

Response to Reviewer 1

General response

We thank the Reviewer for the careful reading of the manuscript and for the constructive feedback. In response, we strengthened the methodological transparency of the review (including a PRISMA-style description of the screening and eligibility process), added a dedicated paragraph on threats to validity, expanded the discussion of partial observability and practical POMDP remedies, and made the presentation of quantitative performance outcomes more systematic. We also improved figure/table captions, corrected typographical issues, standardised notation, and verified the reference list for duplicates.

Response to comment 1

We agree and have strengthened the methodological transparency of the review. In the revised manuscript, we added a dedicated description of the literature selection protocol, specifying: (i) the databases consulted (Scopus, Web of Science, ScienceDirect, ResearchGate, and Google Scholar), (ii) the search terms used (including combinations of “offshore wind”, “maintenance”, “reinforcement learning”, “deep reinforcement learning”, “predictive maintenance”, and “O&M optimisation”), (iii) the search window (up to 2025), and (iv) clear inclusion and exclusion criteria. We also report that this protocol yielded 54 papers after full-text screening, and we clarify that “deliberately narrowed” refers to focusing on studies with explicit DRL formulations for maintenance decision-making while excluding adjacent O&M domains unless directly informing maintenance planning. These additions improve reproducibility and reduce selection bias, in line with PRISMA-style expectations for transparent screening and eligibility reporting.

Response to comment 2

We agree that performance comparisons benefit from consistent contextualisation. To address this, we updated the manuscript to directly relate reported benefits to the modelling choices and DRL methods used, supporting clearer cross-study comparison and helping practitioners map approaches to expected outcomes. Where studies did not report explicit quantitative improvements or where baseline/metric definitions were not comparable, we did not infer values.

Response to comment 3

We agree and have expanded Section 4.2 to clarify both why offshore wind maintenance planning problems are often more naturally modelled as POMDPs and how DRL implementations address partial observability in practice. We now explain that decision-makers typically do not observe the true system state, but rather noisy, delayed, and incomplete measurements. We added a structured discussion outlining three practical

remedy classes in DRL: (i) history-based state augmentation (concatenated observation/action windows), (ii) recurrent DRL (LSTM/GRU) as an implicit belief surrogate (including DRQN-style formulations), and (iii) transformer-based memory to capture long-range dependencies. We additionally note that, beyond implicit memory, explicit belief estimation can be integrated through filtering methods when models are available, or via learned latent-state approaches (e.g., deep variational RL). This expansion links uncertainty sources to concrete modelling and policy-representation choices.

Response to comment 4

We thank the Reviewer for this constructive suggestion. We agree that the discussion benefits from a concrete wind-specific illustration of why binary “replace / not replace” actions can be overly restrictive, and how multi-level (imperfect) maintenance better reflects real decision-making. To address this, we added a dedicated paragraph in Section 9 providing a wind-turbine-specific example based on the gearbox literature. In particular, we now cite and summarise the study by Aafif et al. (2022), which evaluates multiple preventive maintenance levels (imperfect maintenance with partial restoration) and compares them against a replacement-only strategy. This makes the argument more concrete by linking multi-level actions to distinct restoration effects and cost trade-offs.

Response to comment 5

We agree and have added a dedicated paragraph explicitly discussing limitations of DRL for offshore wind maintenance planning. The revised text addresses: (i) limited interpretability of deep neural policies (“black-box” behaviour), (ii) high data and computational requirements, and (iii) simulation-to-real transfer challenges, particularly under safety and operational constraints. The discussion is also reinforced in the section on real-world applicability by clarifying modelling assumptions and validation limitations.

Response to comment 6

We thank the Reviewer for this suggestion. To make the visual elements more self-contained, we revised the captions of Figures 5–6 and Table 2 (and, more generally, all figures and tables) to define acronyms, clarify the meaning of plotted distributions, and provide clearer interpretation guidance. We also ensured that the underlying set of reviewed studies supporting each classification can be traced via **Table 3**, which provides a consolidated mapping between categories and the referenced studies. Where distributions are reported, we explicitly indicate (or cross-reference) the corresponding counts to improve transparency.

Response to comment 7

We thank the Reviewer for noting this. We carefully checked the bibliography and removed duplicated entries from the reference list.

Response to comment 8

We thank the Reviewer for noting this. We standardised the notation and consistently use **POMDP** throughout the revised manuscript.

Response to comment 9

We thank the Reviewer for this helpful suggestion. We expanded Section 4.2 by explicitly acknowledging that, beyond SCADA/CM signals, offshore maintenance decision-making can also be informed by inspection- and SHM-driven observations derived from non-destructive evaluation (NDE) techniques. We added a short paragraph citing Civera and Surace (2022), which reviews NDE methods for wind-turbine condition and structural health monitoring and discusses practical deployment modalities (including robotic and UAV-based surveys). We use this reference to further motivate partial observability in practice and to strengthen the rationale for memory- and belief-based DRL policy representations.

Response to comment 10

We thank the Reviewer for noting this. The manuscript has been checked for typographical issues and corrected where identified (including terminology issues such as “convective” vs “convolutional”, where applicable).

Response to comment 11

We agree that explicitly stating threats to validity strengthens the rigour and interpretability of a review. Accordingly, we added a dedicated “Threats to validity” paragraph to the Conclusions, addressing: (i) publication bias toward positive simulation results and the underreporting of negative findings, (ii) reproducibility constraints due to proprietary data/simulators and inconsistent reporting of training setups (e.g., reward shaping, hyperparameters, random seeds), and (iii) limitations to generalisability because many studies rely on stylised environments and may not transfer directly to operational settings or other domains without validation. This complements the PRISMA-style protocol description by clarifying residual risks affecting evidence synthesis and comparability.

Response to Reviewer 2

General response

We thank the Reviewer for the careful reading of the manuscript and for the detailed, constructive comments. In the revised manuscript, we strengthened the positioning of the

review relative to prior literature, corrected internal cross-referencing and reference-list duplication, expanded the discussion of DRL limitations and real-world applicability, clarified the treatment of policy-gradient methods in discrete settings, improved the interpretation and scope framing of Figures 1 and 7, reorganised the single-agent DRL section using a structured algorithm-family overview plus a comparison table, and made statements more precise by adding quantitative values where available.

Response to comment 1

We agree and strengthened the Introduction by adding a concise synthesis of the most relevant prior review literature and clarifying how our review differs. Specifically, we now summarise: (i) reviews on predictive/prescriptive O&M and maintenance optimisation in offshore wind, (ii) reviews on machine-learning-based condition monitoring and prognostics, and (iii) broader surveys of reinforcement learning in wind energy and power systems. We explicitly highlight that, while these works cover data-driven O&M, prognostics, and RL applications in wind/power systems, none provides a dedicated synthesis of DRL for offshore wind-farm maintenance planning, nor do they systematically compare DRL architectures, modelling assumptions, and offshore-wind-relevant constraints. We then state our review's unique contribution and organising dimensions.

Response to comment 2

We thank the Reviewer for noting this. In the revised manuscript, cross-references to figures, tables, and citations now compile correctly and appear as clickable links in the PDF. We also carefully checked the bibliography and removed duplicated entries from the reference list.

Response to comment 3

Thank you for your feedback, we agree and added a dedicated paragraph explicitly discussing limitations of DRL for offshore wind maintenance planning. The revised text addresses: (i) the limited interpretability of deep neural policies, (ii) high data and computational requirements, and (iii) simulation-to-real transfer constraints under safety and operational requirements. These points are supported with appropriate references. We also reinforce real-world applicability limitations in Section 8.

Response to comment 4

We thank the Reviewer for highlighting the ambiguity. To address it, we revised the wording to avoid associating policy-gradient methods exclusively with continuous control. The revised text now states that policy-gradient methods can be advantageous in various decision-making settings, particularly those characterised by long horizons, stochastic

dynamics, sparse/delayed rewards, and partial observability. We also clarify that successful PPO applications exist in discrete maintenance-planning formulations, while avoiding any unsupported claim of inherent superiority over value-based methods in discrete settings.

Response to comment 5

We thank the Reviewer for noting that Figure 1 and Figure 7 required clearer interpretation and scope framing. We revised the manuscript to explicitly state that Figure 1 is intended to communicate the opportunities as mechanisms that can mitigate the challenges of offshore wind O&M, and we added an explanatory paragraph mapping each opportunity to the challenges it can address. This includes clarifying that proactive scheduling informed by operational data and forecasts is presented as an opportunity (consistent with the Reviewer's example). For Figure 7, we now explicitly frame the listed items (e.g., aerodynamics/wake knowledge) as domain knowledge inputs that inform O&M decision-making and modelling.

Response to comment 6

We agree and reorganised the relevant part of the manuscript. We introduced a comparison table (Table 2) summarising the strengths and limitations of the three main single-agent DRL algorithm families (value-based, policy-gradient, actor-critic) and revised the surrounding text to introduce and reference the table. This improves readability and enables rapid comparison of algorithm suitability for offshore wind O&M decision contexts.

Response to comment 7

We thank the reviewer for noting this. The typo was corrected, and the surrounding text was revised as part of the restructuring described in Comment 6.

Response to comment 8

We adopted the Reviewer's suggestion and rephrased the sentence to: "As wind farms scale up and the environment increases in size, ..." to improve clarity.

Response to comment 9

We thank the reviewer for the constructive feedback, we replaced the qualitative statement with a quantitative comparison. The revised text now reports the explicit cost values and the percentage reduction stated in the cited study, this way, quantitative improvements (including lifecycle cost reductions) are consolidated and become more comparable across studies alongside algorithmic and modelling characteristics.

Response to comment 10

We revised the paragraph to remove any implication of guaranteed performance improvement from adding domain specifics. The updated text explains that incorporating domain-specific information can improve realism and better align DRL formulations with offshore wind physics and operational constraints, but it may also increase model complexity and training burden, and conflicting inputs can impede learning. We therefore state that the contribution of domain knowledge is problem-dependent and contingent on the quality, relevance, and reliability of the information provided.

Response to comment 11

We fully agree and revised the paragraph to emphasise that reported gains should be interpreted within the constraints of the simulation environment and modelling assumptions. We explicitly note limitations related to simplified wake modelling (including Jensen-model limitations), fixed maintenance durations, deterministic task lists, and the absence of operational logistics constraints such as vessel/access restrictions. We also stress that more comprehensive validation with higher-fidelity wake and metocean representations, and realistic logistics, is needed before inferring field-level applicability.

Response to comment 12

We agree and revised the presentation of the DRL algorithm families using a structured bullet-point format, complemented by the added comparison table. This improves readability and helps readers quickly distinguish the main families and their typical use in offshore wind O&M formulations.