupling of components joint space. This decoupling alleviates the computational complexity from the original joint space $|\mathcal{S}|^{N_C}$ to a space $|\mathcal{S}| \cdot |\mathcal{S}_\alpha| \cdot N_C$ that grows linearly with the number of considered components N_C. Note that the state space includes now the states of the hyperparameter, which should also be properly discretized. The increase of the state space due to the incorporation of the hyperparameter(s) is however less significant than when considering the whole joint space.

The graphical representation of the proposed hierarchical structure is illustrated in Fig. 5.1, applicable to deterioration processes modeled either as a function of the deterioration rate or conditional on a set of parameters [81]. In any case, the deterioration process$d$ is encoded conditional on the hyperparameter(s) $\boldsymbol{\alpha}$, along with the deterioration rate τ or parameters $\boldsymbol{\theta}$. Evidence collected through observations $\mathbf{o_{d_t}}$ does not only serve for updating the damage state, but also for updating the hyperparameters. Since the hyperparameters are common factors to 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 available. The reliability of the system is represented in Fig. 5.1 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.

This Gaussian hierarchical structure is a mathematically motivated model induced by the convenient formulation available for normal random variables, e.g. the conditional and joint distributions of normal random variables are also normally distributed. 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 $\sqrt{1-\lambda_i^2}$:

$$Y_i = \sqrt{1-\lambda_i^2}X_i + \lambda_i\alpha \tag{5.4}$$

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

$$cov(Y_i, Y_j) = (1-\lambda_i^2)cov(Z_i, Z_j)+\sqrt{1-\lambda_i^2}(\lambda_j)cov(Z_i, \alpha)+\sqrt{1-\lambda_j^2}(\lambda_i)cov(Z_j, \alpha)+\lambda_i\lambda_j cov(\alpha, \alpha) \tag{5.5}$$

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

$$\rho(Y_i, Y_j) = \lambda_i\lambda_j \tag{5.6}$$

If all the components are equi-correlated, then $\lambda_i = \lambda_j = \sqrt{\rho(Y_i, Y_j)}$. Furthermore, the introduced Gaussian structure can be generalized for the case of unequal correlated components, as long as Eq. 5.6 is satisfied. For more complex correlation configurations, one hyperparameter α might not be sufficient to satisfy Eq. 5.6, and in that case, one can incorporate additional hyperparameters $\boldsymbol{\alpha}$, at the expense of a higher computational cost. When multiple hyperparameters $\boldsymbol{\alpha}$ are included, the best fit for $\rho(Y_i, Y_j) = \sum_{k=1}^{m}(\lambda_{ik}\lambda_{jk})$ can be found via optimization procedures, e.g. least squares. Once the Gaussian correlation structure is specified through the parameters $\boldsymbol{\lambda}$, the cumulative distribution of Y_i conditional on the hyperparameter(s) $\boldsymbol{\alpha}$ can be defined as:

$$F_{Y_i|\alpha}(Y) = \Phi\left[\frac{Y_i - \lambda_i\alpha}{\sqrt{1-\lambda_i^2}}\right] \tag{5.7}$$

For the cases in which the deterioration process is modeled by random variables other than Gaussian and considering that a Nataf transformation is applicable, then Eq. 5.7 can be redefined as:

$$F_{d_i|\alpha}(d) = \Phi\left[\frac{\Phi^{-1}[F_d(d_i)] - \lambda_i\alpha}{\sqrt{1-\lambda_i^2}}\right] \tag{5.8}$$

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

$$b(s_{d_i}|\alpha) = F_{S_{d_i}|\alpha}(s^+_{d_i}) - F_{S_{d_i}|\alpha}(s^-_{d_i}) \tag{5.9}$$

5.3.2 Belief update under deterioration dependence

We reformulate the belief update algorithmic scheme introduced in Section 5.2 for a system under deterioration dependence among components. 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) $p(o_{t+1}|\alpha)$, later necessary to update $b(\alpha)$, can be easily computed by marginalizing out the states other than α:

$$p(o_{t+1}|\alpha, a_t) = \sum_{s\in\mathcal{S}}\left[p(s_{t+1}|\alpha)\, p(o_{t+1}|s_{t+1}, a_t)\right] \tag{5.10}$$

Bayesian inference is then conducted on the hyperparameters:

$$p(\alpha|o_{t+1}, a_t) \propto p(\alpha)p(o_{t+1}|\alpha, a_t) \tag{5.11}$$

After updating the conditional beliefs and common influencing factors, the marginal deterioration beliefs can be computed by marginalizing out the hyperparameters α as:

$$b(s_{t+1}) = \sum_{\alpha\in\boldsymbol{\alpha}}\left[p(s_{t+1}|\alpha)\, p(\alpha)\right] \tag{5.12}$$

The effect of maintenance actions on the Gaussian dependence structure has not been explored, up to the knowledge of the authors, in the existing literature [6, 33]. Whereas the originally defined deterioration dependence is preserved if no maintenance interventions are planned, actions might affect the underlying correlation structure. For instance, an initially correlated crack size parameter, associated to common manufacturing processes among the components, will differ after a repair intervention is undertaken. In that case, the

Algorithm 2 Belief update under deterioration dependence

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

transition model $p(s_{t+1}|s_t, a_t)$ of the involved components should be defined accordingly, removing or modifying the deterioration dependence if necessary.

5.3.3 System structural reliability and system cost model

As input to the I&M decision-making problem (Section 5.2), utilities specified according to the decision maker preferences, are assigned to inspection r_{ins} and repair r_{rep} actions. A penalization associated to the annual risk of a system failure r_F is also defined, assigning a monetary value r_f, to the event of structural failure $p_{F_{sys}}$ over consecutive time steps (usually years):

$$r_F = (p_{F_{sys,t+1}} - p_{F_{sys,t}})r_f \tag{5.13}$$

The system structural failure event is specified by a binary variable $p_{F_{sys}}$, failure and survival states, conditional on the belief state $\mathbf{b}(\mathbf{s})$ of the structural components, as illustrated in Fig. 5.1 by the node F_{sys}. In principle, $p_{F_{sys}}$ could be directly defined as a function of the components belief state. In practice, however, $p_{F_{sys}}$ remain only conditional to the event of component failure p_F of the underlying components, specified as:

$$p_F = \sum_{\mathcal{S}_F \in \mathcal{S}} b(s) \tag{5.14}$$

where $\mathcal{S}_F$ corresponds to the state subset classified as failure. Within the deep reinforcement learning approach introduced in Section 5.4, $P_{F,sys}$ can be computed a priori via closed-form procedures and supported by efficient matricial algorithms [92]; or it can be computed following a model-free scheme, obtaining $p_{F_{sys}}$ through a simulator. 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 explored in Section 5.5.

In most structural systems, from bridges to offshore platforms or wind farms, inspection and repair interventions 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. 5.3, adding a fixed campaign cost r_{camp}, incurred every time a campaign is planned, along with inspection r_{ins} and repair r_{rep} costs, assigned to the individual components:

$$r_{tot} = r_{camp} + \sum_{l=1}^{N_C} \left\{ r_{ins}^{(l)} + r_{rep}^{(l)} \right\} + r_F \tag{5.15}$$

5.4 Optimal I&M planning via deep reinforcement learning

I&M planning decision problems, formulated as (factored) POMDPs (Sections 5.2 and 5.3), can be solved in theory by dynamic programming algorithms, e.g. via exact value iteration or policy iteration algorithms [76]. In practice, however, exact value iteration can be applied to only small state space problems due to the complexity associated to finding a policy as a function of the belief state b, which is a probability distribution over states. Recently, I&M planning decision problems, at the component level, formulated as POMDPs and characterized with high-dimensional state spaces have been efficiently solved via point-based POMDP algorithms [22, 81, 84]. Point-based solvers exploit the fact that the value function $V(\mathbf{b})$ (Eq. 5.2) is piece-wise linear convex and can be thus parameterized by a set of $\alpha \in \Gamma$ vectors, each of them associated with a specific action $a \in A$. The optimal policy can be therefore defined in terms of a set of α vectors [59]:

$$\pi^*(\mathbf{b}) = argmax_{\alpha \in \Gamma} \left[\sum_{s \in \mathcal{S}} b(s)\alpha(s) \right] \tag{5.16}$$

State-of-the-art point-based POMDP solvers mainly differ on the approach to sampling belief points, in which a Bellman backup operation is conducted, e.g. in [31, 32, 75]. The reader is directed to [24] 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 reasonably small systems, the dimensionality becomes a limiting factor in high dimensional state, action, and observation space settings, typical in structural systems. Deep reinforcement learning provides then an attractive solution, as value function or policies can be parameterized with artificial neural networks.

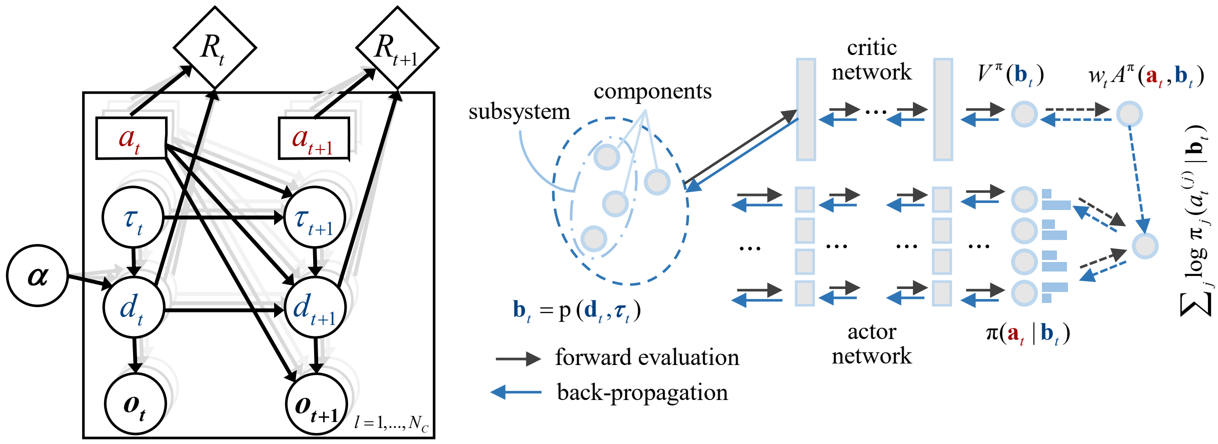


Fig. 5.2. On the left: representation of a factored POMDP derived from the deterioration rate dynamic Bayesian network introduced in Fig. 5.1. The deterioration process d, influenced by the deterioration rate τ and partially observed through o_d, is controlled by the action decision node a. A reward R is collected as a result of taking action a 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.

The interested reader is directed to [93, 94] for a well elaborated introduction on deep reinforcement learning (DRL). In our proposed approach, we integrate the factored POMDP formulation introduced in Section 5.2 with a Decentralized Deep Multi-agent Actor-Critic (DDMAC) scheme, adopted from [28], casting an efficient algorithmic platform for inspection and maintenance planning of structural systems under deterioration, reliability and cost dependence.

Each component of the system is controlled by a stochastic policy $\pi(a|\mathbf{b}, \boldsymbol{\theta}^\pi)$ provided by a group of multi-agent actor networks, as illustrated on the right side of Fig. 5.2 with light blue bars. In most applications, deep reinforcement learning (DRL) policies after training are nearly deterministic, suggesting one action in particular, whereas stochastic policies have proven to be optimal in constraint environments [29]. In our scheme, we consider agents acting as independent units, i.e. the actions taken by one actor do not impact on the state of the other components:

$$\pi(\mathbf{a}|\mathbf{b}) = \prod_{l=1}^{N_C} \pi_l(a^{(l)}|\mathbf{b}) \tag{5.17}$$

The input to the actor networks corresponds to the marginal belief states of all components along with the time step encoded as a one-hot vector. For instance, if the environment is described by the factored POMDP represented on the left side of Fig. 5.2, the actor networks receive the deterioration belief states $\mathbf{b}(s_d)$ and deterioration rate states $\mathbf{b}(s_\tau)$ for all components, plus an input indicating the time step $t \in t_N$. If deterioration

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

Initialize replay buffer
Initialize actor and critic network weights θ^π, θ^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 $\mathbf{a}_t \sim \boldsymbol{\mu}_t = \{\pi_j(\cdot|\mathbf{b}_t, \theta^\pi)\}_{j=1}^{N_c}$
    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 $l = 1, 2, ..., N_c$
    Compute beliefs $\mathbf{b}_{t+1}$: UPDATEBELIEF$(\mathbf{b}_t, \mathbf{a}_t, \mathbf{o}_t)$
    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 $\boldsymbol{\theta}^\pi$ according to gradient:
      $\mathbf{g}_{\boldsymbol{\theta}^\pi} \simeq \sum_i w_i \{\sum_{j=1}^{N_c} \nabla_{\boldsymbol{\theta}^\pi} \, log \, \pi_j(a_i^{(j)}|\mathbf{b}_i, \theta^\pi)\} A_i^\pi$
    Update critic parameters $\boldsymbol{\theta}^V$ according to gradient:
      $\mathbf{g}_{\boldsymbol{\theta}^V} \simeq \sum_i w_i \nabla_{\boldsymbol{\theta}^V} V^\pi(\mathbf{b}_i|\theta^V) A_i^\pi$
  end for
end for

dependence is included through a hierarchical Gaussian model, as explained in Section 5.2, then conditional beliefs $\mathbf{b}(s_d|\boldsymbol{\alpha})$ and hyperparameters beliefs $\mathbf{b}(\boldsymbol{\alpha})$ 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 $\mathbf{b}$, computed by following the steps listed in Algorithm 2. While ReLU activation functions seem to work well for the hidden layers of the actor networks, 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 by according to the noisy rewards collected from a batch of previous experiences, following an offline-training approach that offers more sample efficiency than online-training algorithms. A replay buffer [95] 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 E. 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 $c > 0$ [28]:

$$\mathbf{g}_{\boldsymbol{\theta}^\pi} = \mathbb{E}_{\mathbf{a}_t \sim \boldsymbol{\mu}} \left[w_t \left\{ \sum_{i=1}^{N_c} \nabla_{\boldsymbol{\theta}^\pi} \, log \, \pi_i(a_t^{(i)}|\mathbf{b}_t, \boldsymbol{\theta}^\pi) \right\} A^\pi(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^V) \right] \tag{5.18}$$

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

$$A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^V) \simeq r(\mathbf{b}_t, \mathbf{a}_t) + \gamma V(\mathbf{b}_{t+1}|\boldsymbol{\theta}^V) - V(\mathbf{b}_t|\boldsymbol{\theta}^V) \tag{5.19}$$

The value function is approximated by the critic network, as illustrated on the right side of Fig. 3. Whereas the critic network receives the same input as the actor network (components marginalized beliefs and 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 $A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^V)$, acting therefore, as a critic who is determining how optimal the action taken by the actor network is. The training of the critic network also follows a time difference approach, collecting experiences from the replay buffer, and adjusting the critic parameters $\boldsymbol{\theta}^{\boldsymbol{V}}$ according to the gradient:

$$\mathbf{g}_{\boldsymbol{\theta}^V} = \mathbb{E}_{\mathbf{a}_t \sim \boldsymbol{\mu}}\Big[w_t \nabla_{\boldsymbol{\theta}^V} V^{\pi}(\mathbf{b}_t|\boldsymbol{\theta}^V) A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^V)\Big] \tag{5.20}$$

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 high dimensional state, action and observation spaces. Moreover, the obtained DDMAC policies are intrinsically influenced by the system-effects (Section 5.3) 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 5.2 enables the identification of optimal I&M policies, that intrinsically include the following system-effects:

i) Deterioration dependence among components: 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) $\mathbf{b}(s_d|\boldsymbol{\alpha})$. Under deterioration dependence, information collected by inspecting one component provides information to the other components, influenced by the specified deterioration correlation. The influence of this system-effect on the policy is explored via numerical experiments in Section 5.5, both for a 9-out-of-10 system and for a steel frame structural system subject to fatigue deterioration.

ii) System structural reliability: the utilities associated to the failure risk are computed at the system level, penalizing with a negative reward r_F that is defined as a function of the components health, see Eq. 5.13. The actors, even if acting individually, are trained with respect to the overall 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 a steel frame structural system (Section 5.5).

iii) Inspection and maintenance cost model: a campaign cost r_{camp} can be included, in some applications, as a fixed cost if any component is inspected or repaired, plus an additional inspection or repair cost for each inspected or repaired component, see Eq. 5.15. Since DDMAC collects rewards at a system level, a campaign cost model might influence the resulting I&M policies, arranging inspection and repair actions at the same time step, as observed in the numerical experiments conducted for the 9-out-of-10 system (Section 5.5).

5.5 Numerical experiments

DDMAC inspection and maintenance policies are tested for a 9-out-of-10 system under fatigue deterioration, exploring different deterioration dependence and cost model settings. A second set of numerical experiments is conducted to investigate the efficiency of DDMAC policies for a 2D steel frame, commonly known as Zayas frame, used as a benchmark structural system for offshore engineering collapse analyses [96, 97]. 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 [63]:

$$d_{t+1} = \left[\left(1 - \frac{m}{2}\right) C_{FM} S_R^m \pi^{m/2} n + d_t^{1-m/2} \right]^{2/(2-m)} \tag{5.21}$$

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 5.1. Following a through-thickness failure criterion [91], the failure limit state at time step t can be formulated as:

$$g_t = d_c - d_t \tag{5.22}$$

Table 5.1. Random variables and deterministic parameters utilized to model the fatigue deterioration for the components in the numerical experiments.

Variable	Distribution	Mean	Standard Deviation
$ln(C_{FM})$	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 5.2. Description of the discretization scheme implemented for the factored deterioration rate POMDP.

Variable	Interval boundaries
Deterioration rate model	
S_d	$0, \exp\left\{\ln(10^{-4}) : \dfrac{\ln(d_c) - \ln(10^{-4})}{\lvert S_d\rvert - 2} : \ln(d_c)\right\}, \infty$
S_τ	$0 : 1 : 30$

The fatigue deterioration is encoded in a deterioration rate DBN model, and ultimately shaping a factored POMDP, as shown on the left side of Fig. 5.2. The continuous crack depth d is discretizated into 30 s_d states conditional on 31 fully observable deterioration rates s_τ states. The intervals and state space utilized for this deterioration rate model are listed in Table 5.2.

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 [81].

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

The system explored in this application is composed of ten components, each of them subjected to 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), characterized thus with more redundancy than 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, as a function of the failure state of all components, by following the recursive method proposed in [98].

Description of the I&M decision problem

A total of eight I&M planning scenarios are investigated, exploring different deterioration dependence among components, as well as cost model settings. 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. 5.3. After a Gaussian hierarchical structure with two hyperparameters is optimized, by computing the $\boldsymbol{\lambda}$ parameters with the objective of satisfying Eq. 5.6, an approximated correlation structure is obtained with relatively small errors, as shown on the right side of Fig. 5.3. 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, with an additional computational expense, as explained in Section 5.3. For each of the aforementioned environments, specified with different deterioration dependence, two I&M cost models are further investigated: an I&M cost model that charges inspection and repairs individually; and an I&M cost model in which an initial campaign cost is activated, 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 $\mathcal{S}_d$ and 31 deterioration rate states $\mathcal{S}_\tau$, the state space of the system sums, therefore, up to a total of 9,300 states, for the experiments that do not include deterioration correlation ($\rho_{eq} = 0$). Otherwise, experiments under equal correlation (ρ_{eq}) sum up to 74,480 states, rising up to $59.5 \cdot 10^7$ states 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 modelled equally for all components, only one hyperparameter is sufficient to satisfy Eq. 5.6, 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 the importance of optimizing the number of hyperparameters included in the model, as the state space grows exponentially with the

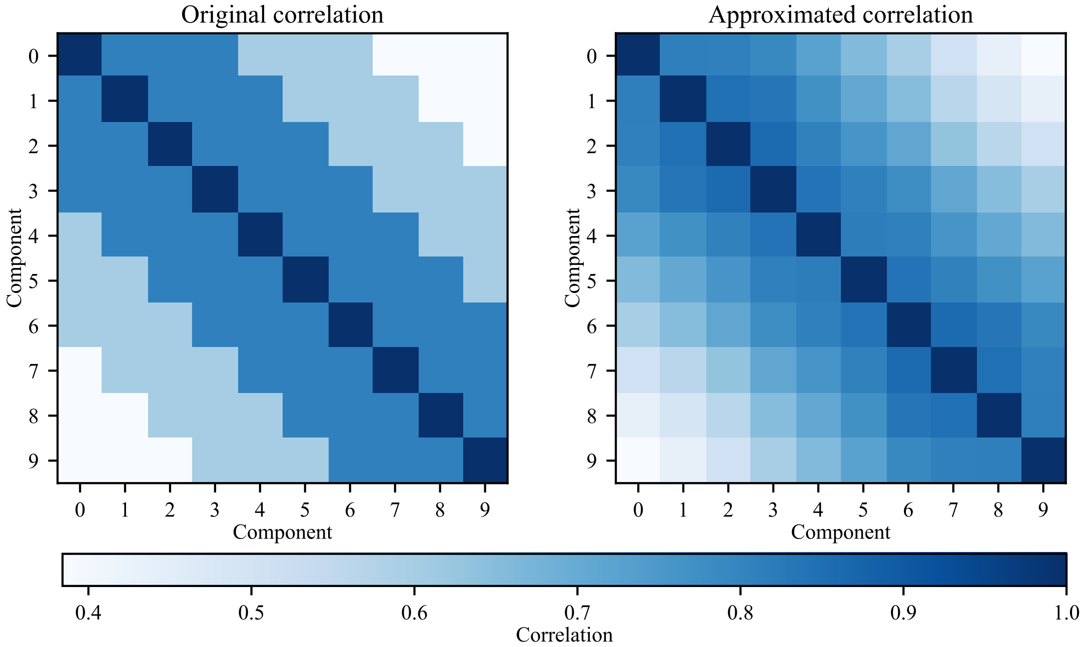


Fig. 5.3. Representation of the initial crack size dependence among the components of the unequally correlated 9-out-of-10 system. The original deterioration correlation is represented by the colored matrix on the left. The approximated correlation structure, resulting from the derived Gaussian hierarchical model with two hyperparameters, is displayed on the right colored matrix.

number of considered random variables. By formulating the POMDPs' transition model as dynamic Bayesian networks, their dimensionality is reduced from $|\mathcal{S}_d|^2|\mathcal{S}_\tau|^2$, in a flat structure, to $|\mathcal{S}_d|^2|\mathcal{S}_\tau| + |\mathcal{S}_\tau|^2$ 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 enables the decoupling of the joint space at the system level, which would grow exponentially for a flat POMDP structure, and instead grows linearly. For instance, the setting under unequal deterioration dependence in a joint space would be described by 930^{10} states, and instead is defined by $930 \cdot 10 \cdot 80^2 + 80 \cdot 2 \simeq 59.5 \cdot 10^6$ states in the hierarchical model, with two hyperparameters 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.

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 / no-inspection* is not included 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 repairs are charged per component, $r_{ins} = -1$ and $r_{pr} = -20$, respectively, while 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 $r_{pr} = -20$ money units, respectively. The consequence of a system structural failure is penalized with $r_f = -10,000$ 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 heuristics decision rules, inspired by [6], 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 inspected, 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 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.

Results and discussion

The life-cycle expected costs obtained by evaluating the investigated policies are displayed in Fig. 5.4, 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: 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 the explored numerical examples, DDMAC outperforms the optimized heuristics, yielding life-cycle cost savings 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 less inspection, repair, and failure expected costs 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 detection instance.

With regard to deterioration dependence, highly correlated environments result generally in less expected total costs, as observed in Fig. 5.4. Information collected on one component, in environments under deterioration correlation, also provides information to the other components in the system.

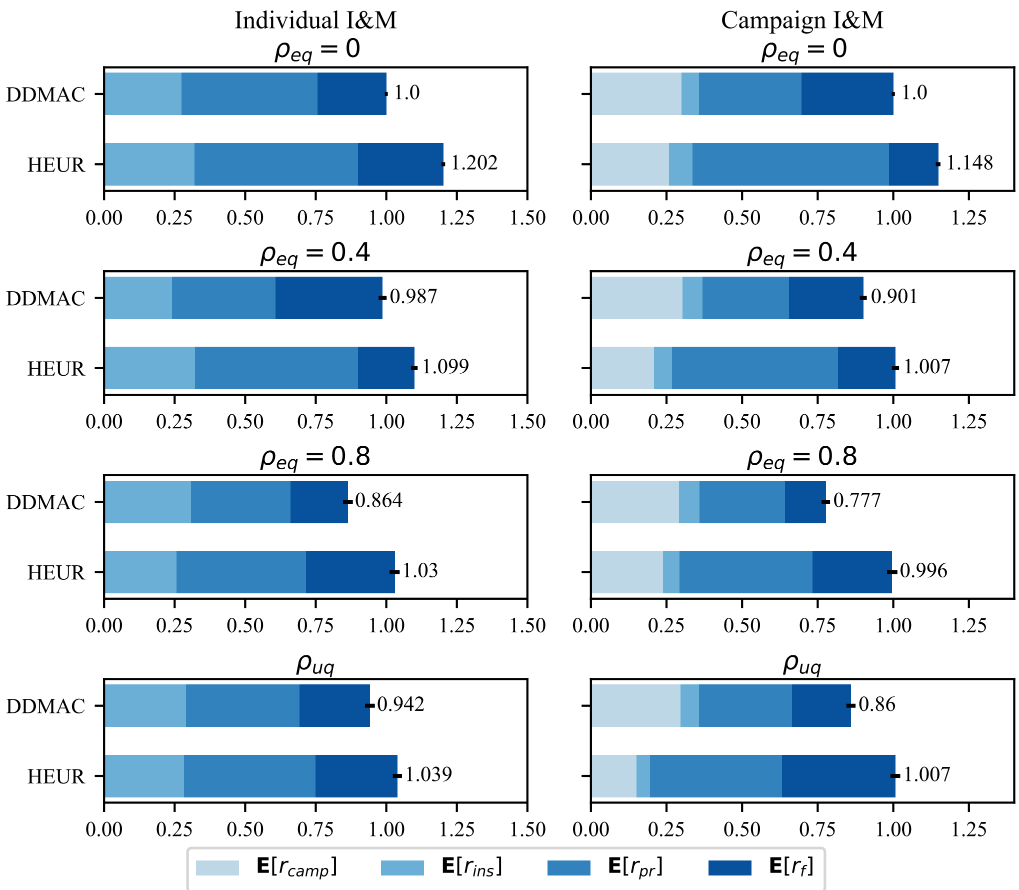


Fig. 5.4. Expected cost results of all the numerical experiments conducted for the 9-out-of-10 system, divided into campaign $\mathbf{E}[r_{camp}]$, inspection $\mathbf{E}[r_{ins}]$, perfect-repair $\mathbf{E}[r_{pr}]$ and failure $\mathbf{E}[r_f]$ expected costs. On the left, DDMAC and heuristic policies, 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.

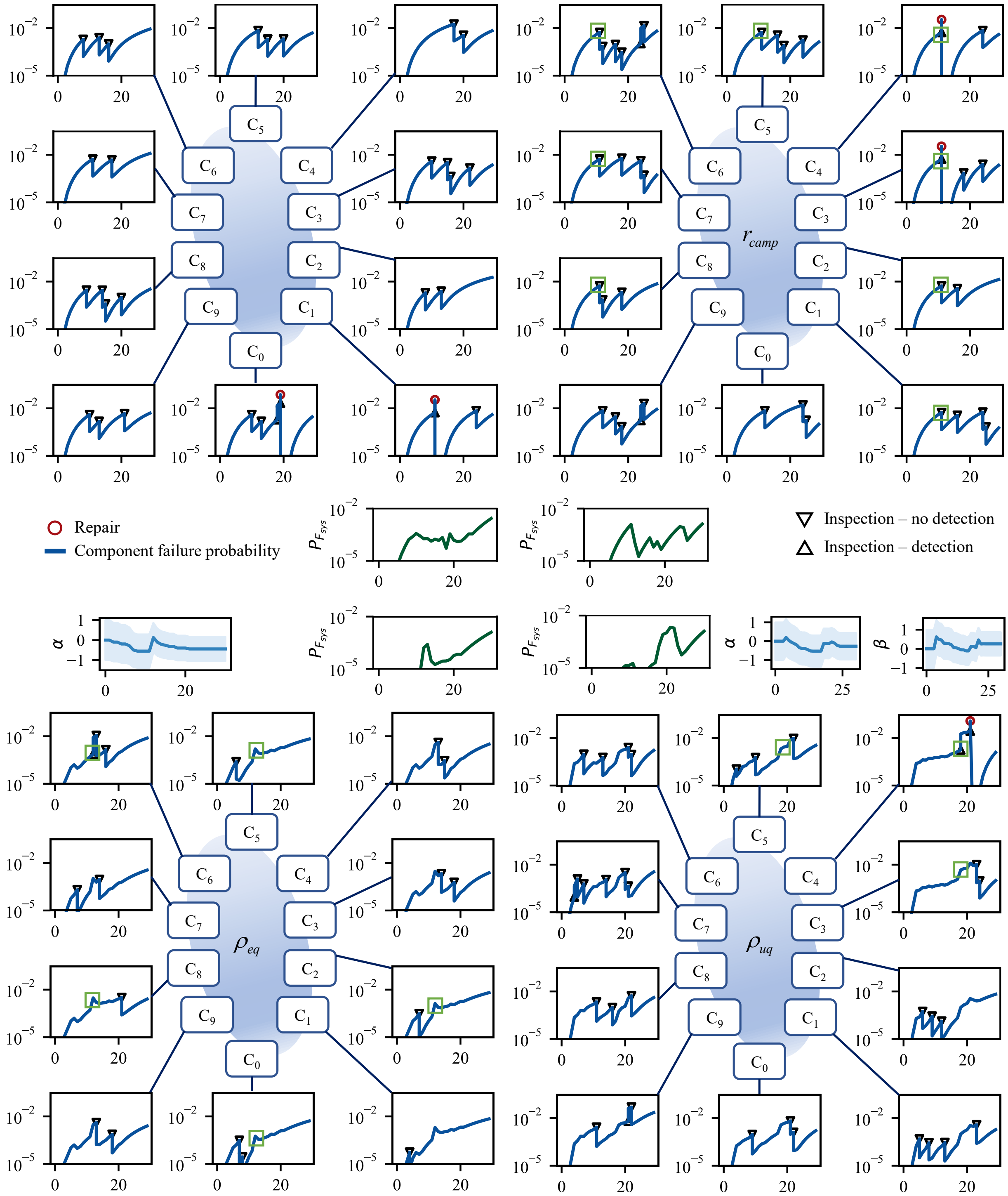


Fig. 5.5. 9-out-of-10 system policy realizations: (Upper-left) uncorrelated deterioration & individual cost model; (Upper-right) uncorrelated deterioration & campaign cost model; (Lower-left) equally correlated environment ($\rho_{eq} = 0.8$) & individual cost model; (Lower-right) unequally correlated deterioration & 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.

For instance, a crack detection observed on component 5 for the case of equal correlation $\rho_{eq} = 0.8$, leads to an increment of failure probability on the other components, as clearly displayed by green rectangles on the lower-left corner of Fig. 5.5. Components are, however, informed from other inspected components according to their degree of correlation. This effect can be better 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 square of Fig. 5.5. In this case, components 3 and 5, highly correlated with component 4, as indicated in Fig. 5.3, are clearly affected by the observed crack detection.

Moreover, highly correlated deteriorating environments induce more variability on the expected total costs, as shown by the black error bars in Fig. 5.4. The variability can be attributed to the very different policies that result depending on whether the collected inspections observe cracks or not. If a crack is detected on one component, the other components' failure probability will increase, and repairs actions or additional inspection will be planned thereafter, whereas 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 less inspections, as it would be expected due to the additional information gained through the underlying correlation among components, but highly correlated environments might plan more inspections, while resulting in significant failure risk reductions, as displayed for the case with $\rho_{eq} = 0.8$ in Fig. 5.4.

To further investigate the effect of including campaign utilities within the cost model, a histogram over $3 \cdot 10^5$ policy realizations is shown in Fig. 5.6 for a DDMAC policy in which inspections and repairs are charged separately (light blue) and another DDMAC policy adding the expense of campaign actions (dark blue).

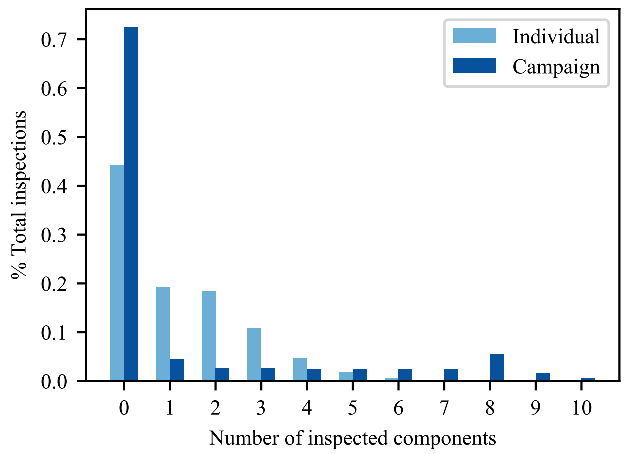


Fig. 5.6. Comparison of DDMAC policies specified with either individual (light blue) or campaign (dark blue) I&M cost models. For each case, the number of inspected components per time step is represented in a histogram over 10,000 policy realizations.

The deterioration environment for both DDMAC policies does not consider deterioration correlation among components. The emphasis of Fig. 5.6 is on the number of components inspected every occasion an inspection is planned. If inspection and repairs are paid individually for each component, only one or two components are usually inspected, being rare to inspect more than five components per time step. In contrast, eight inspected components becomes the predominant inspection decision if an initial campaign cost is included in the cost model. The system-effect can be visualized in the policy realizations shown at the top of Fig. 5.5, in which green rectangle marked components inspected at the same year for the *campaign* DDMAC policy. The policy for the campaign cost model tends thus to group inspections and repair actions at the same year, avoiding if possible unnecessary campaign costs associated with one or two inspection campaigns. In few cases, campaigns are planned for only one or two components according to the DDMAC *campaign* policy, contrasting with the static inspection decision rules imposed by heuristics, in which a specific number of inspection will be fixed for all the inspection campaigns. Based on this reasoning, we demonstrate the capability of DDMAC to adjust the I&M policy according to the specified cost model, whether charging campaign costs or inspection and repairs individually, and providing a more flexible and adaptive policy than static heuristic policies.

5.5.2 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 components and the effects of including a campaign cost within the cost model. In this second application, we 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, target of many offshore structural analysis benchmarks. Zayas is composed of two columns, which along with 13 braces, sustain a rigid beam at the top. Geometry and material properties of the Zayas frame are described in [99].

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 an independence among components' deterioration. The state space for the former includes 30 crack states along with 31 deterioration rate states, for each of the 22 hotspots, resulting therefore in a total 20,460 states; while the state space for the latter climbs up to approximately $1.6 \cdot 10^6$ states, including 80 states from one discretizatized hyperparameter.

The benefits associated with the proposed decoupled hierarchical structure are very significant, the state space would increase otherwise up to 930^{22} states if the joint space of all hotspots was explicitly considered.

Similarly as for the experiments reported in Section 5.5, 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*. The inspections provide binary crack indications (detected or no-detected), equally modeled for each component by the observation model described earlier in this Section. With respect to the cost model, inspections and repairs (planned individually for each component) cost $r_{ins} = -1$ and $r_{pr} = -15$ money units, respectively, while the system failure cost is defined as $r_f = -50,000$ money units. All the costs are discounted to the present value by a $\gamma = 0.95$ factor. DDMAC's architecture is fairly 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 exploration settings are equally defined as for the first application. The investigated heuristic-based policies are based on the same set of decision rules introduced in the former numerical experiments, amounting in this case, to all the combinations of inspections intervals Δ_{ins} and inspected hotspots per campaign n_C. Both DDMAC and heuristic policies are evaluated over 10,000 episodes and the results, in terms of expected total costs, are showcased in Fig. 5.7.

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 [33, 100], a total of 22 hotspots are considered, located at the joints between the braces or columns, each brace having associated either one or two hotspots, as illustrated in Fig 5.8. Each brace contains either one or two hotspots or critical locations for fatigue deterioration, resulting in a total of 22 hotspots. 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 5.5).

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 under the action of gravity. At a component level, the health of each hotspot is described by the vector $\mathbf{x_h}$, in which x_h is a binary variable with $x_h = 0$ indicating a hotspot failure and $x_h = 1$ corresponding to a hotspot survival. The failure probability of a hotspot $p_F^{(h)}$ corresponds thus to the probability of being in state $x_h = 0$. At an element level, the state of each brace is represented as well by a binary vector $\mathbf{x_{el}}$, attributing $x_{el} = 0$ 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. $p(x_{el} = 0)$, can be therefore computed as a series system:

$$p_F^{(el)} = 1 - \prod_{h \in el} \left[1 - p_F^{(h)}\right] \tag{5.23}$$

At a global 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 8192 (2^{13}) non-linear progressive collapse, also denoted static push-over, simulations have been run with the assistance of the computer code 'USFOS' (available within the software package Sesam), before the training of DDMAC, and providing explicitly the failure probability of the system conditional on all element state combinations. The element configuration for each collapse simulation is arranged according to the element state vector $\mathbf{x_{el}}$, removing the braces associated with a failed state $x_{el} = 0$. For instance, if only the first element is in a failed state, the collapse analysis for that case is conducted by removing the failed element from the original frame. The resistance $L_{col}(\mathbf{x_{el}})$ of each element state combination is retrieved from the conducted collapse simulations.

The collapse event of the frame is defined as the probability of the horizontal load exceeding the structural system resistance $p(L > L_{col})$. In this case, the horizontal is load is modeled as a lognormal random variable with mean μ_L = 70kN and 25% coefficient of variation, while no uncertainty is associated with the resistance, being a reasonable assumption when the load is highly uncertain in comparison with the resistance [100]. The failure probability of the system p_{Fsys} conditional on the element state vector $\mathbf{x_{el}}$ can be defined directly from the probability density function of the load:

$$p_{Fsys}^{(x_{el})} = 1 - \int_{L_{col}}^{\infty} f_L(x)dx \tag{5.24}$$

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 10^{-4}. The state of the frame is, however, computed conditional on the state of all the elements, and for that the probability of being in each state combination should be computed. We follow the iterative procedure proposed in [92] 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 $p_F^{(el)}$:

$$\begin{aligned} \mathbf{q_{[1]}} &= \left[p_F^{(1)} \quad \bar{p}_F^{(1)}\right]^T \\ \mathbf{q_{[i]}} &= \begin{bmatrix} \mathbf{q_{[i-1]}} \cdot p_F^{(i)} \\ \mathbf{q_{[i-1]}} \cdot \bar{p}_F^{(i)} \end{bmatrix} \end{aligned} \tag{5.25}$$

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

$$p_{Fsys} = \sum_{\mathbf{x_{el}}} \left[p_{Fsys}^{(x_{el})} \cdot q^{(x_{el})} \right] \tag{5.26}$$

Results and discussion

The comparison between DDMAC and optimized heuristics follows the same trend as for the 9-out-10 structural system experiments. In terms of expected life-cycle costs, DDMAC policies outperformed heuristic-based policies in the two tested settings, as shown in Fig. 5.7, with costs 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. Under deterioration dependence, i.e. the initial crack size among the hotspots is correlated, observation collected at one hotspots provide information to other hotspots, updating the belief damage state. The belief is updated for both detection and no detection observation outcomes, as illustrated in Fig. 5.8. At year 12, a crack is detected at the lower X-brace, and this observation leads to a higher failure probability of the other components, marked with a green rectangle in the diagram, effect that can be observed clearly in the updated mean of the hyperparameter α. In most of the policy realizations, however, the likely observation outcome is no-detection, explaining this way the risk reduction.

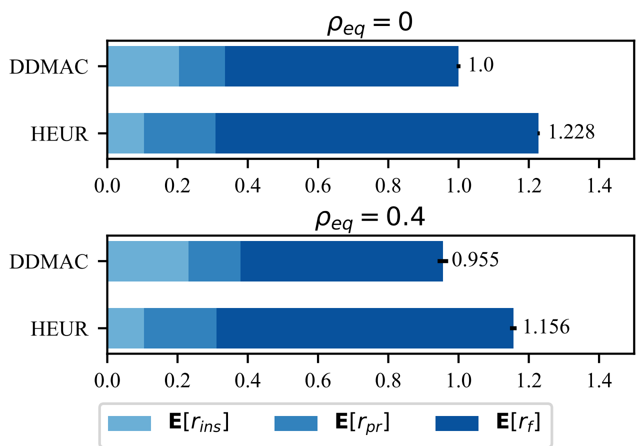


Fig. 5.7. Expected cost results of all the numerical experiments conducted for the Zayas frame, divided into inspection $\mathbf{E}[r_{ins}]$, perfect-repair $\mathbf{E}[r_{pr}]$ and failure $\mathbf{E}[r_f]$ expected costs. DDMAC and heuristics are compared under an uncorrelated deterioration environment at the top, and under an equally correlated environment ($\rho_{eq} = 0.4$) at the bottom.

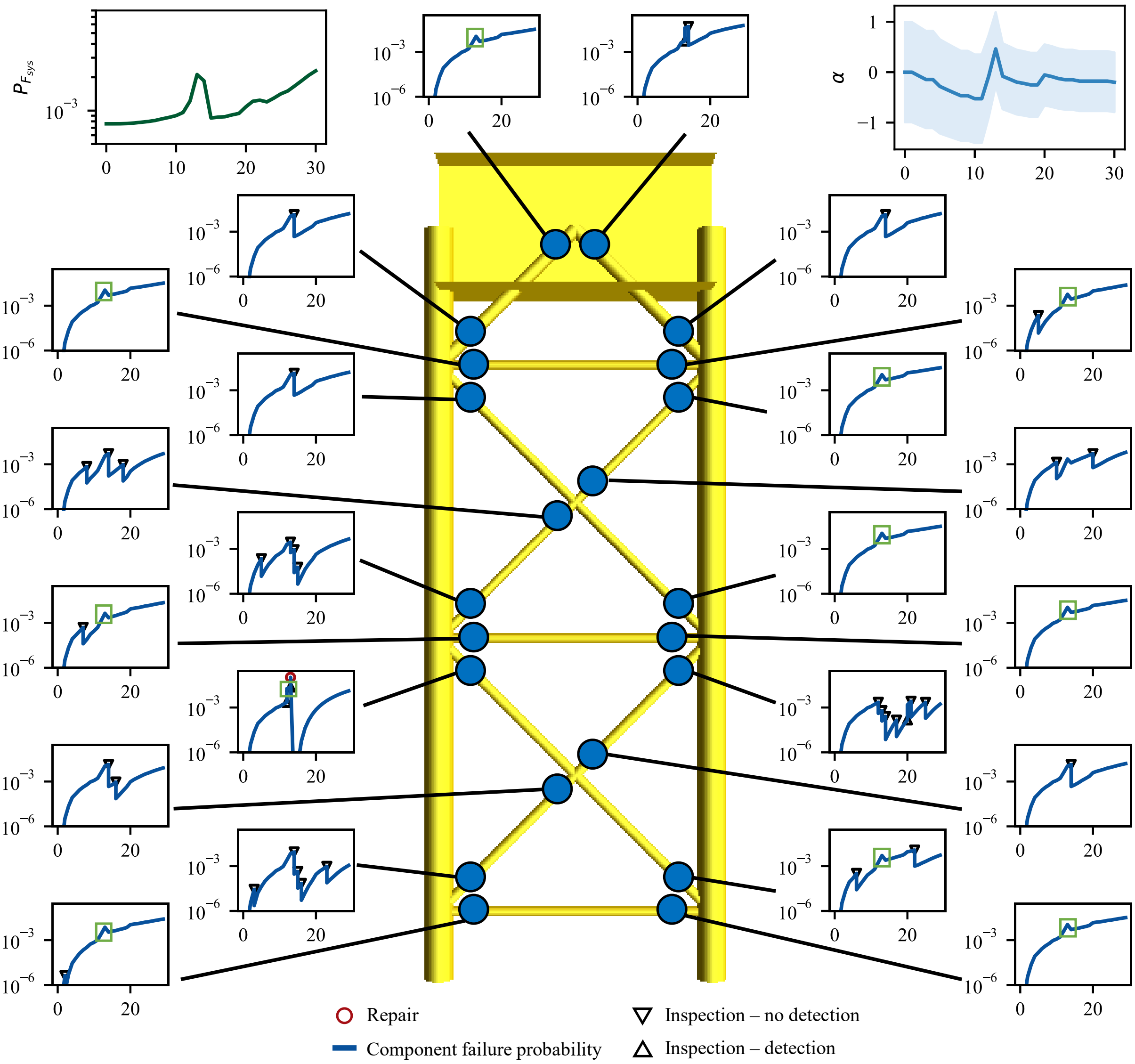


Fig. 5.8. Zayas frame policy realization in an equal deterioration correlation environment ($\rho_{eq} = 0.4$). The failure probability of each hotspot is depicted by a blue line, inspection indications are represented by upwards (detection) or downwards (no-detection) triangles, and perfect repairs are circled in red. At the system level, the failure probability and system-effects 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.

DDMAC intrinsically includes the importance of each hotspot to 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 [101], 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 ($\sim el$).

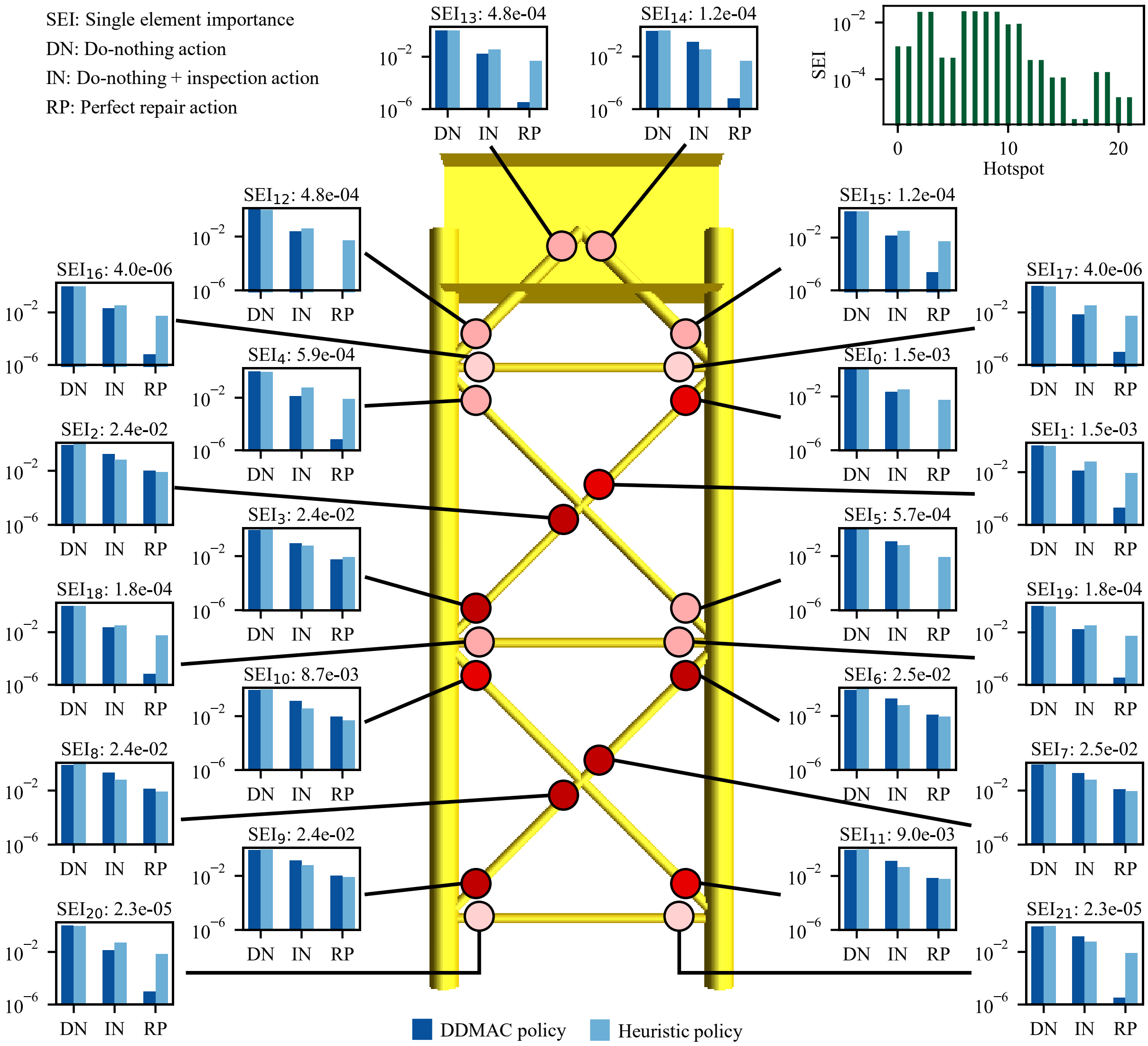


Fig. 5.9. Histograms of DDMAC and heuristic actions over 10,000 Zayas frame policy realizations, each episode with a span of 30 years. The single element importance metric (SEI) associated with each fatigue hotspot is indicated at the top of each histogram and summarized at the green top-left diagram. The relative importance of each hotspot is also represented by color, with a darker red being a more critical element to the structural reliability of the system.

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 the importance of each hotspot as:

$$SEI_h = p_{F_{sys}}^{(\sim h)} - p_{F_{sys}} \tag{5.27}$$

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. 5.9, along with histograms of the actions taken at each component during $3 \cdot 10^5$ DDMAC (dark blue) and optimized-heuristic (light blue) realizations. As represented by the dark green bar diagram at the top-right corner of Fig. 5.9 and in agreement with the findings reported in [100], the critical hotspots are located at the X-braces, whereas the less critical hotspots are the ones connecting the horizontal braces. While *do-nothing* action predominates and *inspection* actions are distributed similarly among components, the balance of *repair* actions among hotspots differs for DDMAC and heuristics. DDMAC plans repairs mainly for important elements to the global structural reliability, i.e. with a high SEI, such as hotspots 6 and 7, whereas less important components to the system structural reliability are less repaired. In contrast, the heuristic-based policy plans repairs nearly balanced among components, disregarding the influence of each hotspot to the reliability of the system. We can therefore conclude that DDMAC policies are able to capture the system-effect attributed to the importance of each element to the system structural reliability.

5.6 Concluding remarks

This paper introduced an efficient algorithmic platform for optimal decision-making under uncertainty of engineering systems exposed to deteriorating environments. The decision-making problem is formulated as a factored Partially Observable Markov Decision Process (POMDP), in which the dynamics are encoded in Bayesian network conditional structures. The experiences collected by simulating the specified POMDP are fed to a multi-agent actor-critic, who is able to identify optimal strategies in high dimensional state, action, and observation spaces, commonly found in practical structural systems. In particular, we demonstrated through numerical experiments that the proposed approach provides more optimal inspection and maintenance (I&M) policies than state-of-the-art policies based on heuristics and enables a systematic treatment of 'system-effects', that is reflected in the identified strategies.

POMDP-based policies, approximated in high-dimensional settings by the decentralized deep multi-agent actor-critic (DDMAC), map the current belief state of the system to a probability distribution over the possible actions. The stochastic policies are thus prescribed as a function of the belief state, which is a sufficient statistic defined as the, dynamically updated, probability distribution over states, and constituting the history of actions and observations. Constructing the policies based on a sufficient statistic feature enables more optimal decision-making strategies than static optimization methods or adaptive heuristic approaches, constrained by the limited space explored during the policy search. POMDP-based policies provide an additional flexibility to the decision maker, who can opt for an alternative decision at some point, and the policy will be automatically adapted thereafter, yielding near-optimal results.

Furthermore, DDMAC policies are approximated by actor neural networks, whose weights are adjusted 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 intrinsically consider the following system-effects:

- In deterioration dependent environments, observing the state of one component provides indirect information, modulated by the degree of correlation, to the other components of the system. In the tested I&M planning scenarios, more correlated environments resulted in a reduction of expected costs, usually characterized with less expected failure risks. As structural systems are designed according to high reliability standards, demanding a low failure risk, observations mostly indicate benign structural states, which is translated in highly correlated environments, as a global reduction of failure risk. In contrast to independent deterioration settings, more variability in the expected costs is observed in dependent environments, in which very different I&M policy scenarios can be experienced. If benign observations derived to a healthy system belief state, less interventions actions are planned, whereas a single negative observation on a component state is reflected as a global warning to the system belief state, resulting in a very intrusive intervention policy.

- A clustering effect, on inspections and repairs, is observed in settings that are specified with a campaign cost model, i.e. a fixed cost is activated if any component is repaired or inspected. In this case, policies seek to either avoid planning single or few inspections and repairs at one time step; instead, inspection and maintenance actions are frequently grouped, saving the additional campaign cost associated to inspecting 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 allocated to critical elements, whereas less important components to the global reliability were seldom repaired.

In this work, the deterioration environment is formulated as a discrete state POMDP, in which exact Bayesian inference can be conducted. Further research should be focused on the development of continuous state POMDPs or optimization procedures that would allow a reduction of the state space dimensionality, facilitating therefore the treatment of very large systems. Also, research efforts are encouraged on the derivation of optimality bounds for deep reinforcement learning policies, similar to those available for POMDP point-based solvers.

Authorship contribution statement

Morato, P. G.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Andriotis, C. P.:** Conceptualization, Methodology, Validation, Formal analysis, Software, Writing - Review & Editing. **Papakonstantinou, K. G.:** Methodology, Validation, Formal analysis, Resources, Writing - Review & Editing, Supervision. **Rigo P.:** Supervision, Project administration.

Acknowledgements

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 and Dr. Andriotis would further like to acknowledge that this material is also based upon work supported by the U.S. National Science Foundation under Grant No. 1751941. Finally, the authors would like to acknowledge the support provided by DNV GL Digital Solutions for granting access to the software package 'Sesam'.

Chapter 6

Optimal Management of Offshore Wind Structural Systems via Deep Reinforcement Learning and Bayesian Networks

---

Paper Morato, P. G., Andriotis, C. P., Papakonstantinou, K. G., Hlaing, N. and Rigo, P. (2021). Optimal Management of Offshore Wind Structural Systems via Deep Reinforcement Learning and Bayesian Networks. *Engineering Structures*, under internal review.

Abstract Structural systems are exposed to loads and deterioration mechanisms throughout their planned lifetime. The uncertainties associated with the prediction of such deterioration processes can be reduced by collecting information from monitoring and inspections. To systematically quantify the benefit of a monitoring scheme, most existing methods compute the gain or loss in life-cycle expected cost of the considered monitoring system, assuming the sensors are continually operational. In this work, we include monitoring choices within the sequential decision-making problem by formulating the monitoring observation model conditional on the sensors' health, i.e. monitoring observations can only be collected if the sensors are operative. We additionally provide recommendations for the modeling of the overarching decision-making problem under uncertainty and imperfect observations as a Partially Observable Markov Decision Process (POMDP), with a strong emphasis on the management of offshore wind substructures subject to fatigue deterioration. Due to the high-dimensional state, action, and observation spaces featured by practical structural systems, we approximate POMDP policies with actor neural networks, who guided by critic neural networks, can identify optimal monitoring, inspection, and maintenance policies. The proposed algorithmic platform is tested for the management of offshore wind substructures, both at the offshore wind turbine and offshore wind farm levels, in which each examined structural detail is subject to various degrees of fatigue intensity and accessibility constraints. In the investigated setting, observations can be collected from both structural response monitoring and non-destructive inspection techniques. The

results reveal that the advanced actions and observations decisions patterns provided by POMDP-based deep reinforcement learning policies offer substantial savings compared to corrective, calendar, and heuristic decision rules strategies.

6.1 Introduction

Engineering systems consist of various components that jointly perform an intended function. In terms of structural reliability, the system should be able to resist the combined exposure to loads and deterioration mechanisms throughout its lifetime. Estimation of this reliability can be challenging from a computational and modeling standpoint, due to the high-dimensional spaces formed by multi-component deterioration processes. In addition, system-level responses and decisions are naturally defined on combinatoric spaces of component states and actions that quickly become practically intractable even for small systems. To circumvent these complexities, most inspection and maintenance planning approaches resort to component level decision rules, which are, however, often detached from global system metrics and optimality. Within this context, offshore wind systems, exposed to harsh marine conditions and experiencing fatigue and corrosion deterioration due to the combined action of wind and wave loads, require efficient system management strategies. Such management solutions should be able to optimally balance joint information decisions and effective maintenance actions, in order to optimize sought objectives, while also minimizing the inherent structural failure risks.

The majority of the offshore wind maintenance planning methods available in the literature rely on failure rate data collected from records of failure statistics for projecting reliability estimates, e.g. [102, 103, 104, 105, 106]. However, failure statistics of offshore wind substructures are usually unavailable, and instead, system structural reliability estimations are a priori drawn through simulations, often based on physical models. Throughout the operational lifetime, information collected from inspections and/or monitoring systems can be used to reduce the underlying epistemic uncertainties. Early inspection and maintenance planning methods for the management of offshore wind substructures [8, 107] are mainly derived from risk-based inspection planning schemes previously developed for the management of oil and gas offshore structures [52, 62, 108, 109]. An offshore wind substructure failure event is, however, associated with less consequences than the failure of an oil and gas platform, due to the limited likelihood of human fatalities in unmanned operated facilities [110]. In this case, traditional risk-based inspection planning strategies defined based on prescribed inspections and maintenance decision rules, e.g. undertaking repairs after observing a crack, might results in less optimal policies [81].

More recently, offshore wind monitoring schemes have received increasing attention, registering a myriad of research efforts in the literature [111, 112]. Whereas the engineered structural health monitoring systems are, in most cases, not intertwined with quantitative decision-making strategies, some proposed inspection and maintenance planning methods, relying on classical decision theory [1], assess the expected benefits of monitoring schemes through value of information analyses [113]. Some investigations compute the conditional value of monitoring, reducing the decision-making to a posterior analysis [90, 114], while other approaches include monitoring within the pre-posterior decision-making problem [9, 60], considering the monitoring equipment to be fully operational during the total extent of the planned horizon.

In this study, we explore the efficiency of an integrated monitoring, inspection and maintenance planning framework, incorporating dynamic Bayesian networks (DBNs) in actor-critic deep reinforcement learning (DRL), capable of finding optimal strategies in high-dimensional state, action, and observation spaces [28]. Beliefs about system states and deterioration are stochastically updated through inspection action outcomes, or observations, while maintenance actions are probabilistically controlling the component and system conditions. In particular, in the proposed method, we dynamically propagate the system uncertainty through conditional DBN formations, conducting Bayesian inference when information from inspections or monitoring becomes available [81]. Optimization of this inspection and maintenance planning decision-making problem adheres to stochastic optimal control premises and Partially Observable Markov Decision Processes (POMDPs), a principled mathematical framework for optimal decision-making under uncertainty. The computational complexity of identifying the POMDP optimal policy increases with the number of states, actions, and observations. With the implemented DRL approach, this complexity is alleviated, since the actor parametrizes the optimal policy of the system with a multi-agent network, and the critic approximates the value function reflecting the cumulative long-term joint reward of the system, both conditioned on marginal component beliefs. Experience from policy realizations generated by an epsilon-greedy learning scheme is collected in a replay memory, which is then sampled for the training of both actor and critic networks.

To demonstrate the efficiency of the proposed approach, we apply the methodology to the case study of an offshore wind farm management, in which the structural components are subjected to fatigue deterioration. Monitoring, inspection and maintenance actions can be planned for any individual component or a group of components, incurring an initial campaign cost plus an increment for each individually inspected or repaired component. Besides non-destructive inspections, information about the system response can be also monitored by strain gauges or other instrumentation, further enabling probabilistic inference of mechanics-based structural response metrics.

The expected total rewards of the obtained DRL policies are compared with conventional decision rules, verifying the successful applicability and efficiency of the developed methodology to approximate globally optimal monitoring, inspection, and maintenance policies in complex, large-scale structural system settings.

6.2 POMDP formulation of the I&M planning decision problem

We adopt here the formulation introduced in our earlier work [115], defining a factored POMDP with transition and observation model specified as Bayesian networks, as the underlying mathematical foundation for the inspection and maintenance decision-making problem. Along with the summary of the proposed POMDP formulation, recommendations are also provided for the management of offshore wind substructures at the system level, in which the policy optimization is conducted for an offshore wind substructure constituted by various components. At the offshore wind substructure level, the structural reliability of the system depends on the deterioration process experienced by the constitutive structural components, which may be characterized by harsher or more benign deterioration rates. Also, the inspection and maintenance interventions might be charged differently, e.g. a weld located below the water level demands remotely operated vehicles or skilled divers, who are otherwise not required if the structural detail is above the water level. In any case, the main goal is to optimally control the risk of adverse events, e.g. structural failure of the substructure, by timely planning inspections and maintenance interventions.

6.2.1 (Partially Observable) Markov Decision Processses

Due the uncertainties involved in the estimation of the deterioration process, the decision maker, who might also be denoted as intelligent agent, is operating in an uncertain environment. If the agent directly observes the state of the environment at each time step, the decision-making making problem can be formulated as a Markov Decision Process (MDP). If the environment under consideration is non-stationary, state-augmentation techniques [21, 81] can be employed to transform the environment dynamics into those of a Markov Decision Process. In an MDP, the agent takes an action $a_t \in \mathcal{A}$ at time step t, transferring the system from state $st \in \mathcal{S}$ to state $s_{t+1} \in \mathcal{S}$, according to the transition model $\mathcal{T} \doteq p(s_{t+1}|s_t, a_t)$, and receiving thereafter a reward or penalization $c_t \in \mathcal{C}$. The objective of an MDP is to find the optimal policy π^* that maximizes the discounted long-term expected rewards. Formally, the optimal policy π^* provides a mapping from the current state s_t to the optimal actions, and corresponds to the strategy that maximizes

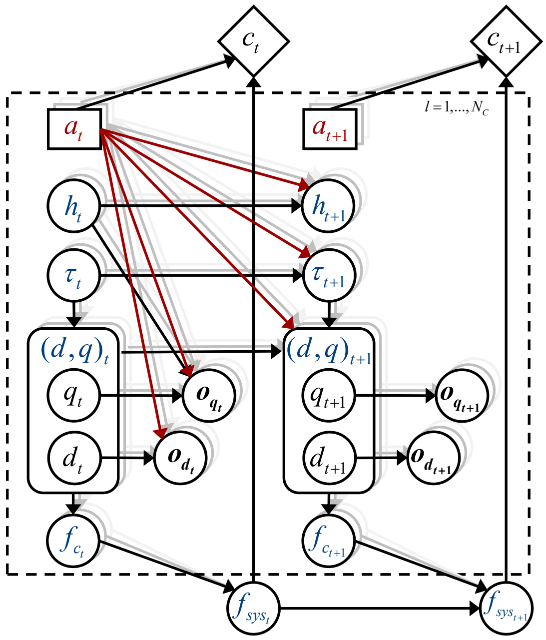


Fig. 6.1. Graphical representation of a POMDP for monitoring, inspection, and maintenance planning of structural systems. The evolution of a deterioration mechanism, parametrized at the component level by the damage d_t, an underlying random variable q_t, the deterioration rate τ_t, and sensors' health h_t, is controlled by the actions a_t and informed by the imperfect observations o_{q_t} and o_{d_t}. The system cost c_t is influenced by the taken actions a_t and the system failure state f_{sys_t}, defined with respect to the component failure state f_t of its constituent components.

the value function V^*, defined as:

$$V^*(s_t) = \max_{a_t \in \mathcal{A}} \left\{ c(s_t, a_t) + \gamma \sum_{s_{t+1} \in \mathcal{S}} p(s_{t+1}|s_t, a_t) V^*(s_{t+1}) \right\} \tag{6.1}$$

Note that if the decision-making problem is specified in terms of costs, the policy optimization consists in minimizing the total expected cost $\mathbb{E}[c_t]$. MDPs can be efficiently solved via dynamic programming by existing algorithms, e.g. value or policy iteration. In many practical applications, however, the agent cannot directly observe the environment, and observations can, instead, be planned to collect further information. For instance, non-destructive inspections provide information about the state of a component, yet the observations are also associated with measurement uncertainties. The generalization of an MDP for planning under imperfect observations corresponds to a Partially Observable Markov Decision Process (POMDP), which constitutes a principled mathematical framework for decision-making under uncertainty. It seems natural therefore to formulate the inspection and maintenance planning as a POMDP. The dynamics of a POMDP corresponds to those of an MDP, specified with the transition model $\mathcal{T} \doteq p(s_{t+1}|s_t, a_t)$, but in this case, observations $o \in \mathcal{O}$ are collected according to the observation model $\mathcal{Z} \doteq p(o_{t+1}|s_{t+1}, a_t)$. Since the states are not fully observable, the policies π are specified as a function of the belief state $\mathbf{b}(s)$, defined as the probability distribution over states $b(s) = p(s), \forall s \in \mathcal{S}$. The value function is thus formulated in terms of the belief state as:

$$V^*(\mathbf{b}_t) = \max_{a_t \in A} \left\{ \sum_{s_t \in \mathcal{S}} c(s_t, a_t) b(s_t) + \gamma \sum_{o_{t+1} \in \mathcal{O}} p(o_{t+1}|\mathbf{b}_t, a_t) V^*(\mathbf{b}_{t+1}) \right\} \tag{6.2}$$

The definition of the policy as a function of the belief state provides POMDPs with benefits with respect to other static or predefined policies, as the belief state is a sufficient statistic that captures the history of dynamically updated actions and observations [81]. We provide hereafter some recommendations for the formulation of the inspection and maintenance planning as a POMDP, with emphasis on the management of offshore wind substructures subjected to fatigue deterioration. The reader is directed to [115] for a more general and detailed overview, with instructions for formulating the inspection and maintenance planning decision-making problem as a POMDP.

If maintenance actions are not undertaken, the environment dynamics corresponds to the evolution of the deterioration process, e.g. fatigue crack growth and/or corrosion deterioration. The deterioration process can be described by various Bayesian networks, e.g. parametric or deterioration rate models, as reported in [81]. In a deterioration rate model, the damage evolution d is modeled conditional on the deterioration rate τ_t, as illustrated in 6.1. The deterioration rate τ_t can be considered as fully observable, evolving to the next deterioration rate τ_{t+1}, under normal deterioration conditions. If instead a perfect repair is performed, then the deterioration rate transitions back to the initial deterioration rate τ_{t_0}. The deterioration rate transition model $p(d_{t+1}|d_t, \tau_{t+1})$ can be estimated from simulated or experimental data, as described in [22]. As mentioned previously, inspections can be planned to collect information about the deterioration state, e.g. non-destructive evaluations or visual inspections, often described by probability of detection curves. The observation model of a POMDP can be directly specified according to probability of detection curves, i.e. $p(o_{d_t}|d_t)$.

6.2.2 POMDP formulation

The monitoring, inspection and maintenance planning decision-making problem can be formulated as a POMDP by formally defining the 7-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{Z}, \mathcal{C}, \gamma)$, in which, the transition and observation models are, in this case, specified by Bayesian networks.

States, actions, and observations

The state space $\mathcal{S}$ of a deterioration rate POMDP, illustrated in Fig. 6.1, includes the damage d, the parameter q, deterioration rate τ and component failure f states, at the component level. The deterioration can be described by alternative models, e.g. representing the deterioration process as a function of the underlying random variables [81]. The state of the sensors is explicitly represented through the h states, including also sensors deterioration rate τ_h, if sensors deterioration is stochastic. Assuming deterioration independence among the components of the system, the space grows linearly with the number of considered components N_c, including two system failure states f_{sys}. The state space dimensionality corresponds thus to $(|d||q||\tau| + |h||\tau_h|)N_c + |f_{sys}|$ states.

The action space $\mathcal{A}$ differs from traditional inspection and maintenance planning settings as monitoring decisions should also be accounted for, e.g. deciding sensor installation or replacement actions. Action and observation decisions can be combined in groups, for example, the action space of a setting featuring the possibility of one maintenance intervention, one inspection technique and one sensor might be shaped with the action-observation combinations: *do-nothing & no-inspection*, *do-nothing & inspection*, *install-sensor & no-inspection*, *install-sensor & inspection*, *repair & no-inspection*, and *repair & inspection*. Some evident sub-optimal actions may, however, be removed from the space, e.g. collecting an inspection after a perfect repair will hardly be an optimal action. Some additional action groups might be considered, depending on the scenario, e.g. *install-sensor, repair & no-inspection* could be also considered within the state space.

At the component level, the observation space $\mathcal{O}$ is defined according to the inspection and monitoring techniques available. For instance, an observation model specified by probability of detection curves includes only two observations, detection and no-detection, while an observation model which imperfectly observes the states includes the same number of observations as states. At the system level, the failure state can be assumed as fully observable in most scenarios, as a system failure, e.g. an offshore wind substructure, will immediately be announced.

Transition model

Both the evolution of the deterioration d_t, deterioration rate τ_t and sensor health h_t are tracked over time, conditional on the selected action a_t. The deterioration rate τ_t, assumed usually as fully observable, evolves to the next deterioration rate state under normal circumstances, and returns to the initial deterioration state after a perfect repair. The evolution of the deterioration rate does not depend on the damage state and can be therefore modelled independently, as $p(\tau_{t+1}|\tau_t, a_t)$. The damage d_t evolution is not only conditional on the selected action a_t, but its evolution is also conditional on the current deterioration rate τ_{t+1}, according to $p(d_{t+1}|d_t, \tau_{t+1}, a_t)$. Also, the damage state return to its initial belief if a perfect repair action is undertaken. The component failure state f_t is conditional on the damage state d_t, and the system failure state is defined as a function of the component failure state. With respect to the sensor health, its evolution is also modelled as $p(h_{t+1}|h_t, a_t)$. If an install-sensor action is assigned, the sensor health transitions to a healthy state, which evolves over time to an inoperative state, if no further replacement interventions are dictated.

Observation model

As previously mentioned, the observation model $p(o_d|d)$ depends on the inspection and/or monitoring techniques available. Non-destructive evaluation measurement quality is often measured through probability of detection curves, and in that case, the observation model can be directly described. The definition of the monitoring observation model $p(o_q|q, h_{t+1}, a_t)$ conditional on the sensor health state is the key attribute of the proposed formulation, i.e. if the sensor state is operative, then monitoring observations are naturally collected, whereas information is not collected if the health state is inoperative. Monitoring decisions are thus included within the sequential decision-making problem.

Cost model

The POMDP cost model $c(\mathbf{b_t}, a_t)$ is defined as a function of the assigned action a_t and belief state $\mathbf{b_t}$. While inspection c_{ins}, monitoring c_{sen}, repair c_{rep}, and replacement c_{repl} costs can be specified independently from the current belief state in most applications, the failure cost c_{fail} is traditionally defined according to the consequences associated with a system failure event, described by the system failure belief state $\mathbf{b}(f_{sys})$. A system failure might result, however, in instantaneous $c_{fail_{inst}}$ and perpetual consequences $c_{fail_{perp}}$. The instantaneous consequences include the economic and environmental damage associated with a failure event, and the perpetual failure cost considers all the losses associated with the inactivity of the system. In practice, the instantaneous loss is charged when the system transitions to a failure state from a survival state, whereas perpetual losses penalize a system that remained in a failure state. The system expected life-cycle cost is defined, discounted by the factor γ, as:

$$c_{tot_t} = \left\{ c_{ins_t} + c_{rep_t} + c_{sen_t} + c_{repl_t} + \mathbf{b}(f_{sys_t}) c_{fail_{inst}} + \mathbf{b}(f_{sys_t}) c_{fail_{perp}} \right\} \gamma^t \tag{6.3}$$

In some settings, the cost model might reasonably consider an additional campaign cost c_{camp}, activated if at least one component is inspected, repaired, or monitored, plus a surplus per inspected, repaired, or monitored component:

$$c_{tot_t} = \left[c_{camp_t} + \sum_{l=1}^{N_c} \left\{ c^{(l)}_{ins_t} + c^{(l)}_{rep_t} + c^{(l)}_{sen_t} \right\} + \mathbf{b}(f_{sys_t}) c_{fail} \right] \gamma^t \tag{6.4}$$

The total expected costs $\mathbb{E}[c_{tot}]$ are also computed at the system level, including campaign c_{camp}, inspection c_{ins}, monitoring c_{sen}, repair c_{rep}, failure c_{fail} and replacement c_{repl} costs:

$$\mathbb{E}[c_{tot}] = \sum_{t=0}^{t_N} \left[\gamma^t \left(c_{camp_t} + \sum_{l=1}^{N_c} \left\{ c^{(l)}_{ins_t} + c^{(l)}_{sen_t} + c^{(l)}_{rep_t} \right\} + c_{fail_t} + c_{repl_t} \right) \right] \tag{6.5}$$

6.2.3 POMDP dynamics

We introduce here the algorithmic steps for computing the POMDP dynamics, i.e. the evolution of the belief state $\mathbf{b}_t$ over time, according to the transition $\mathcal{T}$ and observation models $\mathcal{Z}$ specified as Bayesian networks. The belief update algorithm is applicable to settings in which both information from inspections and monitoring is collected. In a flat POMDP construction, i.e. the dynamics are described by a hidden Markov process, the state space $\mathcal{S}$ would have to be augmented in order to jointly consider all the involved variables. Instead, the state space is here decoupled, considering individual belief states for deterioration $\mathbf{b}_{d_t,q_t}$, deterioration rate $\mathbf{b}_{\tau_t}$, sensor health $\mathbf{b}_{h_t}$, component failure $\mathbf{b}_{f_t}$, and system failure $\mathbf{b}_{f_{sys_t}}$. Moreover, the POMDP is specified through conditional formations, alleviating thus the dimensionality and computational complexity. At the component level, the deterioration rate $\mathbf{b}_{\tau_t}$ transitions according to $p(\tau_{t+1}|\tau_t, a_t)$:

$$b(\tau_{t+1}) = \sum_{\tau_t \in \mathcal{S}_\tau} p(\tau_{t+1}|\tau_t, a_t) b(\tau_t) \tag{6.6}$$

The transition of the deterioration $\mathbf{b}_{d_t,q_t}$ is defined as:

$$\tilde{b}(d_{t+1}, q_{t+1}) = \sum_{d_t \in \mathcal{S}_d} \sum_{q \in \mathcal{S}_q} \sum_{\tau_t \in \mathcal{S}_\tau} p(d_{t+1}|d_t, \tau_{t+1}, a_t)\, b(d_t, q_t)\, b(\tau_{t+1}) \tag{6.7}$$

The sensor health $\mathbf{b}_{h_t}$ also evolves, according to the transition model $p(h_{t+1}|h_t, a_t)$,as:

$$b(h_{t+1}) = \sum_{h_t \in \mathcal{S}_h} p(h_{t+1}|h_t, a_t)\, b(h_t) \tag{6.8}$$

The failure state of the component, indicating whether the component has failed or not, is estimated by specifying the component failure state conditional on the damage state of the analyzed component $p(f_{t+1}|d_{t+1}, q_{t+1})$. Since the damage transition model inherently considers the transition from failed states, we do not introduce an additional link between component failure events at successive time steps. The component failure probability can be computed as:

$$\tilde{b}(f_{t+1}) = \sum_{d_{t+1} \in \mathcal{S}_d} \sum_{q_{t+1} \in \mathcal{S}_q} p(f_{t+1}|d_{t+1}, q_{t+1})\, \tilde{b}(d_{t+1}, q_{t+1}) \tag{6.9}$$

Since the states d_{t+1}, q_{t+1} are partially observable, the belief state can be updated based on information collected from monitoring and/or inspections:

$$b(d_{t+1}, q_{t+1}) = \frac{p(o_{d_{t+1}}, o_{q_{t+1}}|d_{t+1}, q_{t+1}, h_{t+1}, a_t)\, b(d_{t+1}, q_{t+1})\, b(h_{t+1})}{p(o_{d_{t+1}}, o_{q_{t+1}}|\mathbf{b_{d_{t+1},q_{t+1}}}, \mathbf{b_{h_{t+1}}}, a_t)} \tag{6.10}$$

At the system level, the failure state is assumed to be fully-observable, as explained previously. During the estimation stage, the system failure $\mathbf{b}(f_{sys_{t+1}})$ is defined conditional on the component failure $\mathbf{b}(\mathbf{f_{t+1}})$ of N_C components, and the system failure state at the previous time step $\mathbf{b}(f_{sys_t})$:

$$b(f_{sys_{t+1}}) = \sum_{f_{sys_{t+1}} \in \mathcal{S}_{f_{sys}}} \sum_{l \in N_C} p(f_{sys_{t+1}}|f^{(l)}_{t+1}, f_{sys_{t+1}}) b(\tilde{f}^{(l)}_{t+1}) b(f_{sys_t}) \tag{6.11}$$

In practice, the system failure state is fully observed, and both component failure state and component damage state can be inferred conditional on the observed system failure state $\hat{f}_{sys_{t+1}}$. The component failure states conditional on the observed system failure state is described as:

$$b(\mathbf{f_{t+1}}|\hat{f}_{sys_{t+1}}) = \frac{p(\hat{f}_{sys_{t+1}}|\mathbf{f_{t+1}})\, \tilde{\mathbf{b}}(\mathbf{f_{t+1}})}{b(\hat{f}_{sys_{t+1}})} \tag{6.12}$$

The component failure belief $\mathbf{b}(f^{(l)}_{t+1})$ for a component (l) results from marginalizing out all the components other than (l):

$$b(f^{(l)}_{t+1}|\hat{f}_{sys_{t+1}}) = \sum_{\sim l \in N_c} b(f^{(l)}_{t+1}|\hat{f}_{sys_{t+1}}) \tag{6.13}$$

Moreover, the deterioration belief of each component is also inferred conditional on the updated component failure belief as:

$$b(d_{t+1}, q_{t+1}|f_{t+1}) = \sum_{f_{t+1} \in \mathcal{S}_{\{}} \frac{p(f_{t+1}|d_{t+1}, q_{t+1})\, b(d_{t+1}, q_{t+1})\, b(f_{t+1})}{\tilde{b}(f_{t+1})} \tag{6.14}$$

All the algorithmic steps for implementing the belief update are summarized in Algorithm 4.

Algorithm 4 Belief update for a system of N_c components

function UPDATEBELIEF($\mathbf{b}_{d_t,q_t}, \mathbf{b}_{\tau_t}, \mathbf{b}_{h_t}, \mathbf{b}_{f_{sys_t}}, \mathbf{a}_t, \mathbf{o}_{d_t}, \mathbf{o}_{q_t}$)
  for $l = 1, N_c$ **do**
    $b\left(\tau_{t+1}^{(l)}\right) \leftarrow p\left(\tau_{t+1}^{(l)}|\tau_t^{(l)}, a_t^{(l)}\right)\, b\left(\tau_t^{(l)}\right)$ ▷ propagation step
    $\tilde{b}\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right) \leftarrow p\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}|d_t^{(l)}, q_t^{(l)}, \tau_{t+1}^{(l)}, a_t^{(l)}\right)\, b\left(d_t^{(l)}, q_t^{(l)}\right)\, b\left(\tau_{t+1}^{(l)}\right)$
    $b\left(h_{t+1}^{(l)}\right) \leftarrow p\left(h_{t+1}^{(l)}|h_t^{(l)}, a_t^{(l)}\right)\, b\left(h_t^{(l)}\right)$
    $\tilde{b}\left(f_{t+1}^{(l)}\right) \leftarrow p\left(f_{t+1}^{(l)}|d_{t+1}^{(l)}, q_{t+1}^{(l)}\right)\, \tilde{b}\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right)$ ▷ component failure probability
    $b\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right) \leftarrow p\left(o_{d_{t+1}}^{(l)}, o_{q_{t+1}}^{(l)}|d_{t+1}^{(l)}, q_{t+1}^{(l)}, h_{t+1}^{(l)}, a_t^{(l)}\right)\, \tilde{b}\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right)\, b\left(h_{t+1}^{(l)}\right)$
    $b\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right) \leftarrow b\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right) / \left\{p\left(o_{d_{t+1}}^{(l)}, o_{q_{t+1}}^{(l)}|\mathbf{b}_{d_{t+1},q_{t+1}}^{(l)}, \mathbf{b}_{h_{t+1}}^{(l)}, a_t^{(l)}\right)\right\}$
  end for
  $b\left(f_{sys_{t+1}}\right) \leftarrow p\left(f_{sys_{t+1}}|\mathbf{f_{t+1}^{(l)}}, f_{sys_t}\right)\, \tilde{\mathbf{b}}\left(\mathbf{f_{t+1}^{(l)}}\right) b\left(f_{sys_t}\right)$ ▷ system failure probability
  $\hat{f}_{sys_{t+1}} \sim b(f_{sys_{t+1}})$ ▷ system failure state
  $\mathbf{b}\left(\mathbf{f_{t+1}^{(l)}}\right) \leftarrow p\left(\hat{f}_{sys_{t+1}}|\mathbf{f_{t+1}^{(l)}}\right) \tilde{\mathbf{b}}\left(\mathbf{f_{t+1}^{(l)}}\right) / b\left(\hat{f}_{sys_{t+1}}\right)$ ▷ component failure update
  for $l = 1, N_c$ **do** ▷ component belief update
    $b\left(f_{t+1}^{(l)}\right) \leftarrow \sum_{\sim l} \mathbf{b}\left(\mathbf{f_{t+1}^{(l)}}\right)$
    $b\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right) \leftarrow \sum_{f^{(l)}} \left\{p\left(f_{t+1}^{(l)}|d_{t+1}^{(l)}, q_{t+1}^{(l)}\right)\, b\left(d_{t+1}^{(l)}, q_{t+1}^{(l)}\right)\, b\left(f_{t+1}^{(l)}\right) / \tilde{b}\left(f_{t+1}^{(l)}\right)\right\}$
  end for
  return $\mathbf{b}_{d_{t+1},q_{t+1}}, \mathbf{b}_{\tau_{t+1}}, \mathbf{b}_{h_{t+1}}, \mathbf{b}_{f_{sys_{t+1}}}$
end function

6.3 Optimal management of offshore wind structures via deep reinforcement learning

Discovering optimal management policies for structural systems, typically featuring high dimensional state spaces, is a computationally challenging task [115]. In such settings, the management strategy can be instead approximated by an actor network, who parametrizes the policy with artifical neural networks. The actor network can be additionally guided by a critic network, who provides an estimate of the value function, also approximated by neural artificial networks. This deep decentralized multi-agent actor-critic (DDMAC) approach, originally proposed in [28], is capable of handling high-dimensional settings, as already demonstrated in [28, 29, 115]. As illustrated in Fig. 6.2, each accessible component of the structural system is controlled by the stochastic policy $\pi(a|\mathbf{b}, \boldsymbol{\theta}^{\pi})$, defined as a function of the belief state $\mathbf{b}$ and provided by a group of multi-agent actor networks. Within the actor network, each accessible component n_{C_a} is represented by an agent who is acting as an independent unit:

$$\pi(\mathbf{a}, \mathbf{b}) = \prod_{l=1}^{n_{C_a}} \pi_l(a^{(l)}|\mathbf{b}) \tag{6.15}$$

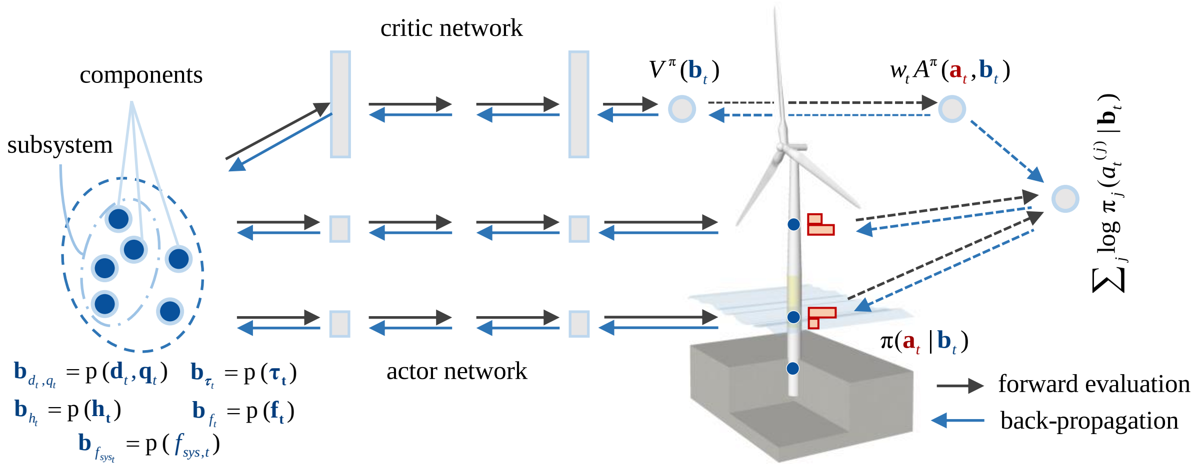


Fig. 6.2. Illustration of the proposed POMDP-based deep reinforcement learning approach for the optimal management of offshore wind substructures. An actor network generates the policy $\pi(\mathbf{a}_t|\mathbf{b}_t)$ as a function of a the dynamically updated system belief states $\mathbf{b}_t$. During the training stage, the weights of the actor network are adjusted according to the collected system costs, guided also by a critic network, who provides an estimate of the value function $V^{\pi}(\mathbf{b}_t)$.

In practice, acting as independent units means that the actions executed by one agent do not influence directly the state of other agents. Note that inaccessible components, i.e. those components in which actions or observations are not possible, might not be modeled by agents, yet their belief state can still be communicated as input to the actor network, providing valuable information about the overall system condition. The POMDPs formulation introduced in Section 6.2 is here integrated with a DDMAC multi-agent approach, casting an efficient algorithmic platform for optimal monitoring, inspection, and maintenance of structural systems. The algorithmic scheme is summarized in Algorithm 5 and applied to the management of offshore wind substructures in Section 6.4.

The actor networks receive as input the belief state of each component's deterioration $\mathbf{b}(d_t)$, underlying random variables $\mathbf{b}(q_t)$, deterioration rate $\mathbf{b}(\tau_t)$, sensor health $\mathbf{b}(h_t)$ component failure states $\mathbf{b}(f_t)$, and system failure state $\mathbf{b}(f_{sys_t})$, as shown in Fig. 6.2. The fully-connected hidden layers of the actor networks might be activated through ReLu functions, whereas the ouput layer is activated by a softmax function, delivering the output policy as a probability distribution over the available actions, as graphically depicted by red bars in Fig. 6.2. The noisy system costs, along with the related belief states, collected through simulations are stored in a replay buffer [95], from which a batch of experiences is stochastically sampled to adjust the actor networks weights. Such offline training scheme results more efficient than online training approaches.

Algorithm 5 Deep Decentralized Multi-agent Actor Critic (DDMAC)

Initialize replay buffer
Initialize actor and critic network weights $\boldsymbol{\theta}^{\pi}, \boldsymbol{\theta}^{V}$
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 $\mathbf{a}_t \sim \mu_t = \{\pi_j(\cdot|\mathbf{b}_t, \boldsymbol{\theta}^{\pi})\}_{j=1}^{n}$
    Collect reward $r(\mathbf{b}_t, a_t)$
    Observe $o_{t+1}^{(l)} \sim p(o_{t+1}^{(l)}|\mathbf{b}_t, \mathbf{a}_t)$ for $l = 1, 2, ..., m$
    Compute beliefs $\mathbf{b}_{t+1}^{(l)}$: $\textsc{updateBelief}(\mathbf{b}_t^{(l)}, \mathbf{a}_t, \mathbf{o}_t^{(l)})$
    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 $A_i = r(\mathbf{b}_i, \mathbf{a}_i) - V(\mathbf{b}_i, \boldsymbol{\theta}^{V})$
    Otherwise $A_i = r(\mathbf{b}_i, \mathbf{a}_i) + \gamma V(\mathbf{b}_{i+1}, \boldsymbol{\theta}^{V}) - V(\mathbf{b}_i, \boldsymbol{\theta}^{V})$
    Update actor parameters $\boldsymbol{\theta}^{\pi}$ according to gradient:
      $\mathbf{g}_{\boldsymbol{\theta}^{\pi}} \simeq \sum_i w_i \{\sum_{j=1}^{n} \nabla_{\theta^{\pi}} \, log \, \pi_j(a_i^{(j)}|\mathbf{b}_i, \boldsymbol{\theta}^{\pi})\} A_i$
    Update critic parameters $\boldsymbol{\theta}^{V}$ according to gradient:
      $\mathbf{g}_{\boldsymbol{\theta}^{V}} \simeq \sum_i w_i \nabla_{\theta^{V}} V^{\pi}(\mathbf{b}_i|\boldsymbol{\theta}^{V}) A_i$
  end for
end for

Specifically, the off-policy gradient estimator is specified with samples retrieved from a behavior policy μ, instead of collecting them directly from π, and rectified with the truncated importance sampling weight $w_t = min\{c, \pi(a_t|b_t)/\mu(a_t|b_t)\}$, with $c > 0$ [28]:

$$\mathbf{g}_{\boldsymbol{\theta}^{\pi}} = \mathbb{E}_{\mathbf{a_t} \sim \mu} \left[w_t \left\{ \sum_{i=1}^{n} \nabla_{\theta^{\pi}} \, log \, \pi_i(a_t^{(i)}|\mathbf{b}_t, \boldsymbol{\theta}^{\pi}) \right\} A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^{V}) \right] \tag{6.16}$$

The optimality of the sampled action a_t, assessed with respect to the current estimated value function $V(\mathbf{b}_t)$, is formulated in a temporal difference learning scheme through the advantage function $A^{\pi}(\mathbf{b}_t, \mathbf{a}_t)$:

$$A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^{V}) \simeq r(\mathbf{b}_t, \mathbf{a}_t) + \gamma V(\mathbf{b}_{t+1}|\boldsymbol{\theta}^{V}) - V(\mathbf{b}_t|\boldsymbol{\theta}^{V}) \tag{6.17}$$

Moreover, the value function is approximated by a critic network, as shown in Fig. 6.2, receiving the same input as the actor networks, yet generating as output a scalar estimate of the total system cost that can be expected. The value function estimate is integrated into the formulation of the advantage function $A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^{V})$, and the critic network therefore acts as a judge of the action selected by the actor networks.

The weights $\boldsymbol{\theta^V}$ of the critic network are also adjusted through a temporal difference learning approach according to the gradient:

$$\mathbf{g}_{\boldsymbol{\theta}^V} = \mathbb{E}_{\mathbf{a_t}\sim\mu}\Big[w_t \nabla_{\boldsymbol{\theta}^V} V^{\pi}(\mathbf{b}_t|\boldsymbol{\theta}^V) A^{\pi}(\mathbf{b}_t, \mathbf{a}_t|\boldsymbol{\theta}^V)\Big] \tag{6.18}$$

6.4 Numerical experiments

The proposed algorithmic scheme is implemented and tested here, devising management strategies for monitoring, inspection, and maintenance planning of offshore wind structural systems under fatigue deterioration. The decision-making problem is firstly formulated as a POMDP, according to the instructions provided in Section 6.2, and the strategies are identified by DDMAC, as explained in Section 6.3, approximating both policies and value functions by artificial neural networks. The resulting DDMAC policies are thoroughly compared against both common and state-of-the-art strategies over 5,000 policy evaluations. Specifically, DDMAC policies are compared, in terms of total expected costs, against the following strategies:

i) Corrective (CORR): a replacement action is immediately undertaken after the occurrence of a system failure event. No parameters are optimized in this case.

ii) Calendar-based (CAL): the fatigue hotspot located in the atmospheric zone is inspected at years 7 and 14, whereas the hotspot located in the splash zone is only inspected at year 13. These inspection decision rules are inspired by the offshore wind design recommendations [116]. In this case, repair interventions are dictated after a crack detection is observed.

iii) Inspection and maintenance planning heuristic decision rules (HEUR-INS): this scheme is adapted from the heuristic decision rules proposed in [6]. Inspections are planned at equidistant intervals, and at each inspection campaign, the number of components inspected are those registering higher hotspot failure probability. Repairs are also undertaken, in this case, after a crack detection is observed. The optimized heuristics are the interval between inspections and the number of components to be inspected.

iv) Monitoring and maintenance planning heuristic decision rules (HEUR-SEN): the previous inspection scheme is readjusted here by planning sensor installation actions, instead of inspections. Sensors are thus installed at equidistant intervals and repairs are executed if the collected monitoring observation exceeds a predefined observation value. The optimized heuristics are the interval between sensor installation actions, the number of components to be monitored, and the monitoring observation threshold that activates repairs.

v) Inspection, monitoring and maintenance planning heuristic decision rules (HEUR-SEN&INS): monitoring, inspection, and maintenance decision rules are combined in this scheme. Sensors are installed at equidistant intervals, inspections are planned if a predefined monitoring observation threshold is exceeded, and repairs are undertaken after a crack detection is observed. The optimized heuristics include, therefore, the interval between sensor installation actions, the number of components to be monitored, and the monitoring observation threshold that activates inspections.

6.4.1 Fatigue deterioration environment

Each offshore wind substructure contains, in the conducted numerical experiments, three representative structural details characterized with different fatigue deterioration, inspection quality, and cost model, depending on their location. All the analyzed structural details are butt welds that connect pipe segments of a tower-monopile substructure inspired by NREL 5MW offshore wind turbine [117]. Considering the fatigue details to be representative butt welds, the substructure failure event occurs if any of the hotspots fails, thus constituting a series system. The specific fatigue deterioration attributes of each component are illustrated in Fig. 6.3 and listed below:

- Hotspot above the water level (AW): This structural detail is located in the tower, above the waterline, and with a plate thickness of 20 mm. Due to the ease of accessibility, a fatigue design factor (FDF) of 1 is assigned, according to industrial design standards [116], and since the hotspot is in the atmospheric zone, the reference SN curves are those in air environmental conditions.

- Hotspot below the water level (BW): This joint is located in the splash zone, below the waterline, and with a plate thickness of 60 mm. In this case, inspections require remotely operated vehicles or skilled divers, and according to industrial design standards [116], a fatigue design factor (FDF) of 2 is assigned. In the splash zone, the corrosive environment accelerates fatigue deterioration, thus the SN reference curves in a corrosion environment with cathodic protection are selected for the fatigue damage computation.

- Hotspot below the mudline (MD): This connection is located in the monopile, below the mudline, and with a plate thickness of 60 mm. Inspections and repairs are not possible and according to fatigue design guidelines a fatigue design factor (FDF) of 3 is assigned [116]. The fatigue damage is estimated with basis on a seawater environment with cathodic protection SN curve.

Table 6.1. Random variables and deterministic parameters for modeling the fatigue deterioration.

Parameter	Distribution	Mean	Std
λ	Deterministic	0.8	-
v (cycles/s)	Deterministic	0.16	-
m_1	Deterministic	3	-
m_2	Deterministic	5	-
Δ	Lognormal	1	0.3
d_0 (mm)	Exponential	0.11	-
Y	Lognormal	1	0.1
m_{FM} (mm)	Deterministic	3	-
AW hotspot			
$C_{1,SN}$*	Normal	12.164	0.2
$C_{2,SN}$*	Normal	15.606	0.2
q (MPa)	Trunc. Normal	10.21	2.55
lnC_{FM}	Normal	-26.445	0.122
d_c (mm)	Deterministic	20	-
BW hotspot			
$C_{1,SN}$*	Normal	11.764	0.2
$C_{2,SN}$*	Normal	15.606	0.2
q (MPa)	Trunc. Normal	7.40	1.85
lnC_{FM}	Normal	-26.043	0.403
d_c (mm)	Deterministic	60	-
MD hotspot			
$C_{1,SN}$*	Normal	11.764	0.2
$C_{2,SN}$*	Normal	15.606	0.2
q (MPa)	Trunc. Normal	6.74	1.68
lnC_{FM}	Normal	-26.122	0.396
d_c (mm)	Deterministic	60	-

*Fully correlated.

In general, a fracture mechanics model can be already utilized at the design stage to assess the fatigue resistance of each structural detail. However, SN curves and Miner's cumulative damage law are usually followed at the design stage of offshore wind structures due to its simplicity. Since fatigue damage is not physically observable, a prior modeling of the fatigue deterioration cannot be combined with inspections within a Bayesian approach.

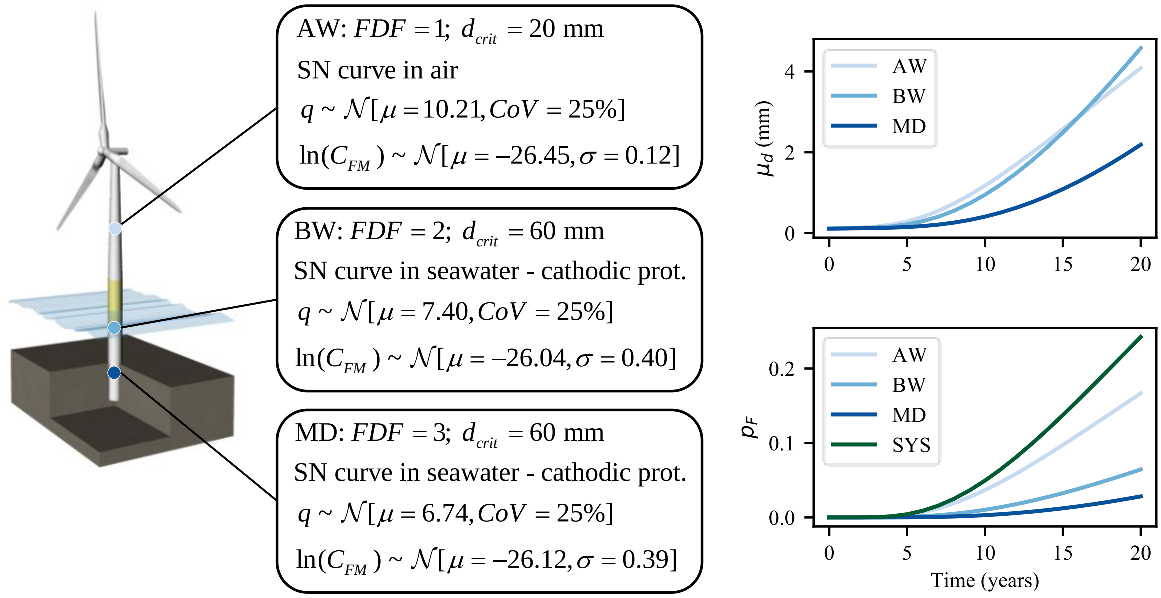


Fig. 6.3. (Left) Fatigue deterioration attributes of each analyzed structural detail, above the waterline (AW), below the waterline (BW), and below the mudline (MD). Each connection is assigned with a fatigue design factor (FDF), SN curve type, stress range scale parameter q, and crack growth parameter C_{FM}. (Right) Evolution of the expected crack size μ_d, component failure probability p_F, and system failure probability (SYS) over time.

Instead, a fracture mechanics model can be calibrated with respect to the structural reliability found with a probabilistic SN-Miner's cumulative damage model, enabling therefore Bayesian updating of crack observations collected at inspections. Note that if a fracture mechanics model is already developed at the design stage, then it can be directly employed for inspection and maintenance planning. In this case and considering the widely usage of Miner's model during the design stage, we calibrate a fracture mechanics model for each analyzed welded joint. All the parameters required for estimating the fatigue deterioration, in terms of both Miner's rule and fracture mechanics, are listed in Table 6.1.

The Miner's rule limit state is formulated by subtracting the cumulative fatigue damage to the damage failure parameter Δ, corrected by the fatigue design factor (FDF). The long-term stress range is assumed to be described by a Weibull distribution with scale parameter q and shape parameter λ, whereas the fatigue resistance is empirically parameterized by bi-linear SN curves, with slopes m_1 and m_2, along with the corresponding SN curve intercepts $C_{1,SN}$ and $C_{2,SN}$. Considering a cycle rate v of 0.16 over a horizon t of 20 years, the fatigue damage limit state g_{SN} can be constructed as a function of the time step as:

$$g_{SN}(t) = \Delta - vt\left[\frac{q^{m_1}}{C_{1,SN}}\gamma_1\left\{1 + \frac{m_1}{\lambda}; \left(\frac{S_1}{q}\right)^{\lambda}\right\} + \frac{q^{m_2}}{C_{2,SN}}\gamma_2\left\{1 + \frac{m_2}{\lambda}; \left(\frac{S_1}{q}\right)^{\lambda}\right\}\right] \tag{6.19}$$

Note that γ_1 and γ_2 correspond to incomplete gamma functions arising from the bi-linear SN curves. With basis on the limit state, the failure probability can be then computed by structural reliability methods [64] or crude Monte Carlo simulations as the event described by $p_F(t) = p[g_{SN}(t) \leq 0]$. The retrieved failure probability over time, or structural reliability $\beta(t) = -\Phi^{-1}[p_F(t)]$, constitute the reference for calibrating the fracture mechanics model. In this case, the crack growth is modeled through a Paris' law model, originally proposed by Ditlevsen [63]. In this Markovian model, the crack size d_{t+1} can be computed as a function of the crack size at the previous time step d_t, Paris' law parameters C_{FM} and m, equivalent stress range S_e and number of cycles in one time step:

$$d_{t+1} = \left[d_t^{\frac{2-m}{2}} + \frac{2-m}{2} C_{FM} (Y\pi^{0.5} S_e)^m n) \right]^{\frac{2}{2-m}} \tag{6.20}$$

Assuming that each hotspot fails when the crack grows further than its plate thickness [91], the fracture mechanics limit state can be formulated as $g_{FM}(t) = d_c - d_t$. The crack growth parameters C_{FM} are then calibrated with the objective of minimizing the difference between the failure probabilities estimated by Miner's and fracture mechanics limit states. The calibration is conducted by least-square optimization and the resulting parameters are listed in Table 6.1 and Fig. 6.3. The equivalent stress range S_e is often considered time-invariant, as the mean of the stress range described by a two-parameter Weibull distribution:

$$S_e = q\Gamma(1 + 1/\lambda) \tag{6.21}$$

In this case, we additionally include a Gaussian noise to the temporal evolution of the scale parameter q, representing the potential variation of offshore wind turbine dynamics due to scouring, rotor imbalance, or other factors. At each time step t, the scale parameter is then influenced by a Gaussian noise ϵ_q with a 4% coefficient of variation, and the fatigue growth model is reformulated as:

$$d_{t+1} = \left[d_t^{\frac{2-m}{2}} + \frac{2-m}{2} C_{FM} \{Y\pi^{0.5} q\epsilon_q \Gamma(1 + 1/\lambda)\}^m n) \right]^{\frac{2}{2-m}} \tag{6.22}$$

As already anticipated in Section 6.2, the continuous distributions are discretized in order to formulate the inspection and maintenance decision problem as a factored POMDP, encoded by discrete dynamic Bayesian networks. The discretization scheme is presented in Table 6.2, listing the intervals and state space of the crack size d, deterioration rate τ and scale factor q.

Table 6.2. Description of the discretization scheme for the crack size d, stress range scale parameter q, deterioration rate τ, sensor condition h, component failure state f, and system failure state f_{sys}.

Variable	Interval boundaries	States
d	$[0, d_0 : (d_c - d_0)/(\lvert S_d\rvert - 2) : d_c, \infty]$	60
q	$[0, 0.11 : 19.89/(\lvert S_q\rvert - 2) : 20, \infty]$	30
τ	$[0 : 1 : 21]$	21
h	$[0 : 1 : 3]$	3
f	[survival, failure]	2
f_{sys}	[survival, failure]	2

6.4.2 Optimal monitoring, inspection, and maintenance planning of offshore wind substructures

The objective of these numerical experiments is the conception of optimal monitoring, inspection, and maintenance strategies for offshore wind substructures under fatigue deterioration described according to the fracture mechanics model introduced in Section 6.4.1. In terms of dimensionality, the state space at the hotspot level features 37,805 states, with 60 crack size states d, 30 q states and 21 deterioration rate states, 3 sensor states, and 2 component failure f states, augmenting to 113,417 states at the offshore wind turbine level, in which the two additional states indicate the failure state of the system.

By formulating the deterioration process through conditional formations, the state space increases linearly with the number of considered hotspots, in contrast to a flat representation, in which the state space would otherwise increase exponentially, resulting in a total space of $\approx 5 \cdot 10^{13}$ states. Note that the health of the sensors is also tracked by three fully-observable states per sensor, with the first two states indicating an active sensor operation, whereas the last state indicates an inactive sensor. In this application, the sensors are assumed to be operative for two years without further maintenance, providing information about the stress range scale parameter q.

In terms of the observation model, the measurement uncertainty associated with inspections is quantified by probability of detection curves, available in offshore wind design standards [79]. From the three studied hotspots, only the structural details above the water line and at the splash zone can be inspected, yet the accessibility of the latter is more complex as divers or remotely operated vehicles might be required.

Eddy current non-destructive inspections are selected as the available inspection technique and the probability of detection curves, for each hotpot, are formulated as:

$$p(o_{d_t}|d_t) = 1 - \frac{1}{1 + (d_t/\chi)^b}, \tag{6.23}$$

where the parameters χ and b are 0.4 and 1.43, respectively, for the AW hotspot, and 1.16 and 0.90, for the hotspot located at the splash zone [79]. Besides crack inspections, strain can also be monitored through operational sensors, providing in turn stress range information q. If the sensors were able to perfectly measure q, then the state of q would be perfectly observed. Yet the existing measurement uncertainty, along with the fact that sensors are not directly measuring the strain at the precise fatigue hotspot location, is accounted by including an unbiased Gaussian noise, characterized with a 15% coefficient of variation with respect to the initial q. In total, there are thus 60 observations available, i.e. the joint of 30 loading observations o_q and 2 crack observations o_d.

The decision-maker disposes, in this setting, of six available actions for each accessible hotspot: (i) do-nothing & no-inspection, (ii) do-nothing & inspection, (iii) install-sensor & no-inspection, (iv) install-sensor & inspection, (v) repair & install-sensor, and (vi) repair. Actions (i), (ii) and (vi) correspond to the same I&M decision problem as the one formulated in Section 6.4.2, whereas actions (ii), (iii) and (v) are directly related to monitoring decisions. Installing or replacing a sensor (action iii) transfers the sensor state from inactive to operative, and this operation can also be conducted at the same time as inspections are collected (action iv), or while undertaking repairs (action vi). Moreover, a system replacement action is automatically planned after the occurrence of an offshore wind substructure failure, and in that case, all the fatigue hotspots return to their intact condition.

DDMAC's architecture contains, for each actor, two hidden fully-connected layers of 100 neurons activated by ReLu functions, along with a critic network featuring two hidden fully-connected layers of 300 neurons, and a softmax function provides output probabilities for the six available actions. The input to both actor and critic networks includes the crack size, stress range scale parameter, deterioration rate, sensor condition, component failure, and system failure belief states. The gradients of the networks are adjusted with respect to the overall system costs, including failure, replacement, inspection, repairs and monitoring costs. A system failure event is associated with a consequence c_f of 600 monetary units, representing capital losses as well as environmental consequences, and the replacement of an offshore wind turbine c_{replac} costs 350 monetary units. Inspection c_i are charged for the AW and BW hotpots with 1 and 4 monetary units, respectively; and the repairs c_{rep} cost 10 and 30 monetary units, respectively. Installing a sensor is assumed to cost for the AW and BW hotspots, 2 and 6 monetary units, respectively.

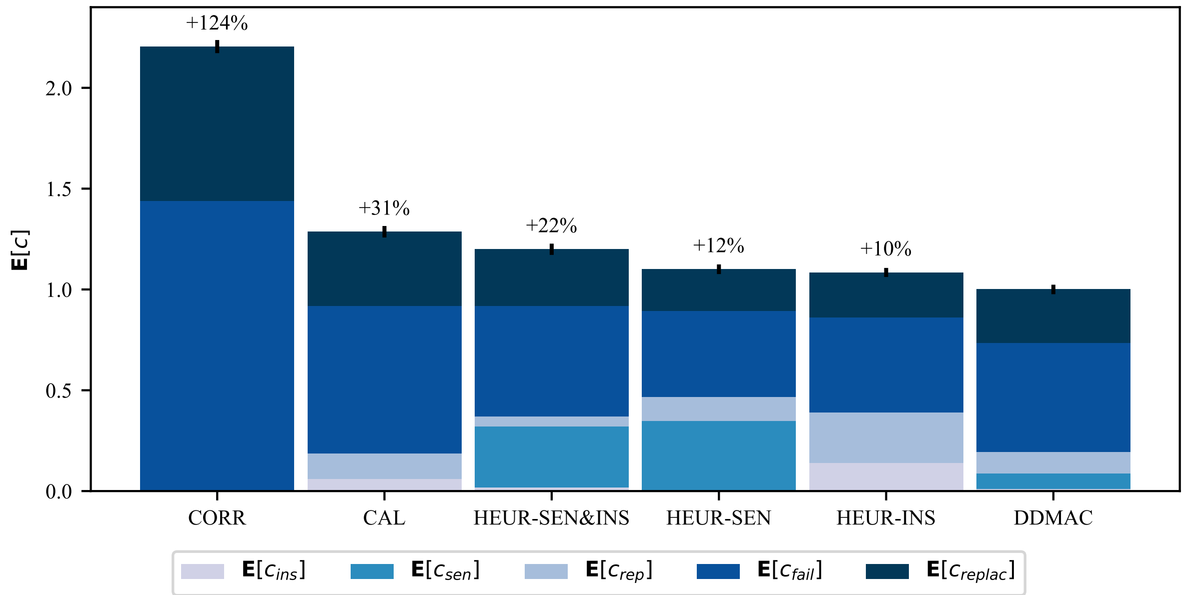


Fig. 6.4. Expected cost results of the numerical experiments conducted for the management of offshore wind substructures, divided into, inspection $\mathbf{E}[c_{ins}]$, sensor installation/replacement $\mathbf{E}[c_{sen}]$, perfect-repair $\mathbf{E}[c_{rep}]$, failure $\mathbf{E}[r_{fail}]$, and replacement $\mathbf{E}[r_{replac}]$ expected costs. The bar chart compares the resulting expected cost from DDMAC, corrective, calendar-based, and various heuristic strategies.

The expected total costs of corrective, calendar-based, risk-based heuristics and DDMAC policies are showcased in Fig. 6.4, highlighting the individual contribution of inspection $\mathbb{E}[c_{ins}]$, sensor installation $\mathbb{E}[c_{sen}]$, repair $\mathbb{E}[c_{rep}]$, failure $\mathbb{E}[c_{fail}]$, and replacement $\mathbb{E}[c_{replac}]$ to the total expected costs $\mathbb{E}[c_{tot}]$. The stacked bars shown in Fig. 6.4 indicate the total expected costs and 95% confidence intervals for each considered strategy over 5,000 policy evaluations. Note that the expected costs are normalized with respect to the result obtained by DDMAC, thereby enabling a direct comparison.

The results reveal that DDMAC policies outperform all the other tested strategies, with cost savings ranging from 124%, for the case of corrective policies, to 10% corresponding to heuristic-based strategies. A corrective maintenance policy leads to high failure and replacement costs due to the lack of maintenance control, whereas the calendar-based approach, inspired by design standards, reduces the failure risk by conducting periodic inspections, followed by repair actions if defects are found. The calendar-based strategy results, however, in higher expected failure and replacement costs than heuristics and DDMAC strategies. Heuristic-based policies provide lower expected costs than the calendar-based scheme, due to the optimization conducted for the selection of the predefined decision rules.

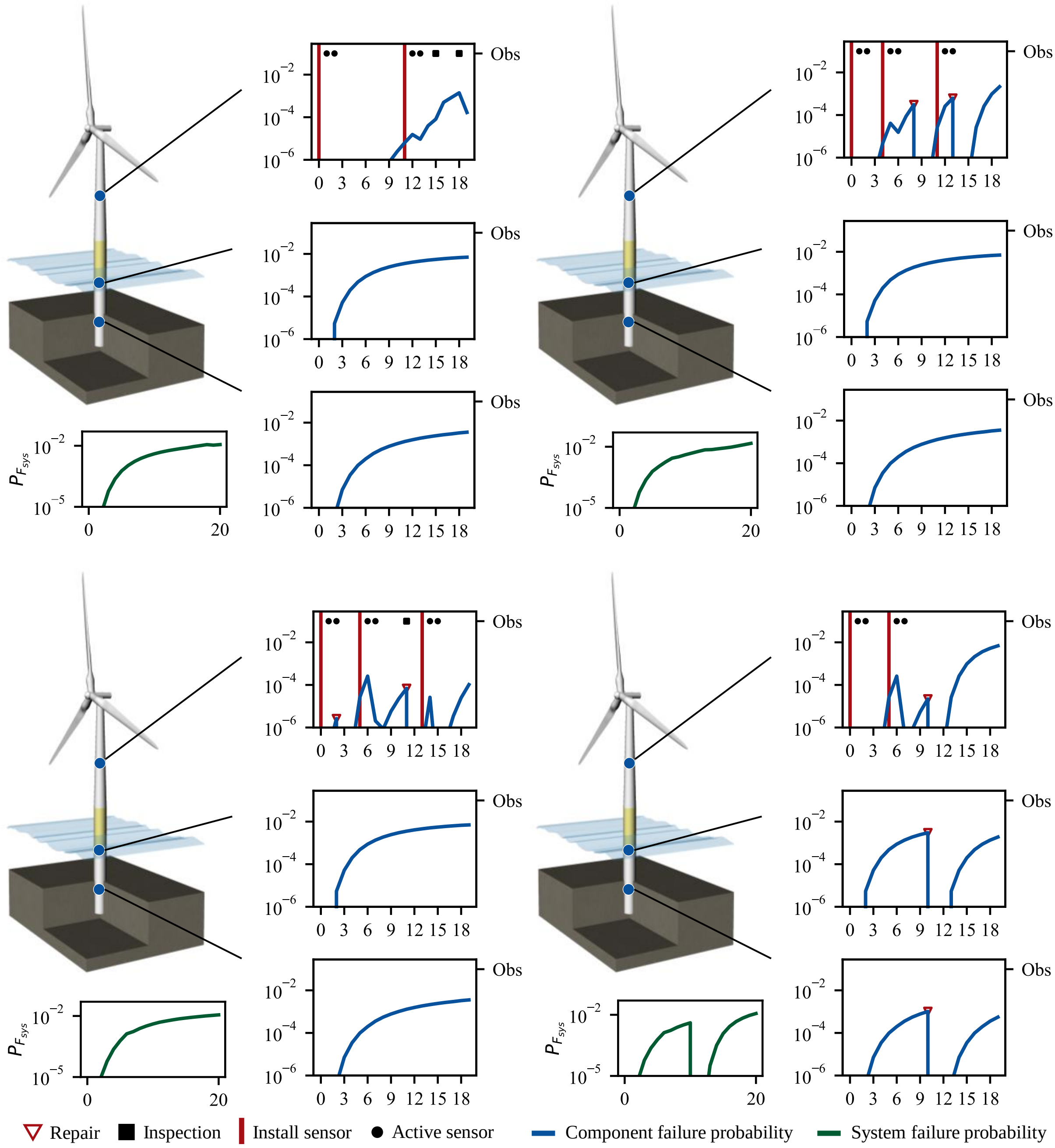


Fig. 6.5. DDMAC policy realizations illustrating the management of an offshore wind substructure subjected to fatigue deterioration. Hotspot and substructure failure probabilities over time are represented in the diagrams with blue and green lines, respectively. Within the component diagrams, maintenance and observation decisions are also depicted with markers, and vertical red lines indicate the installation or replacement of a sensor.

With respect to the total expected costs, inspection and monitoring heuristics yield very similar results, differing on whether inspections or monitoring observations are collected to dictate repair decisions. In contrast, heuristic policies defined by both monitoring and inspection decision rules result less optimal than DDMAC and other heuristics, demonstrating the complexity of defining a set of decision rules capable of combining monitoring, inspection and maintenance decisions, within an immense available policy space.

Within the high-dimensional, observation and action state space characteristic of this monitoring, inspection and maintenance decision problem, DDMAC policies effectively combine monitoring, inspection and repair decisions, yielding decision sequences which otherwise might be difficult to predefined based on engineering judgement. To understand how DDMAC combines monitoring and inspection decisions, Fig. 6.5 illustrates four policy realizations, depicting the failure probability and the status of the sensor over time for each hotspot, indicating inspection, sensor installation and repair actions, as well as the failure probability of each wind structural system. In general, DDMAC policies concentrates on controlling the failure risk of the hotspot AW, located at the atmospheric zone, since it is the weakest link of a series system and the inspection and repair actions are cheaper than at other hotspots. In particular, inspections are sometimes planned after a monitoring campaign, as shown at the upper-left corner of Fig. 6.5 whereas inspections and sensors can be also planned concurrently, as displayed at the lower-left corner of Fig. 6.5. Also, monitoring and repairs might be combined within one episode, as illustrated at the upper-right corner of Fig. 6.5, without planning any inspections. A policy realization registering a system failure is also shown at the lower-right corner of Fig. 6.5 for illustration purposes, indicating that the hotspot located below the mudline can only be repaired if a replacement action is undertaken.

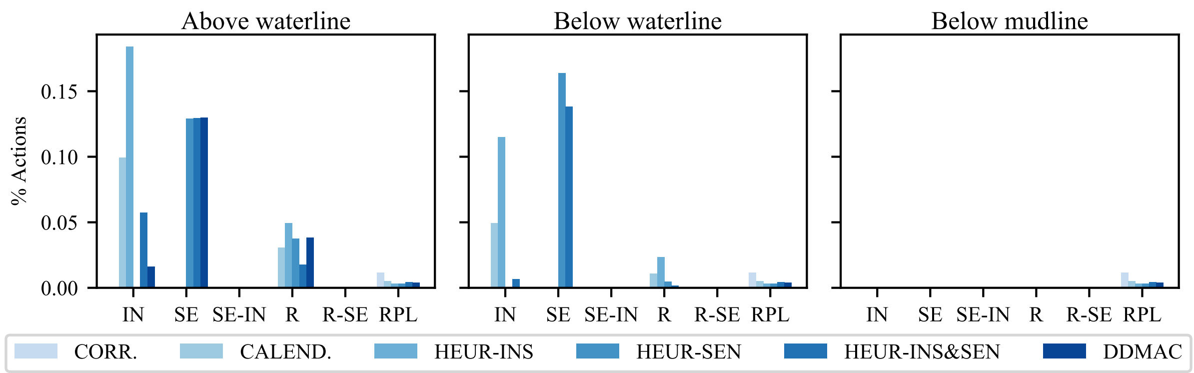


Fig. 6.6. Histogram of the actions assigned by DDMAC, corrective (CORR.), calendar (CALEND.), and heuristic (HEUR) based strategies over 5,000 hotspot policy realizations. The percentage of inspections (IN), sensor installations (SE), sensor installations & inspections (SE-IN), repairs (R), repairs & sensor installations (R-SE), and replacements (RPL) are represented by vertical bars for each hotspot – above the waterline, below the waterline, and below the mudline –.

Finally, the actions experienced, over 5,000 policy evaluations, by all the tested strategies are represented through a histogram in Fig. 6.6. The percentage of experienced inspection (IN), sensor installation (SE), sensor installation and inspection (SE-IN), repair (R), repair and sensor installation (R-SE) and replacement (RPL) actions, for each offshore wind subtructure hotspot, are represented by vertical bars. Note that do-nothing actions are not represented in Fig. 6.6 as they predominate over the other actions, complicating a close examination of results. The histogram shows that DDMAC policies concentrate, mainly, on the uppermost hotspot (AW), as already observed through the policy realizations displayed in Fig. 6.5, whereas the risk-based heuristics often plan inspection, monitoring or repair actions on the hotpot located below the water line (BW). This can be explained by the definition of the heuristic decision rules, in which the inspected, or monitored, components corresponds to those with higher failure probability, disregarding the potential different inspection, monitoring or repair cost between the structural components. The cost of interventions, in this setting, varies significantly for the different hotspots examined, as underwater actions are more expensive that interventions undertaken above the water level. Also, the action histogram reveals the balance between inspection and monitoring actions provided by DDMAC policies.

6.4.3 Optimal management of an offshore wind farm

In this setting, we devise strategies for the management of 55 monopile-type substructures inspired on Belwind offshore wind farm and schematically represented in Fig. 6.7. In contrast to the previous explored setting, the cost model is specified considering that the farm will be managed in groups of five wind turbines, activating a global campaign cost if at least one component of a group is inspected, monitored, or repaired, plus a surplus from individual interventions. The campaign cost c_{camp} represents, in practice, the mobilization cost associated with the charter of crew transfer or offshore supply vessels. The cost model penalizes each campaign with 1 monetary unit, inspections conducted at the hotspots above and below the waterline with 0.8 and 3.8 monetary units, respectively, and sensor installations or replacements with 1.8 monetary units for the hotspot above the waterline and 7.8 monetary units for the one below the waterline. Repair, failure, and replacement costs are equally specified as in the previous setting (Section 6.4.2. One monetary unit is assumed, in this case, to be equivalent to 15,000 €, thereby providing a premise for the cost saving estimates reported in Fig. 6.8.

Besides the cost model, the dimensionality of the decision problem differs from the previous setting, including, in this instance, a total of 15 fatigue hotspots in each group of 5 offshore wind substructures.

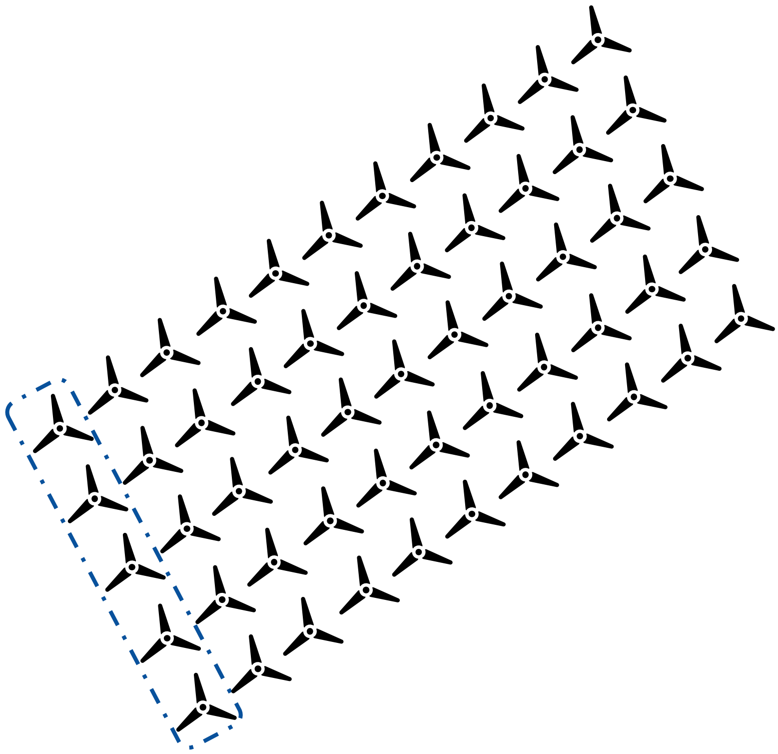


Fig. 6.7. Schematic representation of a symbolic farm arrangement, inspired by Belwind offshore wind park. In the numerical experiments, the policy is optimized for groups of five wind turbines, constituting the proxy for the management of the offshore wind farm.

Since the state space for one offshore wind structure features 113,402 states, i.e. crack, deterioration rate, stress range, sensors health and system failure states, the state space for a group of 5 offshore wind substructures scales up to 567,010 states. Note that the state space increases linearly with the number of considered hotspots due to the proposed factored POMDP representation, which otherwise would increase exponentially with the number of hotspots in a flat representation, reaching up in that case up to, approximately, $4.59 \cdot 10^{68}$ states.

In terms of the neural network architecture, DDMAC features in this setting 10 actor networks, one for each accessible fatigue hotspot, with two fully-connected hidden layers with 100 neurons each. Activation functions, error functions and learning rates are specified equally as in the previous setting. The critic network, this time judging the actions of 10 actors as a function of the 15 considered components, contains two hidden fully connected layers with 400 neurons, also specified with the same activation and error functions as in the previous case (Section 6.4.2). Calendar-based, heuristic, and DDMAC strategies are compared with respect to the resulting total expected cost, retrieved over 5,000 policy evaluations.

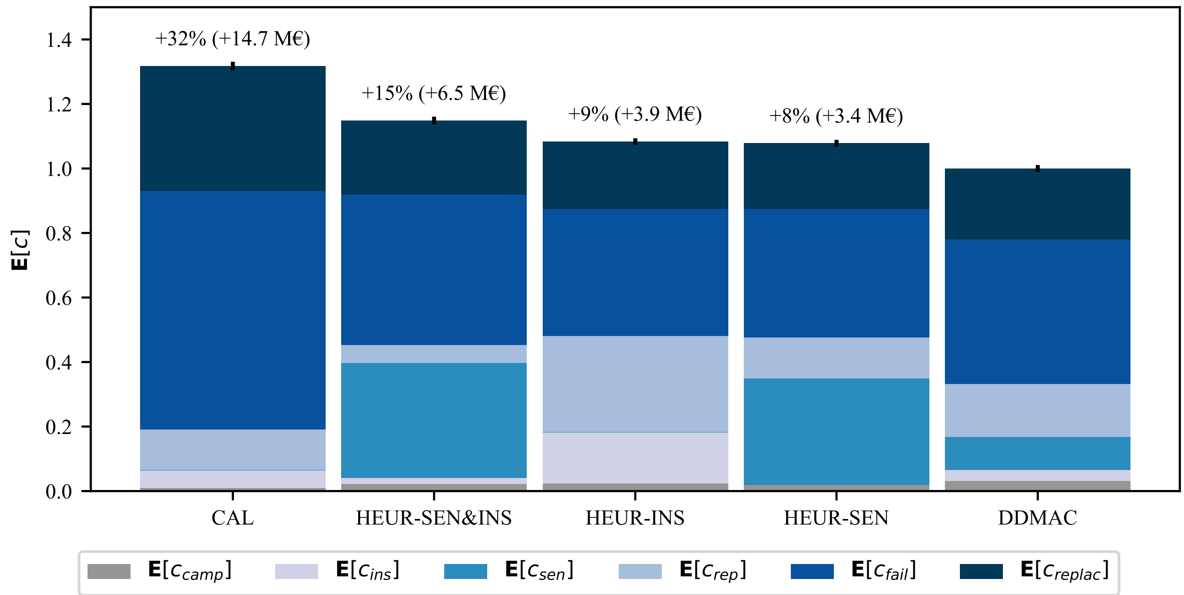


Fig. 6.8. Expected cost results of the numerical experiments conducted for the management of an offshore wind farm, divided into, campaign $\mathbf{E}[c_{camp}]$, inspection $\mathbf{E}[c_{ins}]$, sensor installation/replacement $\mathbf{E}[c_{sen}]$, perfect-repair $\mathbf{E}[c_{rep}]$, failure $\mathbf{E}[r_{fail}]$, and replacement $\mathbf{E}[r_{replac}]$ expected costs. The bar chart compares the resulting expected cost from DDMAC, calendar-based, and various heuristic strategies.

Fig. 6.8 shows the comparison of all the tested policies, emphasizing the contribution of campaigns $\mathbb{E}[c_{camp}]$, inspections $\mathbb{E}[c_{ins}]$, monitoring $\mathbb{E}[c_{sen}]$, repairs $\mathbb{E}[c_{rep}]$, failures $\mathbb{E}[c_{fail}]$ and replacements $\mathbb{E}[c_{replac}]$ to the total expected cost $\mathbb{E}[c_{tot}]$. Note that the cost dependence, induced by the shared campaign cost amongst a group of wind substructures, will not affect the decisions or costs of a corrective strategy, and it is therefore not included here.

In this setting, DDMAC's strategy outperforms again all the other tested policies, providing costs savings ranging from 32% to 8%, thus yielding from 14.7 M€ to 3.4 M€ absolute cost savings for the management of the offshore wind farm. Although calendar and heuristic-based policies consider campaign interventions in the definition of the decision rules, i.e. planning inspections or sensor installations of a group of components at equidistant time intervals, DDMAC optimally allocates monitoring, inspection, and maintenance interventions, providing significant benefits also for cost dependent structural systems.

To better visualize the actions assigned by DDMAC strategies, Fig. 6.9 illustrates one policy realization for the management of 10 wind substructures over the 20-year planned horizon, representing therefore the actions allotted to two groups of five wind turbines. In general, the fatigue hotspots located above the waterline demand more interventions, due to the higher fatigue intensity and lower cost requirements, and repairs are usually not dictated before collecting information from inspections or monitoring.

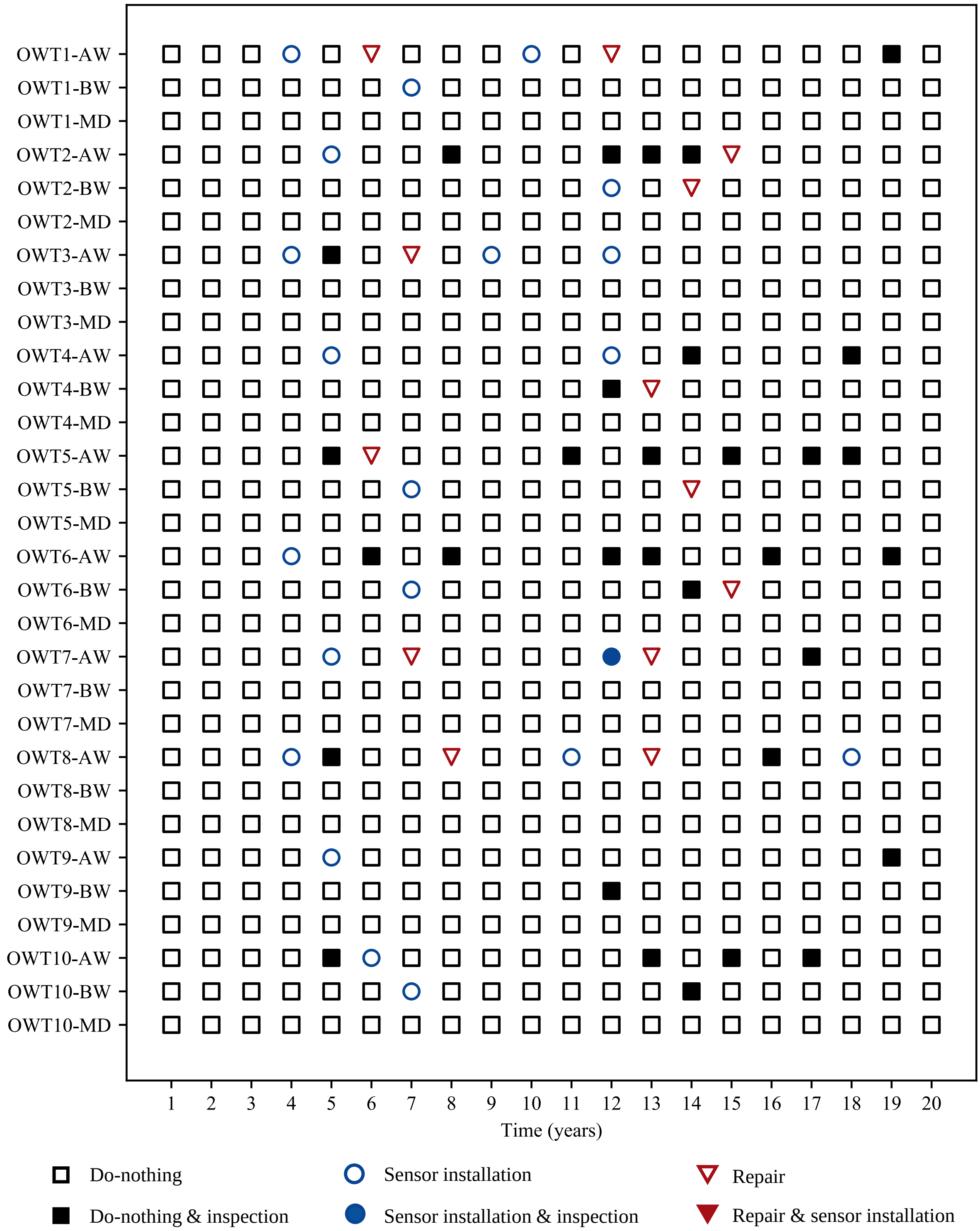


Fig. 6.9. DDMAC policy realization representing the management of an offshore wind farm over a 20-year planned horizon. Distinctive markers depict the actions assigned to each component – above the water line (AW), below the water level (BW), and below the mudline (MD) – of 10 probed offshore wind substructures.

A further analysis of the actions suggested by DDMAC indicates that identifying management strategies as a function of the dynamically updated belief state generates intricate decision patterns, interspersing monitoring, inspection, and maintenance interventions. Due to the high dimensional action space featured by structural systems, one can deduce that policies based on heuristics are not only restricted by the explored set of pre-defined decision rules [81], but the decision pattern of optimal management strategies might be also complex to parametrize by pre-determined decision rules.

6.5 Concluding remarks

This paper introduced an efficient algorithmic scheme for optimal monitoring, inspection, and maintenance planning of structural systems, with emphasis on the management of offshore wind substructures subject to fatigue deterioration.

Whereas previous studies quantify the benefits of monitoring systems through value of information analyses, the proposed formulation enables the introduction of monitoring choices within the sequential decision-making problem. In this paper, monitoring observations are specified conditional on the health of the sensing equipment, i.e. monitoring observations are only collected if the sensor is operational. Furthermore, the treatment of system failures as fully-observable events allows not only the direct inference of the system failure state, but also permits the updating of the underlying random variables.

In terms of policy optimization, we demonstrated that POMDP-based strategies, computed here by a decentralized deep multi-actor critic (DDMAC) approach, can efficiently combine monitoring, inspection, and maintenance actions, providing optimal decision sequences that might be otherwise difficult to predict. Moreover, the formulation of the decision problem as a factored POMDP, specifying the transition and observation models based on Bayesian networks, alleviates the computational complexity associated with handling multiple random variables via flat-POMDP representations [115].

As demonstrated by the conducted numerical experiments, the proposed algorithmic scheme can be applied to the management of offshore wind structural systems, providing significant cost savings compared to corrective, calendar, and state-of-the-art heuristic-based strategies. In particular, the results show that the failure risk of offshore wind substructures subjected to fatigue deterioration can be controlled by combining information from both non-destructive experiments and strain gauges to optimally dictate maintenance decisions. Further research efforts are recommended toward the development and investigation of management strategies for offshore wind structural systems under risk and/or budget constraints.

Authorship contribution statement

Morato, P. G.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Andriotis, C. P.:** Conceptualization, Methodology, Validation, Formal analysis, Software, Writing - Review & Editing. **Papakonstantinou, K. G.:** Methodology, Validation, Formal analysis, Resources, Writing - Review & Editing, Supervision. **Rigo P.:** Supervision, Project administration.

Acknowledgements

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.

Chapter 7

Conclusions and outlook

7.1 Concluding remarks

This thesis investigated methods for optimal inspection and maintenance planning of deteriorating structures with emphasis on the management of offshore wind substructures subjected to fatigue deterioration. The study also reflects on computational aspects, elaborating efficient algorithmic platforms for decision-making under uncertainty and imperfect information. The policies identified by the proposed methods have been thoroughly compared against state-of-the-art inspection and maintenance planning strategies, through numerical experiments conducted both in traditional and detailed settings. This chapter provides closure to the performed research, articulating a cohesive summary with the main findings and contributions drawn throughout the investigation.

As illustrated in Chapters 2 and 3, dynamic Bayesian networks are particularly suitable for inference tasks in probabilistic environments. From medium to high-dimensional state space settings, Partially Observable Markov Decision Processes (POMDPs) transition and observation models can be derived from dynamic Bayesian networks through space-augmentation techniques. In that case, POMDPs can be efficiently solved via state-of-the-art point-based solvers, yielding optimal inspection and maintenance strategies for both finite and infinite horizon settings. If the decision-making problem involves higher dimensional state, action, and observation spaces, constructing POMDP transition and observation models as Bayesian networks not only alleviates dimensionality concerns, but also enables the treatment of structural systems under deterioration, reliability, and cost dependence, as described in Chapter 5.

The factored POMDP representation facilitates, for instance, the modeling of deterioration dependencies amongst the constituent components of a structural system under equal or unequal correlation, by including common source Gaussian hyperparameters through decoupled hierarchical conditional structures.

Furthermore, the integration of POMDPs with a Decentralized Deep Multi Actor-Critic (DDMAC) method casts an efficient algorithmic platform for decision-making under uncertainty, approximating POMDP policies with an actor neural network, who is guided by a critic neural network. Since DDMAC adjusts the weights of actor and critic networks during the policy search according to the collected system costs, the resulting POMDP policies inherently consider system-effects, i.e. deterioration, cost, and reliability dependencies, as verified through the numerical experiments conducted in Chapter 5.

Besides inspections, monitoring information can also be used to reduce epistemic uncertainties throughout the operational lifetime, thus enabling more informed maintenance decisions. The benefits of installing a monitoring system can be systematically quantified through a value of information analysis (Chapter 4), computing the overall expected profit or loss incurred if the monitoring system is installed. Monitoring choices can also be incorporated within the sequential decision-making problem, by modeling monitoring observations conditional on the sensors' health, following the POMDP formulation introduced in Chapter 6. Strategies informed by both inspections and monitoring can thus be devised for the optimal management of engineering systems.

The extensive numerical experiments conducted throughout this investigation demonstrate that POMDP-based policies offer substantial cost savings compared to corrective, calendar, or heuristic-based strategies. Specifically, POMDPs overcome the computational challenges arising from the exponential growth of the policy space with the planning horizon, by defining adaptive policies as a function of a sufficient statistic, i.e. belief state, which intrinsically captures the dynamically updated history of actions and observations. In contrast, heuristic-based strategies are limited by the restricted explored set of predefined heuristic rules out of an immense policy space. Whereas an experienced operator might be able to draw sophisticated decision rules for traditional settings, decision-makers might guide their choices for the conception of more advanced heuristics through the examination of the patterns revealed by POMDP policy realizations.

In settings featuring high dimensional state, action, and observation spaces, the selection of optimal heuristic decision rules becomes even more challenging. As demonstrated in Chapters 5 and 6, POMDP-based DDMAC can identify optimal strategies in such settings. The advantages of POMDP-based DDMAC policies are further showcased, in a more practical example, by devising optimal monitoring, inspection, and maintenance strategies for the management of an offshore wind farm against fatigue deterioration.

In particular, the original contributions can be shortlisted as:

i) integration of DBNs with POMDP to provide and efficient algorithmic platform for decision-making under uncertainty and imperfect observations;

ii) development and investigation of management strategies for deteriorating structures that inherently consider the underlying system-effects through the combination of POMDP formulations with a DDMAC deep reinforcement learning approach;

iii) generalization of Gaussian hierarchical structures for the probabilistic treatment of engineering systems under unequally deterioration dependence;

iv) development of an algorithmic scheme for optimal monitoring, inspection, and maintenance planning of structural systems by explicitly considering the health of the sensors;

v) thorough comparison of POMDP-based policies against corrective, calendar, and heuristics-based policies;

vi) application of the proposed methods for the management of offshore wind substructures subjected to fatigue deterioration, devising strategies at the hotspot, wind turbine, and wind farm level.

7.2 Suggestions for Further Research

Further research directions are here suggested for future scientific explorations of inspection and maintenance planning methods. Most state-of-the-art point-based POMDP solvers rely on hidden Markov models for the specification of the environment dynamics, demanding state-space augmentation procedures if multiple random variables are involved in the definition of the decision problem. Adapting flat hidden Markov models to conditional formations in the definition of the POMDP dynamics might alleviate dimensionality constraints, thus increasing the potential of point-based POMDP solvers for higher dimensional state, action, and observation space settings. Additional research efforts toward the development of optimality bounds during and after the planning stage of deep reinforcement learning methods are also greatly encouraged. Potential deep reinforcement learning improvements may be achieved by conducting further investigations on: (i) the application of other architectures, e.g. hierarchical and/or convolutional neural networks; (ii) the introduction of stochastic and/or deterministic constraints; and (iii) the exploration of alternative reinforcement learning concepts, e.g. natural gradient actor-critic algorithms.

In terms of probabilistic inference, the exploration of inference methods able to propagate uncertainties directly from continuous distributions and conducting probabilistic inference by, for instance, Kalman filters or Gaussian mixtures, is also recommended. With respect to the scope of the forged decision-making problem, the focus of this work was mainly directed toward optimal inspection and maintenance planning of existing structures. The proposed stochastic optimization methods could be applied in the future for decision-making problems featuring multiple objective functions, both at the design stage and for existing deteriorating structures, including not only operational deteriorating processes, but also the occurrence of occasional extreme events. It would also be worth exploring the fusion of multiple monitoring systems, e.g. environmental, vibration-based, and structural response, with data-driven and/or physically-based models, yielding broader optimal monitoring, inspection, and maintenance schemes. Since such applications usually involve data collection from multiple sources and various potential decision-makers, the effects of data reliability and risk perception on the optimum results should also be addressed.

References

[1] H. Raiffa; and R. Schlaifer, *Applied Statistical Decision Theory.* Harvard University Graduate School of Business Administration (Division of Research); Bailey & Swinfen., 1961.

[2] H. Madsen, J. D. Sørensen, and R. Olesen, "Optimal inspection planning for fatigue damage of offshore structures," in *Proceedings of the 5th International Conference on Structural Safety and Reliability: Structural Safety and Reliability*, pp. 2099–2106, American Society of Civil Engineers, 1990.

[3] M. H. Faber, "Risk-based inspection: The framework," *Structural engineering international*, vol. 12, no. 3, pp. 186–195, 2002.

[4] J. Goyet, D. Straub, and M. H. Faber, "Risk-Based Inspection Planning of Offshore Installations," *Structural Engineering International*, vol. 12, pp. 200–208, 8 2002.

[5] J. G. Rangel-Ramírez and J. D. Sørensen, "Risk-based inspection planning optimisation of offshore wind turbines," *Structure and Infrastructure Engineering*, vol. 8, no. 5, pp. 473–481, 2012.

[6] J. Luque and D. Straub, "Risk-based optimal inspection strategies for structural systems using dynamic Bayesian networks," *Structural Safety*, vol. 76, pp. 68–80, 1 2019.

[7] D. Straub, "Stochastic Modeling of Deterioration Processes through Dynamic Bayesian Networks," *Journal of Engineering Mechanics*, vol. 135, no. 10, pp. 1089–1099, 2009.

[8] J. J. Nielsen and J. D. Sørensen, "On risk-based operation and maintenance of offshore wind turbine components," *Reliability Engineering and System Safety*, vol. 96, no. 1, pp. 218–229, 2011.

[9] J. S. Nielsen and J. D. Sørensen, "Computational framework for risk-based planning of inspections, maintenance and condition monitoring using discrete Bayesian networks," *Structure and Infrastructure Engineering*, vol. 14, no. 8, pp. 1082–1094, 2018.

[10] J. Nielsen and J. Sørensen, "Methods for Risk-Based Planning of O&M of Wind Turbines," *Energies*, vol. 7, pp. 6645–6664, 10 2014.

[11] D. Y. Yang and D. M. Frangopol, "Probabilistic optimization framework for inspection/repair planning of fatigue-critical details using dynamic bayesian networks," *Computers & Structures*, vol. 198, pp. 40–50, 2018.

[12] D. Y. Yang, J. Teng, and D. M. Frangopol, "Cross-entropy-based adaptive importance sampling for time-dependent reliability analysis of deteriorating structures," *Structural Safety*, vol. 66, pp. 38–50, 2017.

[13] P. Thoft-Christensen and J. Dalsgård Sørensen, "Reliability of structural systems with correlated elements," *Applied Mathematical Modelling*, vol. 6, no. 3, pp. 171–178, 1982.

[14] S. Ito, G. Deodatis, Y. Fujimoto, H. Asada, and M. Shinozuka, "Non-periodic inspection by bayesian method ii: structures with elements subjected to different stress levels," *Probabilistic engineering mechanics*, vol. 7, no. 4, pp. 205–215, 1992.

[15] G. Deodatis, H. Asada, and S. Ito, "Reliability of aircraft structures under non-periodic inspection: a bayesian approach," *Engineering Fracture Mechanics*, vol. 53, no. 5, pp. 789–805, 1996.

[16] I. Enevoldsen and J. D. Sørensen, "Reliability-based optimization of series systems of parallel systems," *Journal of structural engineering*, vol. 119, no. 4, pp. 1069–1084, 1993.

[17] D. Straub and I. Papaioannou, "Bayesian updating with structural reliability methods," *Journal of Engineering Mechanics*, vol. 141, no. 3, p. 34, 2015.

[18] E. Bismut and D. Straub, "Optimal Adaptive Inspection and Maintenance Planning for Deteriorating Structural Systems Reliability of hydraulic structures considering spatial variability View project AdaptRisk View project Optimal Adaptive Inspection and Maintenance Planning for Dete," no. November, 2019.

[19] E. Bismut, J. Luque, and D. Straub, "Optimal prioritization of inspections in structural systems considering component interactions and interdependence," no. August, pp. 6–10, 2017.

[20] E. Bismut and D. Straub, "Adaptive direct policy search for inspection and maintenance planning in structural systems," in *Proc. 6th International Symposium on Life-Cycle Civil Engineering (IALCCE)*, 2018.

[21] K. G. Papakonstantinou and M. Shinozuka, "Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory," *Reliability Engineering and System Safety*, vol. 130, pp. 214–224, 2014.

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

[23] R. B. Corotis, J. Hugh Ellis, and M. Jiang, "Modeling of risk-based inspection, maintenance and life-cycle cost with partially observable Markov decision processes," *Structure and Infrastructure Engineering*, vol. 1, pp. 75–84, 3 2005.

[24] K. G. Papakonstantinou, C. P. Andriotis, and M. Shinozuka, "POMDP and MOMDP solutions for structural life-cycle cost minimization under partial and mixed observability," *Structure and Infrastructure Engineering*, vol. 14, no. 7, pp. 869–882, 2018.

[25] M. Memarzadeh, M. Pozzi, and J. Zico Kolter, "Optimal Planning and Learning in Uncertain Environments for the Management of Wind Farms," *Journal of Computing in Civil Engineering*, vol. 29, p. 04014076, 9 2015.

[26] M. Memarzadeh, M. Pozzi, and J. Z. Kolter, "Hierarchical modeling of systems with similar components: A framework for adaptive monitoring and control," *Reliability Engineering & System Safety*, vol. 153, pp. 159–169, 2016.

[27] M. Memarzadeh, M. Pozzi, and J. Z. Kolter, "Optimal Planning and Learning in Uncertain Environments for the Management of Wind Farms," *Journal of Computing in Civil Engineering*, 2014.

[28] C. P. Andriotis and K. G. Papakonstantinou, "Managing engineering systems with large state and action spaces through deep reinforcement learning," *Reliability Engineering and System Safety*, vol. 191, 2019.

[29] C. P. Andriotis and K. G. Papakonstantinou, "Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints," *arXiv preprint arXiv:2007.01380*, 2020.

[30] G. Shani, J. Pineau, and R. Kaplow, "A survey of point-based POMDP solvers," *Autonomous Agents and Multi-Agent Systems*, vol. 27, no. 1, pp. 1–51, 2013.

[31] H. Kurniawati, D. Hsu, and W. Sun Lee, "SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces," in *Proceedings of Robotics: Science and Systems IV*, (Zurich, Switzerland), 2008.

[32] T. Smith and R. Simmons, "Focused real-time dynamic programming for MDPs: Squeezing more out of a heuristic," *Proceedings of the National Conference on Artificial Intelligence*, vol. 2, no. January, pp. 1227–1232, 2006.

[33] J. Luque and D. Straub, "Reliability analysis and updating of deteriorating systems with dynamic Bayesian networks," *Structural Safety*, vol. 62, pp. 34–46, 2016.

[34] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[35] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[36] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[37] D. M. Frangopol, "Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges," *Structure and Infrastructure Engineering*, vol. 7, pp. 389–413, 6 2011.

[38] M. G. Stewart and D. V. Rosowsky, "Time-dependent reliability of deteriorating reinforced concrete bridge decks," *Structural Safety*, vol. 20, pp. 91–109, 1 1998.

[39] F. Akgül and D. M. Frangopol, "Lifetime Performance Analysis of Existing Steel Girder Bridge Superstructures," *Journal of Structural Engineering*, vol. 130, pp. 1875–1888, 12 2004.

[40] D. V. Val, M. G. Stewart, and R. E. Melchers, "Effect of reinforcement corrosion on reliability of highway bridges," *Engineering Structures*, vol. 20, pp. 1010–1019, 11 1998.

[41] D. V. Val and R. E. Melchers, "Reliability of Deteriorating RC Slab Bridges," *Journal of Structural Engineering*, vol. 123, pp. 1638–1644, 12 1997.

[42] T. Moan, "Reliability-based management of inspection, maintenance and repair of offshore structures," *Structure and Infrastructure Engineering*, vol. 1, pp. 33–62, 3 2005.

[43] I. Lotsberg, G. Sigurdsson, A. Fjeldstad, and T. Moan, "Probabilistic methods for planning of inspection for fatigue cracks in offshore structures," *Marine Structures*, vol. 46, pp. 167–192, 3 2016.

[44] P. H. Wirsching, "Fatigue reliability in welded joints of offshore structures," *International Journal of Fatigue*, vol. 2, pp. 77–83, 4 1980.

[45] P. Schaumann, S. Lochte-Holtgreven, and S. Steppeler, "Special fatigue aspects in support structures of offshore wind turbines," *Materialwissenschaft und Werkstofftechnik*, vol. 42, pp. 1075–1081, 12 2011.

[46] W. Dong, T. Moan, and Z. Gao, "Fatigue reliability analysis of the jacket support structure for offshore wind turbine considering the effect of corrosion and inspection," *Reliability Engineering and System Safety*, vol. 106, pp. 11–27, 2012.

[47] B. Yeter, Y. Garbatov, and C. Guedes Soares, "Fatigue damage assessment of fixed offshore wind turbine tripod support structures," *Engineering Structures*, vol. 101, pp. 518–528, 2015.

[48] C. P. Andriotis, K. G. Papakonstantinou, and E. N. Chatzi, "Value of structural health information in partially observable stochastic environments," *arXiv preprint arXiv:1912.12534*, 2020.

[49] W. Fauriat and E. Zio, "Optimization of an aperiodic sequential inspection and condition-based maintenance policy driven by value of information," *Reliability Engineering & System Safety*, vol. 204, p. 107133, 2020.

[50] B. R. Ellingwood, "Risk-informed condition assessment of civil infrastructure: state of practice and research issues," *Structure and Infrastructure Engineering*, vol. 1, pp. 7–18, 3 2005.

[51] M. Sánchez-Silva, D. M. Frangopol, J. Padgett, and M. Soliman, "Maintenance and operation of infrastructure systems," *Journal of Structural Engineering*, vol. 142, no. 9, p. F4016004, 2016.

[52] M. H. Faber and M. G. Stewart, "Risk assessment for civil engineering facilities: Critical overview and discussion," *Reliability Engineering and System Safety*, vol. 80, pp. 173–184, 5 2003.

[53] S. Kim, B. Ge, and D. M. Frangopol, "Effective optimum maintenance planning with updating based on inspection information for fatigue-sensitive structures," *Probabilistic Engineering Mechanics*, vol. 58, p. 103003, 10 2019.

[54] S. J. Russell, *Artificial Intelligence : A Modern Approach.* Upper Saddle River, NJ : Prentice Hall., 3rd ed., 2010.

[55] J. D. Sørensen, R. Rackwitz, M. H. Faber, and P. Thoft-Christensen, "Modelling in Optimal Inspection and Repair," in *Proceedings of the 10th International Conference on Offshore Mechanics and Arctic Engineering*, vol. 2, (United States), pp. 281–288, American Society of Mechanical Engineers, 1991.

[56] E. Bismut and D. Straub, "Optimal adaptive inspection and maintenance planning for deteriorating structural systems," *arXiv preprint arXiv:2102.06016*, 2021.

[57] D. Y. Yang and D. M. Frangopol, "Probabilistic optimization framework for inspection/repair planning of fatigue-critical details using dynamic bayesian networks," *Computers & Structures*, vol. 198, pp. 40–50, 2018.

[58] I. Tien and A. Der Kiureghian, "Reliability assessment of critical infrastructure using bayesian networks," *Journal of Infrastructure Systems*, vol. 23, no. 4, p. 04017025, 2017.

[59] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons, 2014.

[60] P. G. Morato, J. S. Nielsen, A. Q. Mai, and P. Rigo, "POMDP based maintenance optimization of offshore wind substructures including monitoring," in *13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP 2019*, pp. 270–277, 2019.

[61] M. H. Faber, "Risk-based inspection: The framework," *Structural Engineering International*, vol. 12, no. 3, pp. 186–194, 2002.

[62] D. Straub, *Generic Approaches to Risk Based Inspection Planning for Steel Structures.* PhD thesis, Swiss Federal Institute of Technology Zürich (ETH), 2004.

[63] O. Ditlevsen and H. O. Madsen, *Structural Reliability Methods.* Department of Mechanical Engineering. Technical University of Denmark., 2007.

[64] H. O. Madsen, S. Krenk, and N. C. Lind, *Methods of structural safety.* Dover Publications Inc., 2006.

[65] D. Straub, "Reliability updating with equality information," *Probabilistic Engineering Mechanics*, vol. 26, no. 2, pp. 254–258, 2011.

[66] I. Lotsberg, G. Sigurdsson, A. Fjeldstad, and T. Moan, "Probabilistic methods for planning of inspection for fatigue cracks in offshore structures," *Marine Structures*, vol. 46, pp. 167–192, 2016.

[67] F. V. Jensen, *Introduction to Bayesian Networks.* Berlin, Heidelberg: Springer-Verlag., 1996.

[68] K. P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, University of California, Berkeley, 2002.

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

[70] M. J. Kochenderfer, C. Amato, G. Chowdhary, J. P. How, H. J. D. Reynolds, J. R. Thornton, P. A. Torres-Carrasquillo, K. N. Ure, and J. Vian, *Decision Making Under Uncertainty: Theory and Application*. The MIT Press, 1st ed., 2015.

[71] J. S. Nielsen and J. D. Sørensen, “Risk-based Decision Making for Deterioration Processes Using POMDP,” in *12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12* (T. Haukaas, ed.), (Vancouver, Canada), p. 8, Civil Engineering Risk and Reliability Association, 2015.

[72] C.-A. Robelin and S. M. Madanat, “History-dependent bridge deck maintenance and replacement optimization with Markov decision processes,” *Journal of Infrastructure Systems*, vol. 13, no. 3, pp. 195–201, 2007.

[73] J. W. Kim, G. B. Choi, and J. M. Lee, “A POMDP framework for integrated scheduling of infrastructure maintenance and inspection,” *Computers & Chemical Engineering*, vol. 112, pp. 239–252, 2018.

[74] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, “Planning and acting in partially observable stochastic domains,” *Artificial Intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[75] M. T. J. Spaan and N. Vlassis, “Perseus: Randomized point-based value iteration for POMDPs,” *Journal of Artificial Intelligence Research*, vol. 24, pp. 195–220, 2005.

[76] J. Pineau, G. Gordon, and S. Thrun, “Point-based value iteration: An anytime algorithm for POMDPs,” in *IJCAI International Joint Conference on Artificial Intelligence*, pp. 1025–1030, 2003.

[77] R. A. Howard, “Information value theory,” *IEEE Transactions on systems science and cybernetics*, vol. 2, no. 1, pp. 22–26, 1966.

[78] E. Walraven and M. T. Spaan, “Point-based value iteration for continuous POMDPs,” *Journal of Artificial Intelligence Research*, vol. 65, pp. 307–341, 2019.

[79] G. DNV, “Probabilistic methods for planning of inspection for fatigue cracks in offshore structures,” *DNV GL, Oslo, Norway, Standard No. DNVGL-RP-C210*, 2015.

[80] M. Memarzadeh and M. Pozzi, “System-level inspection scheduling: An approach based on stochastic future allocation,” *International Workshop on Structural Health Monitoring*, no. September, 2015.

[81] P. G. Morato, K. Papakonstantinou, C. Andriotis, J. S. Nielsen, and P. Rigo, “Optimal inspection and maintenance planning for deteriorating structures through dynamic bayesian networks and markov decision processes,” *arXiv preprint arXiv:2009.04547*, 2020.

[82] G. DNV, “Fatigue design of offshore steel structures,” *Recommended Practice DNVGL-RP-C203*, vol. 20, 2016.

[83] M. Memarzadeh and M. Pozzi, "Value of information in sequential decision making: Component inspection, permanent monitoring and system-level scheduling," *Reliability Engineering and System Safety*, vol. 154, pp. 137–151, 2016.

[84] P. G. Morato, Q. A. Mai, P. Rigo, and J. S. Nielsen, "Point-based POMDP risk based inspection of offshore wind substructures," *Proceedings of IALCCE 2018: The Sixth International Symposium on Life-Cycle Civil Engineering*, pp. 3069–3076, 2018.

[85] R. Rackwitz, A. Lentz, and M. Faber, "Socio-economically sustainable civil engineering infrastructures by optimization," *Structural safety*, vol. 27, no. 3, pp. 187–229, 2005.

[86] D. M. Frangopol and M. Soliman, "Life-cycle of structural systems: recent achievements and future directions," *Structure and Infrastructure Engineering*, vol. 12, pp. 1–20, 1 2016.

[87] D. M. Frangopol, M.-J. Kallen, and J. M. v. Noortwijk, "Probabilistic models for life-cycle performance of deteriorating structures: review and future directions," *Progress in structural engineering and Materials*, vol. 6, no. 4, pp. 197–212, 2004.

[88] D. M. Frangopol and M. Liu, "Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost," *Structure and infrastructure engineering*, vol. 3, no. 1, pp. 29–41, 2007.

[89] K. G. Papakonstantinou and M. Shinozuka, "Probabilistic model for steel corrosion in reinforced concrete structures of large dimensions considering crack effects," *Engineering Structures*, vol. 57, pp. 306–326, 2013.

[90] L. Long, Q. A. Mai, P. G. Morato, J. D. Sørensen, and S. Thöns, "Information value-based optimization of structural and environmental monitoring for offshore wind turbines support structures," *Renewable Energy*, vol. 159, pp. 1036–1046, 2020.

[91] N. Hlaing, P. G. Morato Dominguez, P. Rigo, P. Amirafshari, A. Kolios, and J. S. Nielsen, "The effect of failure criteria on risk-based inspection planning of offshore wind support structures," in *Proc. 7th International Symposium on Life-Cycle Civil Engineering (IALCCE)*, 2020.

[92] J. Song and W.-H. Kang, "System reliability and sensitivity under statistical dependence by matrix-based system reliability method," *Structural Safety*, vol. 31, no. 2, pp. 148–156, 2009.

[93] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[94] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[95] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *arXiv preprint arXiv:1511.05952*, 2015.

[96] E. P. Popov, V. A. Zayas, and S. A. Mahin, "Inelastic cyclic behavior of tubular braced frames," *Journal of the Structural Division*, vol. 106, no. 12, pp. 2375–2390, 1980.

[97] T. Moan, J. Amdahl, T. Granli, and O. Hellan, "Collapse behaviour of offshore structural systems," *Advances in Marine Structures–2*, pp. 469–494, 1991.

[98] R. E. Barlow and K. D. Heidtmann, "Computing k-out-of-n system reliability," *IEEE Transactions on Reliability*, vol. 33, no. 4, pp. 322–323, 1984.

[99] R. Schneider, S. Thöns, and D. Straub, "Reliability analysis and updating of deteriorating systems with subset simulation," *Structural Safety*, vol. 64, pp. 20–36, 1 2017.

[100] R. Schneider, S. Thöns, and D. Straub, "Reliability analysis and updating of deteriorating systems with subset simulation," *Structural Safety*, vol. 64, pp. 20–36, 2017.

[101] D. Straub and A. Der Kiureghian, "Reliability acceptance criteria for deteriorating elements of structural systems," *Journal of Structural Engineering*, vol. 137, no. 12, pp. 1573–1582, 2011.

[102] B. Le and J. Andrews, "Modelling wind turbine degradation and maintenance," *Wind Energy*, vol. 19, no. 4, pp. 571–591, 2016.

[103] R. Dawid, D. McMillan, and M. Revie, "Review of markov models for maintenance optimization in the context of offshore wind," in *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, pp. 269–279, 2015.

[104] M. Florian and J. D. Sørensen, "Risk-based planning of operation and maintenance for offshore wind farms," *Energy Procedia*, vol. 137, pp. 261–272, 2017.

[105] H. Seyr and M. Muskulus, "Decision support models for operations and maintenance for offshore wind farms: A review," *Applied Sciences*, vol. 9, no. 2, p. 278, 2019.

[106] A. May, D. McMillan, and S. Thöns, "Economic analysis of condition monitoring systems for offshore wind turbine sub-systems," *IET Renewable Power Generation*, vol. 9, no. 8, pp. 900–907, 2015.

[107] Z. Guédé, B. Bigourdan, A. Rouhan, J. Goyet, and P. Renard, "Framework for Risk-Based Inspection of an Offshore Wind Farm," in *Volume 2B: Structures, Safety and Reliability*, 2013.

[108] A. Rouhan and F. Schoefs, "Probabilistic modeling of inspection results for offshore structures," *Structural Safety*, vol. 25, no. 4, pp. 379–399, 2003.

[109] J. Goyet, D. Straub, and M. H. Faber, "Risk-based inspection planning of offshore installations," *Structural Engineering International: Journal of the International Association for Bridge and Structural Engineering (IABSE)*, vol. 12, no. 3, pp. 200–208, 2002.

[110] S. Marquez-Dominguez and J. D. Sorensen, "Fatigue Reliability and Calibration of Fatigue Design Factors for Offshore Wind Turbines," *Energies*, vol. 5, no. 6, pp. 1816–1834, 2012.

[111] C. Devriendt, F. Magalhães, W. Weijtjens, G. De Sitter, Á. Cunha, and P. Guillaume, "Structural health monitoring of offshore wind turbines using automated operational modal analysis," *Structural Health Monitoring*, vol. 13, no. 6, pp. 644–659, 2014.

[112] W. Weijtjens, T. Verbelen, G. De Sitter, and C. Devriendt, “Foundation structural health monitoring of an offshore wind turbine—a full-scale case study,” *Structural Health Monitoring*, vol. 15, no. 4, pp. 389–402, 2016.

[113] S. Thöns, “On the Value of Monitoring Information for the Structural Integrity and Risk Management,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 1, pp. 79–94, 2018.

[114] Q. A. Mai, W. Weijtjens, C. Devriendt, P. G. Morato, P. Rigo, and J. D. Sørensen, “Prediction of remaining fatigue life of welded joints in wind turbine support structures considering strain measurement and a joint distribution of oceanographic data,” *Marine Structures*, vol. 66, pp. 307–322, 2019.

[115] P. G. Morato, C. Andriotis, K. Papakonstantinou, and P. Rigo, “Optimal management of deteriorating structures considering system effects: a deep reinforcement learning approach,” *under review*, 2021.

[116] DNV-GL, “DNVGL-ST-0126: Support structures for wind turbines,” *Oslo, Norway: DNV GL*, 2016.

[117] J. Jonkman and W. Musial, “Offshore code comparison collaboration (OC3) for IEA Wind Task 23 offshore wind technology and deployment,” tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2010.