18th EAWE PhD Seminar on Wind Energy
2-4 November 2022
Bruges, Belgium

Interpretation of offshore wind management policies identified via partially observable Markov decision processes

Nandar Hlaing[a], **Pablo G. Morato**[a], **Konstantinos G. Papakonstantinou**[b], **Charalampos P. Andriotis**[c], and **Philippe Rigo**[a]

[a]Naval & Offshore Engineering, ArGEnCo, University of Liege, 4000 Liege, Belgium
[b]Department of Civil & Environmental Engineering, The Pennsylvania State University, University Park, PA 16802, USA
[c]Faculty of Architecture & the Built Environment, Delft University of Technology, 2628 BL Delft, The Netherlands

E-mail: nandar.hlaing@uliege.be

Keywords: Offshore wind turbines; Inspection and maintenance planning; Markov decision processes.

1 Introduction

The installation of offshore wind turbines, profiting from available abundant and stable wind resources, has been steadily increasing in the last decade, yet preserving offshore wind structures in a good condition throughout their lifetime still remains a challenge. Structural components are exposed to deterioration mechanisms (e.g., fatigue, corrosion, among others), and far offshore, inspection and maintenance (I&M) operations can be complex and expensive. Hence the need for efficient optimal I&M planning methods has been increased in order to control the risk of structural failures by timely allocating inspection and maintenance interventions.

Identifying optimal I&M policies demands the solution of a complex sequential decision-making problem under uncertainty and imperfect information. Whereas time-, condition-, or heuristic-based strategies are conventionally followed in the offshore wind industry in order to alleviate the aforementioned computational difficulties, the resulting policies statically select inspection and maintenance actions and/or consist in predefined heuristic decision rules, e.g., equidistant inspections, repairs after detection inspection outcomes, which are optimized by exploring a subset out of the vast policy space. Instead, optimal management strategies can be identified via partially observable Markov decision processes (POMDPs), relying on mathematical principles conceived for planning under uncertainty [1]. POMDP policies, efficiently computed through point-based solvers, provide optimal adaptive I&M strategies that ultimately result in substantial cost benefits compared to their state-of-the-art counterparts [2], also demonstrated in offshore wind inspection and maintenance planning settings [3].

Even if recently reported results demonstrate the benefits of implementing POMDP-based adaptive policies for the management of offshore wind assets, the interpretation and execution of POMDP-based strategies by decision-makers (e.g., designers, operators, etc.) accustomed to calendar- and/or condition-based conventional I&M approaches might be initially challenging. In this work, we analyze and interpret POMDP-based policies with the objective of accelerating their practical implementation by offshore wind asset management decision-makers. Also, we showcase the inherent flexibility and adaptability properties offered by POMDP-based policies in a typical offshore wind inspection and maintenance planning setting, in which a decision-maker opts for an action other than the one suggested in the optimal POMDP policy.

2 Optimal I&M planning for offshore wind structures through POMDPs

A Markov decision process (MDP) is a 5-tuple $\langle S, A, T, R, \gamma \rangle$ controlled stochastic process for optimal planning under uncertainty and perfect information. At every decision step, the agent observes state $s \in S$ and takes an action $a \in A$, then the state randomly transitions to state $s' \in S$ according to a stochastic transitional model $T(s, a, s') =$

$P(s'|s,a)$, and finally the agent receives a reward $R(s,a)$. An MDP policy $(\pi : S \rightarrow A)$ prescribes actions as a function of the current state, with the main objective of identifying the optimal policy $\pi^*(s)$, resulting in the maximum expected rewards (or minimum expected cost).

A POMDP is a generalization of an MDP in which the agent only receives partial information about the current state. In this case, the agent reasons according to the current belief **b**, i.e., a probability distribution over states. A POMDP is defined as a 7-tuple $\langle S,A,O,T,Z,R,\gamma \rangle$ controlled stochastic process. While a POMDP transitional model corresponds to the underlying MDP, an observation model is additionally defined by specifying the probability $Z(o,s',a) = P(o \mid s',a)$ of collecting observation $o \in O$ after taking action a. After taking action a and collecting observation o, the belief **b** is updated via Bayes' rule:

$$b(s') \propto P(o \mid s',a) \sum_{s \in S} P(s' \mid s,a) b(s). \tag{1}$$

Since beliefs are dynamically updated, the current belief, **b** is a sufficient statistic of the past taken actions and collected observations. A POMDP policy therefore maps the current belief **b** to the action. As for an MDP, the goal is to identify the optimal policy $\pi^*(\mathbf{b})$ leading to the maximum expected reward.

The decision-making problem corresponding to the optimal inspection and maintenance planning for offshore wind structures can be adequately formulated as a POMDP, in which the agent reasons in a stochastic environment (i.e., probabilistic deterioration model) and under imperfect information (i.e., measurement uncertainty associated with inspection techniques). Once the optimal POMDP policy $\pi^*(\mathbf{b})$ is identified, the decision-maker (e.g., operator, designer, etc) selects inspection and/or maintenance actions according to the current belief state. As opposed to static decision rules, e.g., calendar- or condition-based maintenance approaches, POMDP policies are inherently adaptive since beliefs are dynamically updated, thus resulting in substantial cost benefits.

2.1 Solving POMDPs

The exact solution of a POMDP demands the identification of optimal actions for each belief state, which as mentioned before, is a continuous probability distribution over states, thus rendering the problem computationally challenging. Whereas value iteration algorithms or grid-based interpolation techniques might work well for solving very low-dimensional state space POMDPs, their application to higher dimensional state space POMDPs remains limited, also due to computational tractability problems. However, the recently developed point-based solvers, by executing Bellman backups only for a set of reachable belief points, have enabled the solution of medium to high dimensional state space POMDPs [4]. Since the value function is generally piece-wise linear and convex, it can be parametrized through a finite set Γ of α-vectors, each of them associated with an action [2]. At a certain belief state, the optimal action is, therefore, indicated by the α-vector that maximizes the value function. Point-based solvers are usually developed for the solution of infinite horizon settings, yet practical applications normally correspond to finite horizon problems, e.g., the operational lifetime of offshore wind structural components is often considered as 20 or 30 years. In that case, the infinite horizon POMDP can be transformed into a finite horizon POMDP through state augmentation techniques [1, 2].

3 Interpretation of POMDP-based management policies

With the objective of facilitating the interpretation of POMDP-based offshore wind management policies, we conduct hereafter an I&M planning case study for a fatigue-sensitive offshore wind structural component, inspired by [3]. The I&M decision problem is formulated as a POMDP, adequately defining the elements of the POMDP tuple, as follows:

- States: The structural component deterioration states correspond to the discretized fatigue crack size. In this study, the crack size is discretized into 40 deterioration states, with the last one indicating a failure state.
- Actions: Three action-observation combinations are considered, (i) Do-nothing/No-inspection (DN-NI), (ii) Do-nothing/Inspection (DN-I), and (iii) Perfect-repair/No-inspection (PR-NI).
- Observations: Inspections provide binary indications, resulting in either 'crack detection' or 'no crack detection'. If an inspection is not performed, no additional information is collected.
- Transition probabilities: The transitional model associated with a Do-nothing (DN) action is estimated through crack propagation Monte Carlo simulations, where the crack growth is computed according to Paris

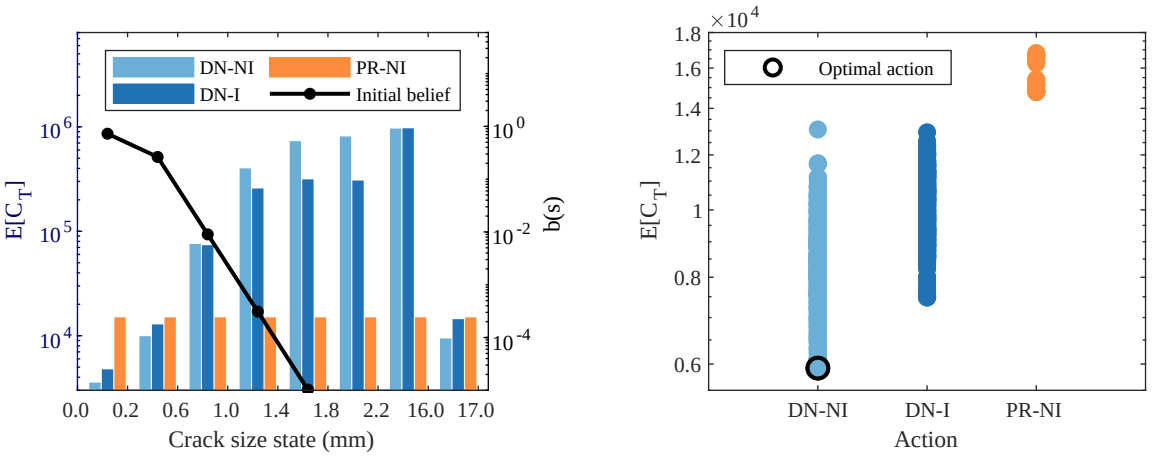
law. If a Perfect-repair (PR) action is undertaken, the structural component deterioration transitions to its initial belief condition state.

- Observation probabilities: The observation model is defined according to the detection probability curve that corresponds to eddy current inspection techniques [5].
- Rewards: At every decision step, the agent collects a reward, $R(\mathbf{b},a)$, which is a weighted sum of the belief probability $b(s)$ and state reward $R(s,a)$. A penalization of one million monetary units is charged at the last state, i.e., failure condition, whereas 1,000 and 10,000 thousand monetary units are assigned as inspection and repair costs, respectively.

In this case study, the structural component lifetime is defined as 20 years and the corresponding finite horizon POMDP is computed via SARSOP point-based solver [6]. The resulting optimal policies are parametrized by a set of α-vectors, and as mentioned before, each α-vector is associated with a specific action. At each decision point, the decision-maker selects the α-vector (and corresponding action) that maximizes the value function $V^*(\mathbf{b})$ (minimizes the total expected cost):

$$V^*(\mathbf{b}) = \max_{\alpha\in\Gamma} \sum_{s\in S} b(s)\alpha(s). \tag{2}$$

The expected total cost associated with each α-vector can be simply computed as the weighted sum of the expected total cost corresponding to a specific deterioration state $\alpha(s)$ and the probability of being in that state $b(s)$. Figure 1a illustrates the expected total cost resulting from three α-vectors, indicating both the corresponding deterioration state values along with the representation of the initial belief $\mathbf{b}_0$. The key observation is that the actions recommended in POMDP-based policies are selected according to the current belief state $\mathbf{b}$, which is dynamically updated after each taken action and collected observation, as mentioned in Section 2. The expected costs associated with all α-vectors available at the initial decision step (i.e., $\mathbf{b}_0$) are additionally represented in Figure 1b. Logically, the optimal decision at this point is DN-NI, and its corresponding value function indicates the total expected cost $\mathbb{E}[C_T]$ for the considered 20-year decision horizon.



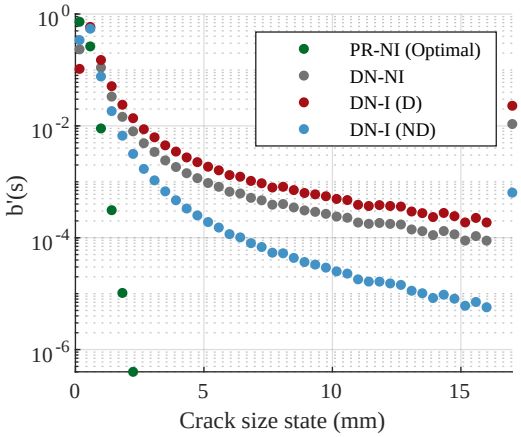
(a) Initial belief (black line) and state values of three α-vectors (colored bars).

(b) Expected total cost associated with each α-vector (and corresponding action).

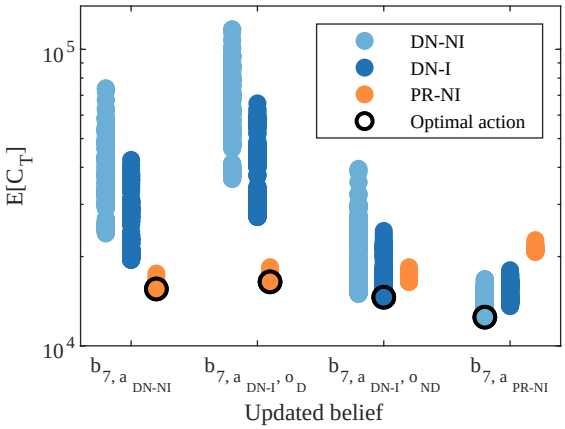
Figure 1: Initial probability distribution over deterioration states (i.e., initial belief, $\mathbf{b}_0$) and expected total cost resulting from each α-vector at the initial decision step.

3.1 What if the optimal policy is not strictly followed?

In this study, we investigate the effect of selecting an alternative action rather than the optimal one suggested in the POMDP policy. We consider that the optimal POMDP policy is followed up to year 7, and at that point, the decision-maker is evaluating the potential economic implications of avoiding a perfect repair maintenance intervention, which is the action suggested in the optimal POMDP policy, as showcased in Figure 2b. Previously, two crack detection inspection outcomes were reported at years 6 and 7, thus indicating a high structural failure risk, which could be effectively mitigated by conducting a repair action (Figure 3a). In that case, the structural condition



(a) Updated beliefs for all action-observation combinations.



(b) Total expected cost for all action-observation combinations.

Figure 2: Updated beliefs and corresponding total expected cost for all action-observation combinations at year 8.

will be restored, and the updated belief will transition to its initial deterioration condition, $\mathbf{b}_0$, as illustrated in Figure 2a with green markers. If the decision-maker opts, however, for an alternative action at year 8, the expected total cost and the regret, i.e. the extra cost associated with potentially suboptimal actions, can be straightforwardly computed through a Bellman backup operation, as:

$$V(\mathbf{b}_7) = \sum_{s \in S} b_7(s) R(s,a) + \gamma V(\mathbf{b}'_7), \tag{3}$$

where $\mathbf{b}$ and $\mathbf{b}'$ correspond to the current and updated beliefs, respectively, and $R(s,a)$ stands for the reward associated with the action taken. Specifically, the potential alternative actions at this decision point are:

- Do-nothing/No-inspection (DN-NI), in which the fatigue deterioration will naturally progress according to the defined transition model, as illustrated with grey markers in Figure 2a.
- Do-nothing/Inspection (DN-I), which can result in either a crack detection or no crack detection outcome. The corresponding updated beliefs are plotted in Figure 2a with red and blue markers, respectively. Since two inspection outcomes can be collected, in this case, the expected total cost estimated can be computed as:

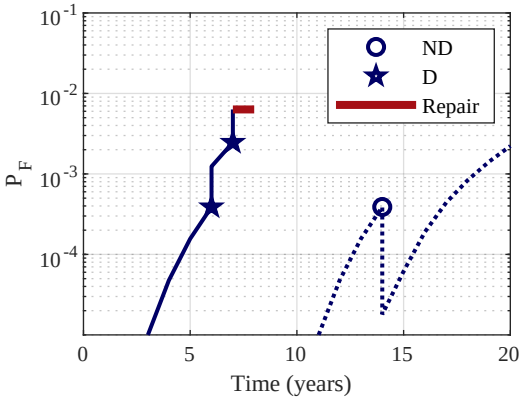
$$V(\mathbf{b}_7) = \sum_{s \in S} b_7(s) R(s, a_{DN-I}) + \gamma \left[\sum_{o \in O} p(o \mid \mathbf{b'}_{7,a_{DN-I}}) \cdot V(\mathbf{b}'_{7,a_{DN-I},o}) \right], \tag{4}$$

  where $p(o \mid \mathbf{b}')$ represents the probability associated with each inspection outcome.

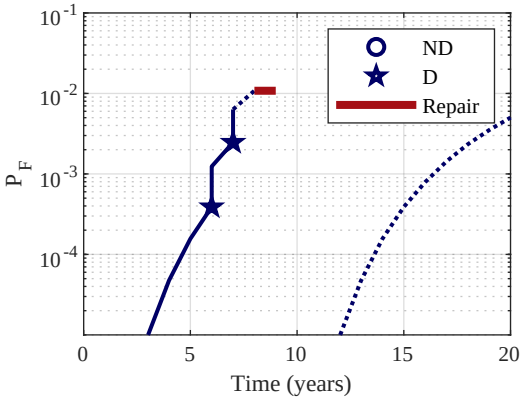
Gathering all action-observation combinations, Figure 2b illustrates the expected cost associated with each action. The suggested action and expected total cost corresponding to each updated belief, $\mathbf{b}'$, can be directly computed through the α-vectors included in the original POMDP policy, as indicated in Equation (2). For instance, if the decision-maker follows the optimal policy and opts for a perfect repair action at year 8, the suggested subsequent optimal action is DN-NI (Figure 3a). Instead, if the decision-maker selects a DN-NI action at year 8, the logical suggested action is an immediate repair action the following year. Note that in the reported results, the total expected cost, i.e., $V(\mathbf{b}')$, is computed from the original POMDP policy. In order to exactly evaluate the economic implication of selecting suboptimal actions, the formulated POMDP can be solved again, considering $\mathbf{b}'$ as the initial belief in a reduced finite decision horizon, which corresponds, in this particular example, to twelve time steps. However, only minor differences in the estimation of the expected total cost between the two aforementioned approaches are observed in this study.

Further examining all alternative action-observation combinations available at year 8, Figure 3 showcases typical resulting policy realizations. As one could expect, a perfect repair is suggested after a DN-NI action is selected at year 8 (Figure 3b), then the structural component condition is restored, and no additional future

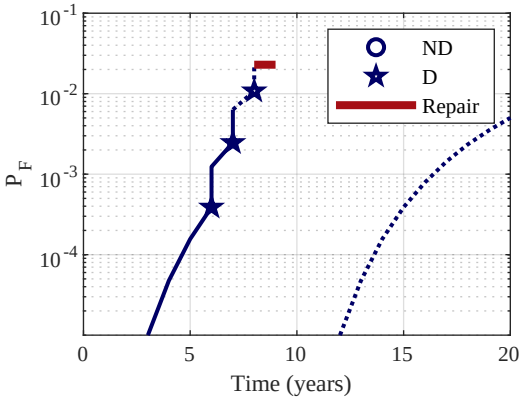
interventions are usually needed. If a DN-I action is taken at year 8, the POMDP policy suggests a subsequent repair action after a crack detection inspection outcome is observed (Figure 3c), whereas if the inspection results in a no detection outcome, a repair is not planned, and instead, the policy realization shows a series of inspections for the remainder of the horizon (Figure 3d).



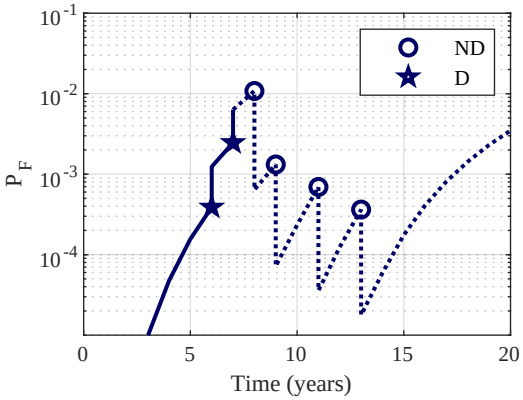
(a) Policy realization for a_{PR-NI} (following the original optimal POMDP policy).



(b) Policy realization for a_{DN-NI}. A repair action is immediately suggested the next year.



(c) Policy realization for a_{DN-I}, o_D. A repair action is immediately suggested the next year.



(d) Policy realization for a_{DN-I}, o_{ND}. A repair action is not subsequently suggested.

Figure 3: Representation of typical policy realizations for all action-observation combinations available at year 8.

In summary, Table 1 lists the expected total cost associated with all available actions at year 8 and their corresponding regret $\mathbb{E}[C_P]$. It can be observed that, in this study, the most suboptimal choice is a DN-NI action, as it results in a yet higher failure risk after year 8, while a repair still needs to be allocated the following year. Interestingly, a DN-I action is less suboptimal, in this case, since subsequent no detection inspection outcomes can still be observed, thus slightly reducing the need of a perfect repair action.

Action at year 8	$\mathbb{E}[C_T]$ (monetary units)	$\mathbb{E}[C_P](\%)$
PR-NI (Optimal)	12,493	-
DN-NI	15,552	24.5
DN-I $\begin{cases} p(o_{ND} \mid \mathbf{b}'_{7,a_{DN-I}}) = 0.5437 \\ p(o_D \mid \mathbf{b}'_{7,a_{DN-I}}) = 0.4563 \end{cases}$	15,428	23.5

Table 1: Regret incurred when selecting alternative actions other than the one suggested in the optimal original POMDP policy.

4 Conclusion

In an offshore wind inspection and maintenance (I&M) planning context, we describe the fundamentals of Partially Observable Markov Decision Processes (POMDPs) -based policies and showcase their inherent adaptive and flexible properties. Through a typical offshore wind I&M planning case study, we also demonstrate that decision-makers following POMDP-based strategies can efficiently and swiftly quantify the effect of selecting alternative actions rather than those suggested in the optimal POMDP policy. Based on the reported benefits offered by POMDP-based policies in terms of optimality [1, 7], adaptability [2, 8], and flexibility [3], along with the interpretability aspects introduced in this work, we encourage the adoption of POMDP-based I&M planning methods in the offshore wind industry.

Acknowledgements

This research is funded by the Belgian Energy Transition Fund (FPS Economy) through PhairywinD and MaxWind projects.

References

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

[2] P. G. Morato, K. G. Papakonstantinou, C. P. Andriotis, J. S. Nielsen, and P. Rigo. Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes. *Structural Safety*, 94:102140, 2022.

[3] N. Hlaing, P. G. Morato, J. S. Nielsen, P. Amirafshari, A. Kolios, and P. Rigo. Inspection and maintenance planning for offshore wind structural components: integrating fatigue failure criteria with Bayesian networks and Markov decision processes. *Structure and Infrastructure Engineering*, 18(7):983–1001, 2022.

[4] 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*, 14(7):869–882, 2018.

[5] DNV. DNV-RP-C210 Probabilistic methods for planning of inspection for fatigue cracks in offshore structures. Recommended practice, Veritasveien 1, 1363 Høvik, Norway, 2019.

[6] H. Kurniawati, D. Hsu, and W. S. Lee. SARSOP: Efficient point-Based POMDP planning by approximating optimally reachable belief spaces. In *Proceedings of Robotics: Science and Systems*, Switzerland, 2008.

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

[8] P. G. Morato, C. P. Andriotis, K. G. Papakonstantinou, and P. Rigo. Inference and dynamic decision-making for deteriorating systems with probabilistic dependencies through Bayesian networks and deep reinforcement learning. *arXiv preprint arXiv:2209.01092*, 2022.