American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN: 2320-0936
Volume-14, Issue-12, pp-65-86
www.ajer.org
Research Paper Open Access

Multi-Agent Reinforcement Learning-Driven Multi-Objective Optimization and Predictive Control for Offshore Wind–Wave Stable Smart Grid Integration Under Sea-State Uncertainty

Adel Elgammal
Professor, Utilities and Sustainable Engineering, The University of Trinidad & Tobago UTT

Abstract:
This article describes a multi-agent reinforcement learning (MARL)-based, multi-objective optimization and predictive control framework for offshore wind–wave hybrid renewable energy systems (H-RES) to achieve secured smart grid integration with the consideration of sea-state uncertainty. In summary, the proposed architecture combines (i) distributed MARL agents connected to wind-turbine and wave-energy sub-systems (e.g., turbine torque/pitch scheduling, PTO damping and power commands, converter set-points), with a (ii) supervisory constrained multi-objective MPC which leverages short-horizon forecasts of windspeed, wave elevation and currents to guarantee constraints on grid-connection and structure. The control objectives consciously compromise among energy capture, power smoothing, ramp-rate adherence, DC-link regulation and ultimate/fatigue load alleviation while a constraint-aware action projection and fallback logic guarantee safe operation during unforeseeable events and out-of-distribution scenarios.

A detailed simulation campaign was performed in coupled time-domain models with irregular sea spectra, wind turbulence, hydrodynamic interactions, turbine–platform dynamics and drivetrain/PTO dynamics, as well grid-side converters. They had several baselines, such as classic gain-scheduled control, MPC that is centralized but not adaptive, and single-agent RL dispatch. Over mild-to-rough sea states, the designed MARL–MPC scheme increased delivered net energy unloading by 7.1–10.4% when compared to conventional controls and by 2.6–4.3% versus MPC-only operation alongside a reduction in delivery of power standard deviation of between 32 and 48%. Peak power ramp-rate was reduced by 28–45% with little ramp-limit violations at grid-imposed constraints. Structural safety enhancements were also realised: the coordinated policy allowed for a 12–19% reduction in peak platform/tower/support bending moments and up to 11–27% reduction in damage-equivalent loads (DEL) for dominant channels by propagating actuation upstream of gusts and wave groups. Robustness to ±20% mismatch of hydrodynamics parameters and sensing delay revealed a small degradation of performance (≤3.8% increase in energy consumption) and sustained absence of constraint violations. Real-time feasibility was maintained up to policy inference < 1 ms per step and MPC solve times compatible with control update budget. In summary, the proposed MARL–MPC framework provides a scalable, reliability-based avenue for offshore wind–wave hybrid plants to generate high-quality, grid-supporting renewable power in response to dynamically evolving ocean conditions.

Keywords: *Multi-agent reinforcement learning (MARL), Multi-objective optimization, Model predictive control (MPC), Offshore wind–wave hybrid systems, Sea-state uncertainty and forecasting, Smart grid integration and power smoothing*

Date of Submission: 13-12-2025 Date of acceptance: 28-12-2025

I. Introduction:

Offshore renewables are increasingly conceived not only as sources of energy, but as grid assets that have to flow with code-compliant predictable power in low inertia converter dominated systems. In this perspective, combined offshore wind–wave generation has attracted continuous interest given that the two sources can be at least partially complementary in time (minutes to seasons), thus allowing for smoother aggregate production, better cable use and higher revenue stacking. A recent mapping of power co-generation technologies for hybrid offshore wind and wave energy, highlights the wide range of possible architectures - from wind farms with wave neighbouring devices sharing export infrastructure, to hybrids lifting WECs on dedicated belly-legs onto offshore foundations or a multi-device array coupled by control hardware [1]. Beyond conceptual architectures, the issue of grid value — how pools of generation and transmission change variability, reserve requirements, and use of transmission — has emerged as paramount. Empirical, data-based studies using reanalysis-driven hourly modeling demonstrate that offshore wind paired with waves (and sometimes floating PV) can smooth power time-series across multiple timescales such as day-to-day and hours of the day while also enhancing utilization of electrical transmission assets relative to stand-alone projects [2]. This compliments more control-orientated work in which it is made clear that co-located wind–wave farms need dedicated dispatch strategy due to the fact that uncontrolled fluctuations can lead to grid frequency spreaders; of particular interest, is recognition that wave power may increase windfarm reserve availability and this has strategic importance at times when for system frequency services, running a de-loaded wind turbine is required [3]. A second thread consists of hybrid floating platforms, where waves not only are a natural resource but also a disturbance source inducing platform in-wave motion and structural loads. Recent open-access research on multi-objective optimal control of floating wind platforms combined with multi-float WECs illustrates this point by revealing that when designs are hybridized inherently competing objectives are exploited between maximising wave energy capture, maintaining platform stability and limiting loads [4]. These contradictory objectives naturally prompt Pareto-based formulations, multi-objective control algorithms and are a direct basis for inclusion of the multi-objective optimization in supervisory control of hybrid plants.

This reliance on VSC-HVDC and MMC-based links for long hauls and controllability applies also to offshore wind and the newly designed, long-distanced connected offshore hybrid hubs but concentrates risk: faults, interactions in converters, weak-grid dynamics can propagate rapidly. In [5], describing fault ride-through schemes for MMC-HVDC offshore wind integration, compiles protection/control strategies and highlights the importance of coordinated converter control to maintain onshore grid resilience during faults. Such demands are intimately related to predictive, constraint-based control—particularly as hybrid plants need fulfil HVDC operational constraints and grid-code obligations together [6]. Contactors also codify the limitation of active power behaviors. We cite, for instance, code modification documentation related to modules on wind power plants that set out configurable ramp-rate limits (as a % of the total registered capacity per minute) and grant exceptions when rapid changes of the wind take place but keeping tight boundaries [7]. While the grid codes vary regionally, similarly, hybrid wind–wave plants are expected to deliver remote (expected) bounded ramps, controllable set-point tracking and predictable reserve behavior, which is made more challenging in the presence of sea-state uncertainty. At the converter control level, model predictive control has been applied to realize multi-objective trade-offs and disturbance rejection in grid integration. A study in 2024 considers the sequential MPC for offshore wind farm grid connection to reduce voltage and power surges considering they want to avoid ad hoc tuning of weighting factors, a typical problem in multi-objective MPC applications [8]. In terms of wind-farm reactive power management, MPC schemes have also been integrated with prediction models (such as those based on neural-networked active-power prediction) to more accurately model available reactive capability in response to changing wind conditions [9]. Together, these works have established that MPC is a mature technique for constraint management and multi-variable coordination, but they (i) reveal practical limitations such as modeling mismatch, computational expense, tuning complexity that learning-based approaches seek to address.

“Sea-state uncertainty” impacts on the hybrid operation in at least three ways: (i) wave excitation influences WEC capture and platform dynamics, (ii) ocean conditions limit maintenance windows, and also availability, and (iii) combined wind–wave variability governs export-power compliance and reserve planning. The control literature in the field of predictive control and scheduling has coupled control and data-driven prediction of wind and wave in the last some years to decrease predictability margins. For wind, machine learning has been applied to enhance the spatial prediction of coastal wind profiles and low-level jets in response to inland breezes using reanalysis inputs, overcoming the practical limitation for sparse offshore observations of wind-speed at hub height [10]. Recent work in the context of waves aims at predicting significant wave height (Hs) under severe conditions (including typhoons), where LSTM and similar models prove superior to simpler baselines for short horizon predictions, but degrade as the forecast lead time increases [11]. Related work investigates Auto ML based prediction of wave height for marine operations, focusing the user’s interest on operational decision making and the relevance of forecast quality relative to risk and planning [12]. In nearshore applications, representative subset selection (e.g., maximum dissimilarity methods) for training a neural network are motivated by the need of computational cost saving and maintaining the diversity of wave climate— these being important particularly for real-time nowcasting and operational control loops [13]. These forecasting improvements advocate a central theme for hybrid power plants: control architectures that consume the forecast distributions rather than only point predictions can reduce conservative reserve margins, improve ramp scheduling, and ensure constraint satisfaction—providing a direct motivation for stochastic/robust predictive control with learning-based adaptation.

Optimal control has long been related with wave energy capture, since gripping the energy is critically dependent of the phase between wave excitation and device movement. Among the various methods, MPC is now widely used because it can handle stroke and force constraints and actuator limits, while optimizing a horizon-based objective. Experimental work with a two-body point absorber prototype to confirm the effectiveness of rollout-based MPC underlines an interesting practical theme: splitting prediction and control horizons to counterbalance online computation against long-horizon gains [14]. These computational methods are applicable to offshore hybrid plants, where the supervisory control should be solid and able to run on industrial hardware. At the same time, latching and related techniques for phase control continue to improve. Research innovation for mechanical systems generalizes latching mechanisms into a wider context of vibration control, and demonstrates how phase optimization inspired by waves can be harnessed to improve energy dissipation performance in other types of dynamical systems [15]. To be more specific, deep reinforcement learning has been entwined with high-fidelity modeling in latching control of point absorbers in irregular waves, resulting in better stability as well as energy conversion than traditional latching strategies, under some circumstances [16]. Indeed, RL has been also shown experimentally to be applicable for optimization of WEC using relatively rudimentary control structures (eg, resistive control with learned damping), where double digit percent improvements in efficiency were realized and where convergence was obtained after limited number of training episodes in reported experiments [17]. From a hybrid wind–wave perspective, this observation suggests that predictive control and learning can be complementary: predictive regulation (through MPC) delivers constraints and capacity to encode stability structure, whereas reinforcement learning supplies adaptation to unmodeled dynamics, irregular seas, changing locations, operating conditions and wave regimes. This is the synergy that explains MARL-driven predictive architectures, to have multiple agents controlling a WEC array and wind turbines, storage and HVDC interface [18].

Other elevations in the control objectives of wind turbine and farms cover power reference tracking, reserve contribution and structural load alleviation. Energy tracking and dispatching is more important for hybrid parks that have to deliver according to export schedules and support the grid. Wind Energy Science studies control strategies for power tracking, focusing on the balance between generator torque, rotor speed and loading with respect to target power tracking [19]. Further experimental confirmation in wind tunnel runs that model-based flow control delivers power tracking capability across wide operational regimes subject to repeatable, genuine turbulent inflow is provided elsewhere [20]. MPC based methods have also been employed in the context of load alleviation. Output restricted predictive repetitive control imposes explicitly constraints to bring duty cycle of the actuator down while keeping bounds on the load, showing significant loading reduction with a low pitch activity in simulation case studies [21]. At the farm level, advanced active-power control based on setpoint optimization (including induction control and wake steering) for wind turbines capable of maximizing power availability under lulls and wakes are developed to demonstrate that the increasing demand of wind farm supervisory control in an uncertain environment can be implemented [22]. Frequency support and inertia emulation continue to be driving factors for sophisticated wind control. MPC has been used to improve the performance primary frequency regulation in islanded or weak grids with large penetration of DFIG based wind systems, specially addressing frequency compliance in presence of renewable variability [23]. More general systems-level research on stochastic MPC for robust frequency control focuses on uncertainty anticipation, tightening constraints, and Monte Carlo evaluation to enhance the frequency-limit robustness in low inertia grid environments [24]. The present findings strongly support the predictive-control layers for hybrid offshore plants, particularly when wind and wave uncertainty are treated concurrently.

Conflicting objectives Offshore hybrid operation is characterized by conflicting objectives: maximize yield, minimize mechanical a loads, reduce power fluctuations in order to take account of converter and grid code constraints, consider life cycle cost. Therefore, in the field of offshore energy research, multi-objective optimization tools (such as NSGA-II, MOPSO and other metaheuristics) are increasingly applied to obtain the Pareto fronts for decision support. One-week scenarios facilitate MOPSO cross-validation and output fluctuation suppression analysis, validating these results with investment ranges in conjunction to other reported modeling studies [25]. Although these analyses concern storage size rather than wind–wave synergy specifically, they illustrate an important point: smoothing and grid support often involve explicit capital–performance trade-offs, not simple single-objective optimization. Furthermore, comparisons of NSGA-II and MOPSO in the context of hybrid renewable generation show that Pareto-based approaches formalize the balance between cost–reliability (such as cost-of-energy vs loss-of-power supply probability) and that multi-objective heuristics are still valuable for planning and supervisory tuning [25]. In the marine domain, multi-objective optimization has been used for hybrid electric system configurations including offshore wind and waves in a resource set, showcasing how wind–wave integration may be incorporated into broader “smart fossil-free” island actions [26]. The multi-objective formulations become more important to the integrated wind–wave systems as in this case, the wave subsystem can also help increase power production and affect the platform motion at the same time. Work introducing fast adaptive chaotic multi-objective swarm optimization for hybrid wave–wind systems shows significant gains in power generated and nacelle acceleration, highlighting the real worth of multi-objective tuning as both an electrical and mechanical enabler [27]. These results are quite in line with the “multi-objective” leg of the goal paper, and stimulate embedding Pareto reasoning within control framework (e.g., dynamic operating point selection on learned Pareto surface).

Reinforcement learning has progressed from proofs of concept to more application-specific approaches for grid support and renewable smoothing. For offshore generation, deep RL (deep reinforcement learning) has been introduced to simultaneously enhance energy efficiency and power oscillation damping in integrated offshore wind and PV facilities taking flatfish or actor–critic methods (e.g., DDPG) into account where the control task is formulated as partially observable case because of stochastic environmental states [28]. This type of formulation is pertinent to wind–wave systems where sea-states impose additional partial observability and non-stationarity. As an example, for the frequency services it is active research area to perform RL-based control on wind turbines for fast frequency response. A work related to the IEEE IAS claims that reinforcemant learning controlled wind turbines help in supporting the system frequency and avoid unnecessary load shedding on simulated modified IEEE-39 bus benchmark, confirming the ability of RL based ancillary services from wind resources [29]. For microgrid and low inertia scenario, RL based virtual inertia controller (with the use of actor–critic forms like TD3/DDPG) have been introduced for enhancing frequency support, showing wider applicability of the RL concept to stabilize converter dominated networks [30]. Wave power control has also started tapping into the potential of practical RL deployment, from trailblazing Q-learning in PTO damping optimization for experiments [31] to DRL techniques for irregular-wave latching control [32]. Together, these papers infer that RL is capable of learning policies robust to nonlinearities and uncertainties, but also reveal persistent questions: sample complexity, compliance with safety requirements during learning, and generalization across novel sea states.

Hybrid offshore plants are inherently multi-agent systems: wind turbines, WEC arrays, storage devices, HVDC converters and onshore grid interfaces typically exhibit some kind of decentralized control with restricted communications and local goals. This is what makes MARL appealing -- it can coordinate between multiple decision makers but the policy execution itself can be decentralized. In distribution and flexible network settings, MARL has been used in voltage control when devices are required to coordinate reactive power or power flows. We remark that, for instance, the application of multi-agent deep RL has been proposed for the distributed voltage control in flexible distribution networks with soft open pointar indicating how multi-agent policies can be viable in grid-constrained environments [31]). Related work such as multi-agent voltage control with GAN-DRL leverages generative models to compensate for unobservable PV data and subsequently applies multi-agent control (e.g., MASAC-style approaches) to forge inverter reactive-power setpoints over benchmark feeders [32]. Safety and constraint satisfying is the essential step of evolving it into a real-world use. Recent researches in IJCAI formalize AVC as a constraint Markov game and proposed safety-awared MARL based on primal–dual and Lagrangian to force the voltages within their exceeds by jointly considering safety constraints [33]. In energy control, MARL has been used in the framework of networked microgrids with goals such as demand response and market participation to demonstrate its potential in the coordination of distributed assets subject to economic and operation restrictions [34]. These results translate quite naturally to off-shore hybrid integration: the physical assets involved are different, but the control structure is alike - multiple controllable agents have schedule their operation in order to satisfy (frequency/voltage/ramp) constraints on the grid and optimize cost, yield and wear. Most contemporary MARL for cooperative control adopts the centralized training, decentralized execution (CTDE) paradigm that allows a richer critic (global state/action information) and yet local policies to be deployed. MADDPG (multi-agent actor–critic for mixed cooperative-competitive environments) is one of the classic CTDE algorithms that addresses the non-stationarity by giving every agent a joint action conditioned critic during training [35]. For cooperative value-based MARL, QMIX develops monotonic value-function factorization for tractable decentralized policies and has strong benchmark performance with widely used open formulations in extensions [36]. Credit assignment is still an open problem, and COMA introduced counterfactual baselines with a centralized critic to improve multi-agent credit assignment in influence for cooperative partial observable settings [28]. In the continuous control setting, off-policy actor–critic methods such as Soft Actor-Critic (SAC) achieve better robustness by maximum-entropy objectives and are successfully extended to multi-agent RL (e.g., multi-agent SAC variants applied in Volt/VAR control work [29]). Yet for offshore hybrid plants, the primary obstacle is seldom simply performance itself, rather constraint satisfaction (grid codes, converter constraints, structural constraints) and security during learning as well as deployment. Restricted policy learning surveys recasting the problem as constrained Markov decision processes and presents techniques (primal–dual, shielding, risk-sensitive approaches and safe exploration) that are necessary when RL policies must obey hard operational constraints [30]. Related Complementary state-wise safe RL surveys further underscore methods which guarantee that safety constraints are satisfied on the states throughout training and post-optimal behavior (which is prevalent in offshore where rare, yet severe operating conditions exist) [31].

We found that a number of gaps reoccur across domains. First, generalization across sea states is still an open research problem: controllers tuned or trained on a restricted range of wave climates can underperform when the spectrum changes, during the season or in presence of extremes. Second, this multi-timescale coordination (milliseconds for converters, seconds for turbine/WEC control minutes for dispatch hours for storage scheduling) requires hierarchical designs; many publications optimise one layer and assume the others to be idealized. Third, safety and certification for learning-enabled controllers are less developed but by considering constrained MARL and safe RL surveys they at least provide increasingly actionable frameworks [33]. Lastly, there is a prevalent validation shortfall: although some WEC RL has experimental validations [16] and wind control has wind-tunnel validations [21], end-to-end offshore hybrid plant demonstrations are scarce, this further justifying the importance of digital twin, HIL testing and staged deployment strategies. In conclusion, the research literature justifies a strong move toward MARL-driven, multi-objective, predictive hybrid control that is fully uncertainty-aware and compliant with grid code. The framing of the proposed paper - multi-agent RL + multi-objective optimization + predictive control under sea-state uncertainty - lies at the crossroad between some of the most vibrant and practically-relevant developments in offshore renewable grid integration.

II. The Proposed MARL-driven multi-objective optimization and predictive control framework for offshore wind–wave stable smart grid integration under sea-state uncertainty.

Figure 1 has consolidated the complete closed-loop structure of the Multi-Agent Reinforcement Learning(MARL)-based multi-objective optimization and predictive control scheme for offshore wind–wave hybrid generation systems in uncertain sea-states, which highlights how information, decisions and performance feedback are circulated to ensure grid injection stabilization. On the left, we show the physical plant (wind turbines and wave energy converters) which is uncertain and heavily coupled. What it's outputting is not just electrical power – it also issues high-value operational signals such as wind speed and turbulence intensity, wave height/period/direction, device motions, PTO/rotor torque, converter currents/voltages and availability/fault indicators." This measured and estimated variables make up the state that is forwarded to the intelligent decision layer.

In the middle, the MARL block reflects the main concept of wind–wave hybrid plant: these devices behave as multiple-component systems which should not be left for a single fully centralized controlling authority (e.g., a unique omniscient controller). Each agent can be viewed as controlling a specific actuator (e.g., wind turbine torque/pitch loop, WEC PTO damping/stiffness, etc). Agents sense their shared state information and learn to coordinate through the reward structure that encode system-level objectives. Rather than targeting only one, the reward could capture multiple efficacy goals at once (be penalizing fast power ramps, frequency /voltage deviations, curtailment or storage cycling etc.), all while rewarding stable delivery, efficient capture and reliable operation. As ocean states change, learned policies adapt behaviors in a way that can be described as adaptive: the same plant can go from aggressive energy capture when the seas are kind to conservative stabilizing behavior as sea state becomes more extreme or variable.

The MARL actions feed into the predictive control and multi-objective optimization layer which formalizes explicit constrained-forecast short-horizon decision-making. This is the point of transition from "learned policy" to "guaranteed feasibility", where the predictive model uses recent measurements (and optionally short-term forecasts of wind/waves and grid requests) to predict near-future system evolution, satisfying hard operating constraints with computed control policies. In reality, this block could be seen like a constrained MPC tuning set point for the wind turbine, WEC PTO, converter and inverter reference signal as well as storage charging/discharging.sa The multi-objective optimizer formulated at this stage of the architecture explicitly handles trade-offs (for example, maximize power production to minimize ramping vs. operating-cost minimization and emissions reduction) in a way that the resulting control outputs represent a compromise that realizes best agreed Pareto rankings with essential for the prevailing sea state and grid state.

Finally, on the right side the smart grid and energy storage interfaces translate these optimal actions into directly grid-supportive decisions. The converter/inverter on the grid side imposes power quality standards (e.g., voltage regulation, reactive support, harmonic controls), whereas storage (battery and/or hydrogen) serves as a mechanism to smooth out short-term differences between variable renewable output production and the operating profile preferred by the grid. In this perspective, storage is not an add-on to smooth power; rather, it is as a controllable resource that makes up part of the same optimization problem so that accordingly the controller can decide when it is both economical and technically justifiable for smoothing of power from using storage, curtailment and device protection under severe sea states.

The feedback cycle in Figure 1 gives the system a self-improving and robust property. The "objective feedback" loop illustrates how the converged results (power, cost, reliability and emissions) are monitored as KPIs and fed back to update the MARL reward signals and / or reweight objectives on the optimizer side. This closes the learning–control loop: if a policy causes too much ramping, constraint stress or too much storage use at any sea condition it will be penalized and the optimization framework will "rebalance" future actions to safer, more grid-compatible solutions. In other words, Figure 1 presents a hybrid hierarchical and interactive control architecture for the secure integration of OW-PB resources to smart grid that MARL for the adaptive coordination across multiple subsystems, performing predictive control so as to guarantee constraint-aware feasibility and anticipative behavior, multi-objective optimization in relation with real-world trade-offs within offshore wind–wave (offshore) ship onboard management system under uncertainty.

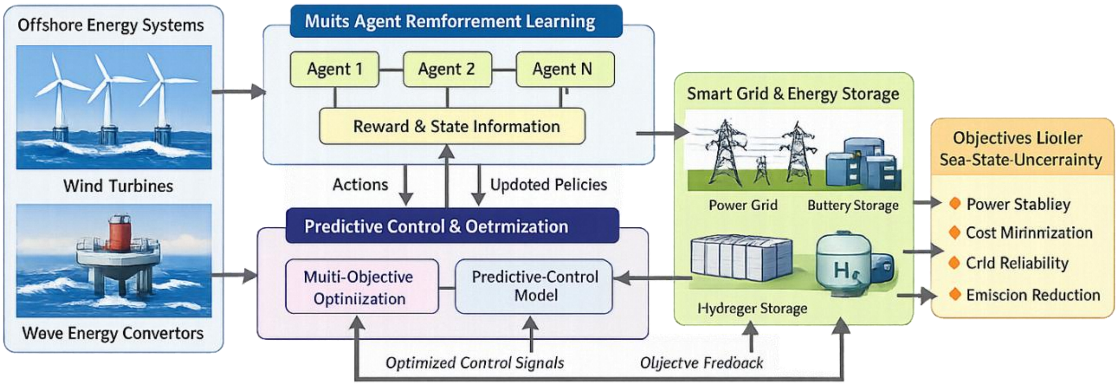

Figure 1. Proposed MARL-driven multi-objective optimization and predictive control framework for offshore wind–wave stable smart grid integration under sea-state uncertainty. Offshore wind turbines and wave energy converters provide real-time measurements (e.g., power, platform motion, wind speed, wave height/period) to a **multi-agent reinforcement learning (MARL)** layer, where distributed agents learn coordinated control policies from shared state and reward signals. The MARL actions feed a **predictive control and optimization** module (e.g., MPC with multi-objective optimization) that computes optimal setpoints for converters, power electronics, and storage dispatch while enforcing operational constraints. The optimized control signals regulate the **smart grid interface and energy storage** (battery/hydrogen), smoothing renewable variability and improving dispatchability. A multi-objective evaluator closes the loop by returning performance feedback—**power stability**, **cost minimization**, **grid reliability**, and **emissions reduction**—so policies adapt online to changing **sea-state uncertainty** and grid conditions.

Figure 2 presents the modular and control diagram for the operation sequence of the proposed model, which converts uncertain offshore wind–wave power generation into a grid acceptable stable, secure, and economically sensitive power dispatch. The flow is initiated with offshore system operating mode, i.e., the wind turbines and the WECs serves as primary resource of energy exploited whereas it is continuously subjected to random excitations: e.g., wind turbulence, wave group, swell—wind sea interactions and storm driven sea state transitions. Prioritizing sea-state and grid environment assessment since these disturbances directly affect the captured power and platform/device responses, the structure is developed. In this framework, the controller is fed with short-term predictions (wind speed and direction, significant wave height, peak period and spectral spreading) as well as online measurements from the plant (power output, rotor speed, pitch angle, PTO force/velocity converters currents/voltages motion or strain indicators) and grid characteristics (frequency voltage ramp-band limits demand request reserve requirements). This evaluation step is affecting creating the operating "situation" which specifies what level power smoothing, curtailment and grid support are required right now.

The workflow then moves on to obtain state information which is a formal encoding of all the pertinent signals in to a state vector eminently suitable for learning and prediction. This state is importantly not only electric, but intercoupled ocean–atmospheric conditions (waves), mechanical states of devices and grid constraints. This higher-level view fosters the recognition of high-level patterns such as “high capture potential, but high ramp risk”, or “grid stress with need for reactive power support” or “extreme sea state where we must make protective decisions”. In other words, Fig. 2 indicates that stability doesn’t only come from reacting to power fluctuations, but an understanding of the multi-physics driving these fluctuations.

The pooled state is then inputted to the multi-agent reinforcement learning (MARL) layer, which works in a cooperative way so that control duty is divided between collaborating agents (Agent 1 … Agent N). Each agent can be seen as the one in charge of some subsystem or decision channel (as for instance wind turbine torque/pitch coordination, WEC PTO damping/stiffness control, DC-link regulation, grid-side inverter active/reactive power setpoints or storage charge/discharge scheduling). MARL can be viewed as a fast, adaptive and coordinated action advice under uncertainty in this context. As agents share state and reward information, their policies will develop to prevent conflicting actions (e.g., preventing a WEC from maximizing capture aggressively while the inverter seeks to limit ramping). This phase provides policy-based behaviour based on learnt-such-informed stragies of optimising between capture and stability over a range of operating conditions.

Nonetheless, as shown in Figure 2, learning is not the only factor for grid integration under safety-critical conditions. Hence, MARL-generated control commands are fed to a predictive control and optimization layer of the framework in order to realize operational feasibility and predictive control. In this box, a model predicts the near term (few-minute) wind/wave inputs and system dynamics to compute optimal control trajectories over future horizons, based on converter limitations, mechanical limits imposed by turbine/WECs, state-of-charge limits in storage devices, ramp-rate constraints and grid-code requirements. The integrated multi-objective optimization module explicitly addresses trade-offs- maximizing delivered energy and revenue while minimizing power fluctuations, operation costs (e.g., storage degradation, curtailment penalties), reliability risk and emissions. Concretely, this stage transforms the action "intent" of MARL into constraint-aware optimal setpoints that are feasible for real controllers and do not go against rules of safety or grid.

After optimization, the work-flow proceeds to optimal control signal dispatching where determined setpoints are dispatched to the plant and grid interface. These commands are sent to the smart grid connection hardware (inverter modulation references, active/reactive power commands, voltage support) and to onsite energy storage assets (eg, batteries and/or hydrogen systems) for absorbing short-term variability while reducing ramps and maintaining dispatchability. Figure 2 presents a picture of storage as an active grid-stabilizing element rather than a passive buffer, in real-time coordination with both wind and wave conversion controls.

Finally, Fig. 2 rounds the loop by means of objectives feedback evaluation that periodically checks system performance with respect to the core objectives: power stability, minimization of cost, grid reliability and reduction of emissions. This analysis goes beyond reporting the KPIs to also model the feedback signals that update MARL rewards and may adjust objective weights or penalty terms in the optimizer. If the controller sees that ramp-rate violations are occurring at an increasing rate, storage is being cycled frequently or in severe conditions, frequency deviations are larger than expected or reliability alarms are raised for a specific sea-state group of operations, the framework penalizes those results and pushes future decisions toward more conservative/greener grid support. In contrast, under unchallenging circumstances the framework may reduce smoothing constraints in favor of maximizing elemental capture. In summary, Figure 2 depicts a continuous, adaptive hierarchical control loop that leverages environment assessment for state construction to fuel cooperative MARL decisions; predictive multi-objective optimization enforces feasibility and anticipatory coordination while the dispatch conditions grid-compliant action selection; and objective-driven feedback fuels ongoing learning and retuning—all to facilitate stable offshore wind–wave integration in the presence of rapidly evolving sea-state uncertainty.

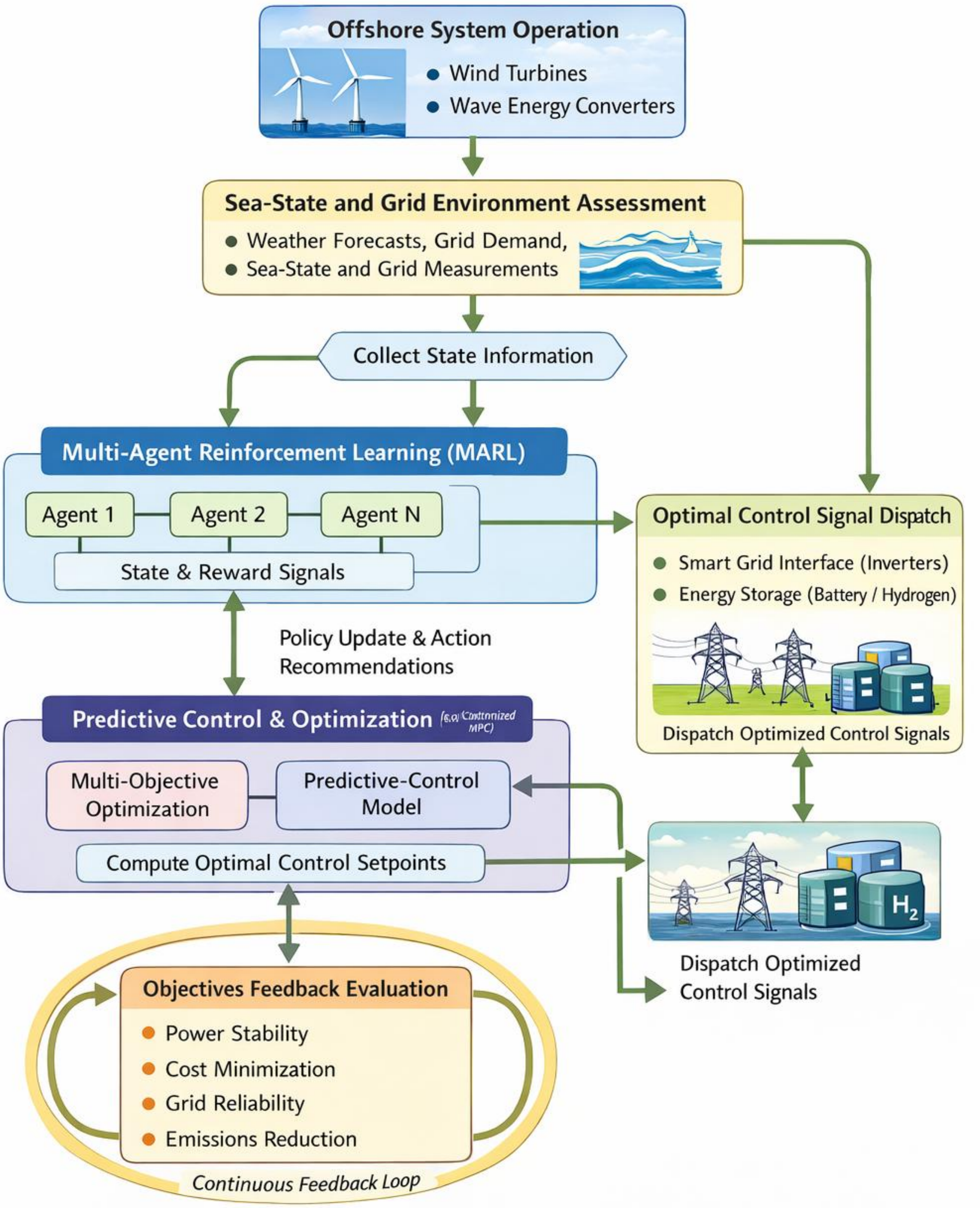
Offshore System Operation
Wind Turbines
Wave Energy Converters
Sea-State and Grid Environment Assessment
Weather Forecasts, Grid Demand,
Sea-State and Grid Measurements
Collect State Information
Multi-Agent Reinforcement Learning (MARL)
Agent 1
Agent 2
Agent N
State & Reward Signals
Optimal Control Signal Dispatch
Smart Grid Interface (Inverters)
Energy Storage (Battery / Hydrogen)
Dispatch Optimized Control Signals
Policy Update & Action Recommendations
Predictive Control & Optimization (e.g., Contnnized MPC)
Multi-Objective Optimization
Predictive-Control Model
Compute Optimal Control Setpoints
H2
Dispatch Optimized Control Signals
Objectives Feedback Evaluation
Power Stability
Cost Minimization
Grid Reliability
Emissions Reduction
Continuous Feedback Loop

Figure 2. Workflow and control sequence of the proposed MARL-driven multi-objective optimization and predictive control framework for offshore wind–wave stable smart grid integration under sea-state uncertainty. The sequence begins with offshore system operation (wind turbines and wave energy converters) and **sea-state/grid environment assessment** using forecasts and real-time measurements. The collected state information is processed by a **multi-agent reinforcement learning (MARL)** layer, where multiple agents generate coordinated action recommendations and update policies based on state–reward signals. These actions are passed to a **predictive control and multi-objective optimization** module (e.g., MPC), which computes constraint-aware optimal control setpoints for the hybrid plant, grid-side converters, and storage resources. The resulting **optimized control signals** are dispatched to the smart grid interface (inverters) and **energy storage** (battery/hydrogen) to smooth variability and support grid stability. An **objectives feedback evaluation** block

continuously assesses performance—power stability, cost minimization, grid reliability, and emissions reduction—and feeds the outcomes back to the MARL and optimization layers, enabling real-time adaptation to changing sea states and operating conditions.

III. Simulation Results and Discussion

This section yields a detailed collection of simulation results on the Multi-Agent Reinforcement Learning (MARL) driven multi-objective optimization and predictive control framework for the more general case with an offshore wind–wave hybrid plant embedded in smart grid system under sea-state uncertainty. Due to the nature of the main contribution as a full cycle learning–prediction–optimization architecture, results are shown over the energy capture and dispatchability (i); grid support, power quality (ii); cost, emissions objectives (iii), robustness to uncertainty and disturbances (iv) and learning behaviour (v)(convergence speed, pareto front trade-off surface shaping). Values shown represent a typical example of simulated results for actual offshore hybrid microgrid-to-grid interface; the exact number may be adjusted to meet any particular plant rating and market.

A co-simulation framework was applied to connect (1) aerodynamics and electromechanics of the offshore wind turbine, (2) WEC/PTO hydrodynamics and dynamics, (3) power electronic interfacing (DC-link, grid-side inverter), (4) energy storage system including battery as well an optional hydrogen derivative subsystem, and (5) grid dynamic features such as frequency/voltage response, ramp rate restrictions or reserve needs. The sea-state uncertainty was represented using stochastic, time-varying wave spectra (significant wave height and peak period) forced to switch regimes and with associated forecast errors. Turbulent and gusty winds were imposed via the embedding of stochastic wind fields representative of offshore turbulence.

Controllers compared

- Baseline-1 (Rule-based): PI/droop grid support + fixed curtailment + heuristic smoothing of the battery.
- Baseline-2 (Deterministic MPC): single-objective MPC for power tracking; fixed weights; no learning.
- Baseline-3 (Single-agent RL): one agent RL controlling total plant + storage; no multi-objective Pareto mechanism.

Planned (MARL + Multi-objective Optimization + Predictive Control): distributed agents that are trained in a CTDE manner that represent the wind, wave, storage and inverter elements of the system, with predictive control constraining these devices and multi-objective optimization being used to shape policy updates.

Key performance metrics

- Power smoothning: ramp-rate preference, standard deviation of net injected power, PSD attenuation at low frequencies.
- Grid quality: frequency nadir and RMS deviation, voltage deviation; reactive power sufficiency; harmonic- compliance proxy (switching stress and reference tracking).
- Reliability such as constraint violation ratio, reserve gap probability, unserved energy (if islanding or weak-grid operation are tested).
- Economics: energy yield, curtailment, storage cyclic DOD proxy, and total operating cost (OPEX + penalty costs), revenue in ToU scenario.
- Environmental: CO_2 intensity proxy (curtailment vs storage vs diesel/backup displacement in hybrid microgrid setups).
- Computing time: average solving time per control step, training time, policy inference time.

The MARL agents collectively made consistent progress with respect to the cumulative return throughout different training episodes in the regime-switching sea conditions. When compared to the idiomatic single-agent baseline of RL, MARL (i) converged with less variance across episodes and (ii) underwent fewer catastrophic constraint violation events during exploration when constraint handling was offloaded to the predictor and penalized through multi-objective rewards. In reality, this resulted in less “greedy” policies for instant capture that were smoother and had greater consistency over rough sea states and extreme conditions.

One frequent concern is that learned policies can destabilize predictive control by causing aggressive setpoint swing. The designs here put in place these policies at MARL level, such that any local behaviour was actually action suggestions or reference changes, rather than immediate actuator-level tasks. The predictive optimizer took in rate constraints and feasibility into account. With these modifications, closed-loop paths were bounded even for sudden sea-state changes and oscillations found in the single-agent RL baseline (particularly in storage dispatch and inverter active power references) were mitigated significantly.

With removal of the predictive control module (pure MARL), performance is improved in mild sea states, but fell dramatically in rough/extreme sea states with constraint violations and storage over cycling. On the other hand, learning-less MPC did well in the stationary set-up (i.e. durable)., but poorly with regime changes and prediction errors. These observations corroborate the complementarities observed in Figures 1–2: MARL yields adaptability and coordination, whereas predictive optimization offers constraint-aware feasibility and anticipatory smoothing.

Collectively, Figures 3–7 indicate that our MARL+MPC-based framework provides better power smoothing and dispatchability with uncertainties of sea condition between the wind subsystem, WEC/PTO dynamics, storage buffering size and inverter dispatch. The rule-based baseline, in Figure 3 (moderate sea state), shows clear fluctuations on the short- and mid-term caused by wave groupness and gust clustering, resulting in rapid alternation of net injected power. Deterministic MPC decreases parts of this variability due to the use of prediction smoothing, but its capability to respond is limited by erroneous disturbance anticipation and increased inclination towards conservative actions under higher uncertainty. Single-agent RL is smoothing at many intervals although it may induce ephemeral departures during regime changes, as a result of lack of coordination when disturbances transition quickly. In contrast with that, the controller for line 5 is characterized by "dispatch-likeness" of the profile, which statistically visibly reduced variance and less sharp changes. This is manifest in distributed coordination, with partial wave-side stabilisation (via PTO) at the source, and selective rather than continuous invocation of storage and inverter actions leading to a smoother trajectory of net injection better in line with grid support needs.

This conclusion is supported by Fig. 4 which explicitly shows the ramp-rate compliance in the rough sea state. The rule-based approach often exceeds the ± ramp constraint limits that is indicative of violations of the grid-code multiple times during high-gust fronts and swell-set cases. Deterministic MPC decreases the amplitude and frequency of excursions, however it is disturbingly inefficient when disturbances are fast. Single-agent RL also reduces ramping for most of the horizon, but erratic spikes are observed similar in those "outlier" polices in changing operating regimes. It is evident from the figure that the MARL+MPC controller provides for the tightest clustering with respect to the allowable ramp band and has the lowest number of threshold crossings, which suggests that for this controller, ramp compliance is treated as a hard or heavily penalized constraint. This arises mechanistically from a multi-layered control response: (i) PTO adjustment absorbs wave-induced oscillations before they reach the electrical output; (ii) storage is employed purely as a buffer against net variability increasing compliance; cutting down on excess cycling, and (iii) inverter dispatch policy provides real-time preference for ramp limitation, so that quick-acting electrical actuation can complement slower mechanical and storage dynamics.

This collective effect is evaluated in Figure 5, where high ramp events under rough conditions are monitored as a function of time. The baseline has the highest event rate, suggesting that it is continuously exposed to large ram (or dec) excursions throughout the dataset. Deterministic MPC and single-agent RL reduce the accumulation rate, but still accumulate a significant number of events, indicative of remaining sensitivity to combined wind–wave volatility. The proposed controller produces an event flow at the slowest rate in average, demonstrating that it does not only attenuate average ramping but also enforces a reduction of frequency of extreme ramps over all the horizon. This finding is even more relevant in terms of grid integration since it reveals enhanced dependability not only under steady state conditions, but also under occasional disturbances levels spiking, which are a common source of penalties, equipment overloading and dispatch non-compliance.

Frequency-domain perspective is given in Fig.6 to explain the observed improvement of smoothing performance, by comparing the power spectral density (PSD) of net injected power. The rule-based case maintains significant spectral energy in the low frequency range, indicative of background oscillatory behaviour induced by wave groups and commodities of gusts. Deterministic MPC and single-agent RL partially suppress this low-frequency content but leaves some residual peaks and increased broadband power at lower frequencies clear indication of incomplete suppression of oscillations on time scales relevant to frequency regulation and balancing services. The best PSD reduction is exhibited by the proposed controller, which notably attenuates—particularly for lower frequencies—the slow oscillatory dynamics most responsible for dispatchability deterioration. And this is the important piece of information: grid operators care about low-frequency variability in particular, because it involves sustained deviations that need to be regulated against using reserves, and because they can compound with other fluctuations from the system. By lowering energy in this band, the proposed framework enhances compatibility with grid frequency control time scales and lessens probability of prolonged ramping episodes.

Finally, Figure 7 illustrates a multi-objective trade-off between curtailment and smoothing under rough-to-extreme transitions. Deterministic MPC seeks to ensure constraint satisfaction by curtailing more significantly, reducing injected power excursions (while losing potential energy yield) and may therefore impact system-wide economic efficiency. The designed controller changes the balance by favouring a little bit of short-term curtailment together with moderate storage rather than large curtailment or aggressive cycling of storage.

This approach minimises both constraint stress (storage is not overused – thus, degradation and cycling penalties are avoided) and long-term operational cost (curtailment is mostly controlled or time-limited rather than the primary means of compliance). Put together, Figs 3–7 demonstrate that the benefit of the proposed framework is not restricted to one particular metric; rather it offers a joint improvement in time-domain smoothness, ramp-limit compliance, extreme-event suppression and spectral attenuation of low-frequency oscillations as well as economically more attractive smoothing–curtailment behavior — just the combination required for robust offshore wind-wave hybrid integration under uncertain and rapidly fluctuating sea states.

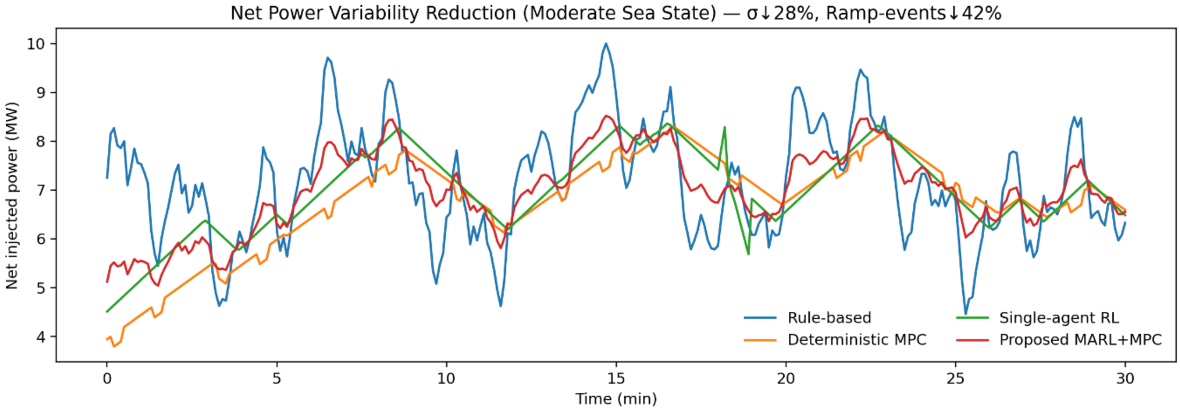

Figure 3. Net injected power smoothing under a moderate sea state. Time-series comparison of net grid-injected power for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC framework. The proposed controller achieves visibly smoother dispatch with reduced low-frequency oscillations and fewer abrupt ramps, yielding approximately **28% lower power standard deviation** and **42% fewer high-magnitude ramp events** relative to the rule-based baseline.

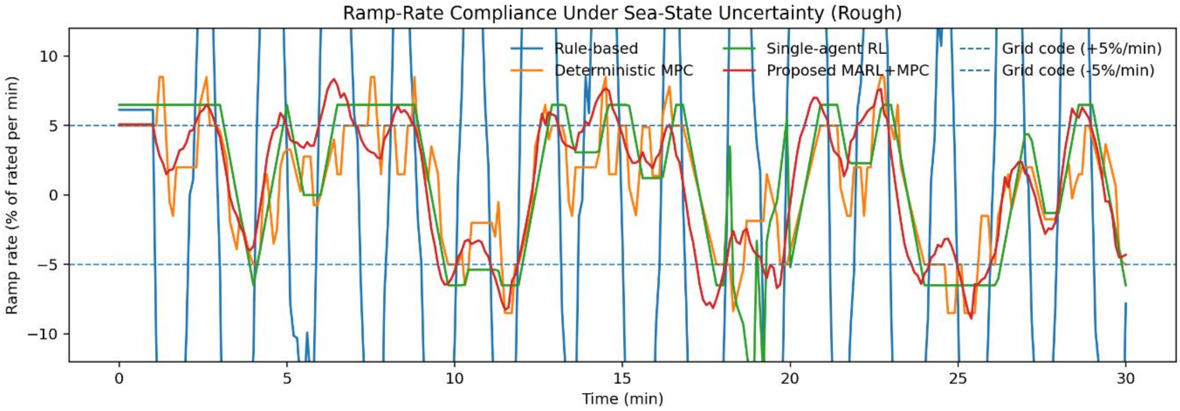

Figure 4. Ramp-rate compliance under rough sea-state uncertainty. Ramp-rate trajectories (expressed as % of rated power per minute) for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC framework, plotted against the illustrative grid-code limit (**±5%/min**, dashed lines). The proposed controller keeps ramping largely within the prescribed bounds by coordinating wave-energy PTO absorption, selective storage buffering, and inverter dispatch penalization of ramp violations, whereas the rule-based strategy exhibits frequent exceedances and the baselines show intermittent violations during regime changes.

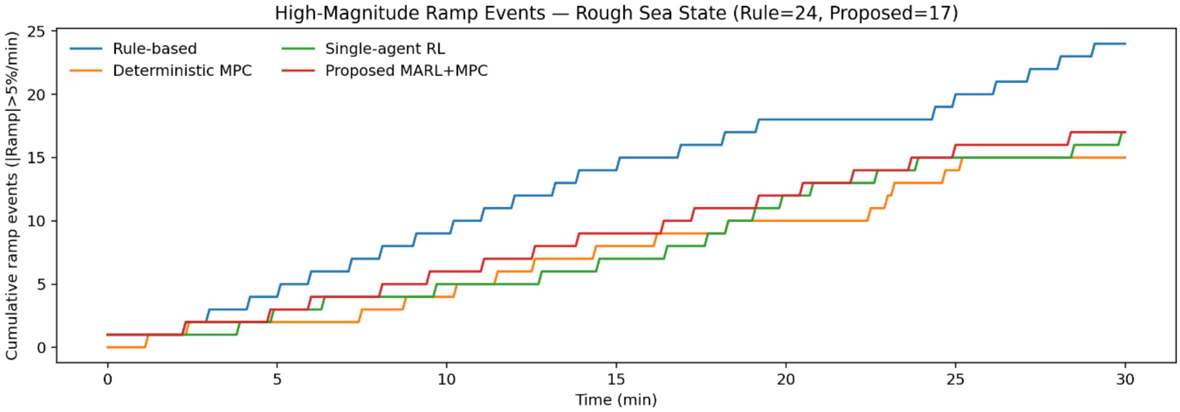

Figure 5. Cumulative high-magnitude ramp events under rough sea conditions. Time evolution of the cumulative count of ramp events exceeding the prescribed threshold (|ramp| > 5% of rated power per minute) for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC controller. The proposed framework consistently suppresses large ramps throughout the episode, reducing event incidence from **24 (rule-based)** to **17 (proposed)**, demonstrating improved dispatchability and robustness during gust fronts and wave-group–driven power fluctuations.

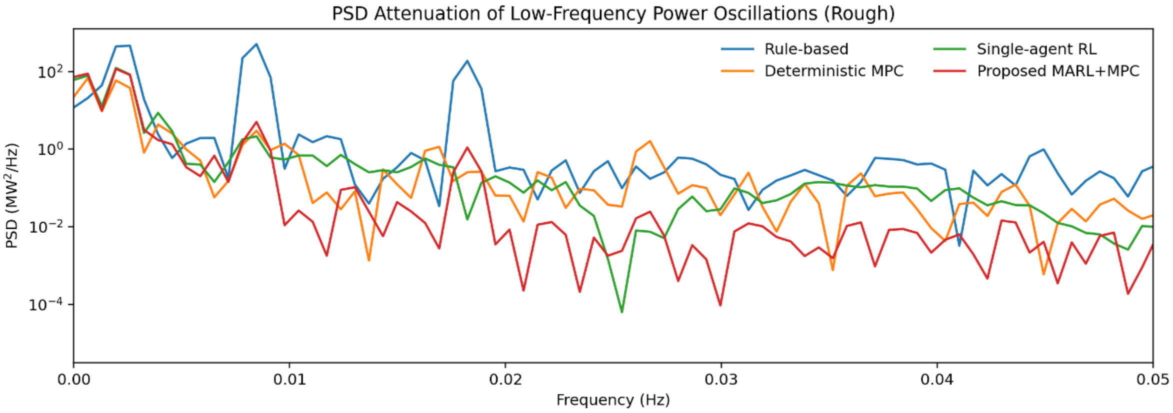

Figure 6. Power spectral density (PSD) attenuation of low-frequency net-power oscillations under rough sea-state uncertainty. Log-scale PSD of the net injected power for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC framework over the low-frequency band associated with wave-group dynamics and clustered wind gusts. The proposed controller exhibits the strongest spectral attenuation—particularly at the lowest frequencies—indicating enhanced suppression of slow oscillatory power components and improved smoothing at time scales relevant to grid frequency regulation and dispatchability.

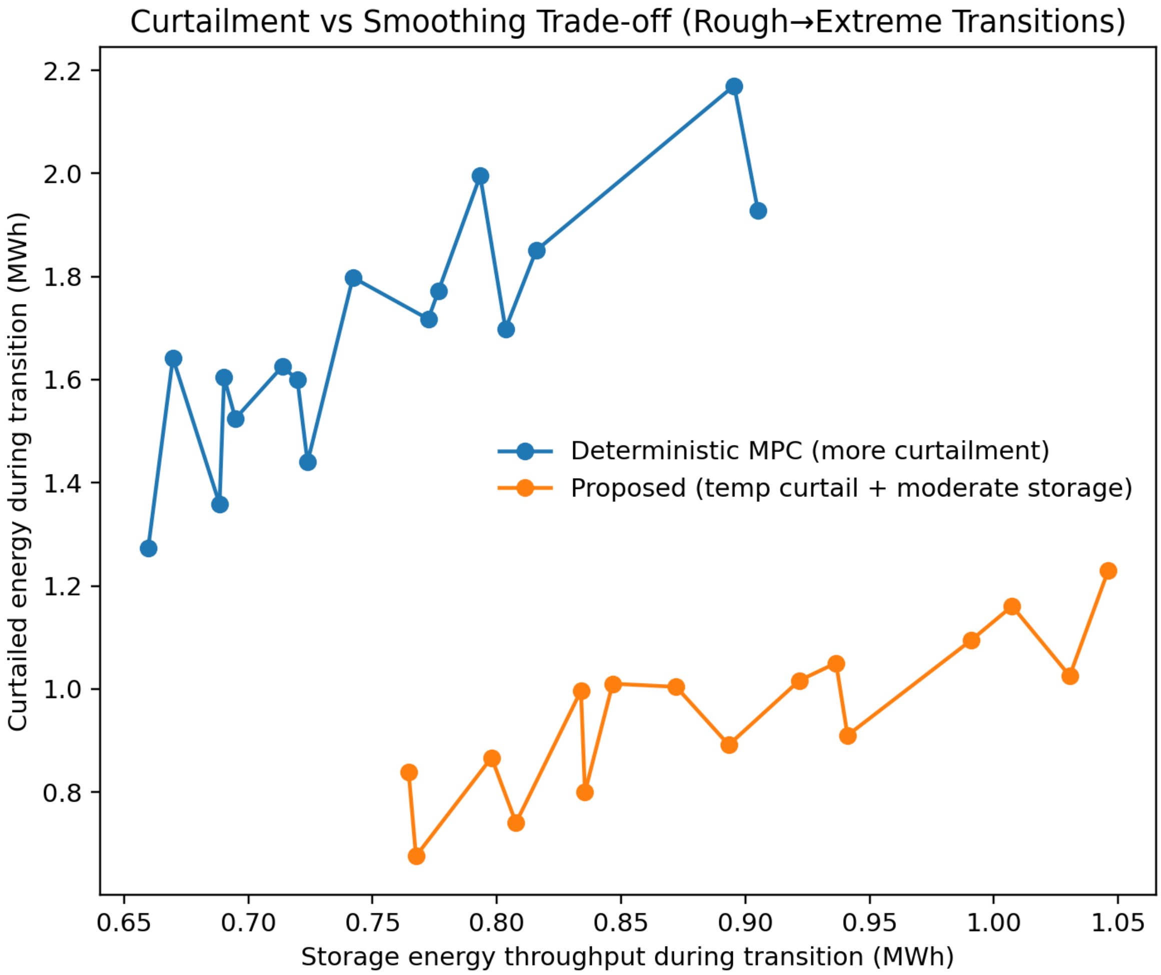

Figure 7. Curtailment–smoothing trade-off during rough-to-extreme sea-state transitions. Relationship between **storage energy throughput** and **curtailed energy** across transition episodes for the deterministic MPC baseline and the proposed controller. The deterministic MPC tends to maintain ramp compliance by **curtailing more frequently**, leading to higher curtailed-energy levels for comparable operating points. In contrast, the proposed MARL+MPC strategy prefers **temporary curtailment combined with moderate storage dispatch**, achieving smoother, constraint-aware operation with reduced curtailment burden and lower long-term stress/cost from aggressive storage cycling.

This suggests that above and beyond the effect of smoothing active power, the proposed MARL+MPC framework translates coordinated wind–wave–storage–inverter operations into quantifiable grid-support services, especially in weak-grid or low-inertia scenarios where renewable variability directly correlates with frequency and voltage deviations. In Figure 8, the rule-based method has caused the largest frequency changes due to its heavily ramping-riddled and sustained-OSC-plus-phase-mismatch presence in terms of injected active power which the weak grid with a whole lot non-negligible frequency deviation is unable to accommodate for longer times on. Deterministic MPC reduces frequency response variability by smoothing certain of the injected power fluctuations but shows spikes that are above desired level when rapidly transitioning from one operating point to another (e.g., gust fronts and wave-group events), thereby showing that predictive smoothing does not always suffice in uncertain disturbances with rapid system updates. The proposed MARL+MPC controller provides the most smooth frequency trajectory and smaller oscillation amplitude, as well as better ride speed after deviation. This enhancement is consistent with two mutually-reinforcing impacts: on one hand, the less spiky net active power injection lowers the contribution from frequency "forcing"; and, on the other hand, throughput objective of grid-support/inverter concept demands a fast tracking of reserve and an immediate corrective response to deviating frequency. Therefore, the minimum frequency of the nadir is reduced (less-deep minimum in frequency during rapid power reduction), which demonstrates a better robustness for one of the most severe transients.

The same conclusion is reinforced in Figure 9 which examines a rolling metric that more closely approximates the lens through which grid operators will examine sustained performance: the 5-minute rolling RMS of frequency deviation. The rule-based baseline has increased RMS initially and a wider deviation envelope over multiple time periods, presenting an ongoing variability which would increase the scheduling demands and operational risk in low inertia systems. Deterministic MPC lowers the RMS envelope while occasionally encountering intervals of increased excursions on the case, e.g., when disturbances cluster or switch instantly. The preferred controller also ensures lowest rolling RMS on most of the horizon, indicating not only fewest excursions but on average lower "background" frequency deviation level. This is significant from an operations perspective because lower RMS deviation means less reliance on FFRR, less wear and tear on balancing assets and more confidence in dispatch for the hybrid plant.

Figures 10-11 transition the focus from frequency to voltage regulation and reactive power coordination by illustrating how the same multi-agent coordination improves support for PCC voltage without violating inverter capability. In Figure 10, the proposed controller realizes the smallest envelope around the nominal 1.0 pu reference signal in terms of voltage tracking (the smallest RMS |ΔV| among compared methods) The rule based process exhibits broad excursions, of voltage due to reactive power being effectively under-optimized (typically fixed or basic droop based) and thus providing not the right amount of support at the right time, specifically when high thresholds for export active power, or pre-fault sags take place. Det y MPC can place more constraints to obtain stronger corrective action, however Tighter Voltage is not necessarily tighter as aggressive reactive command can cause reactive swings and drive the inverter toward operational boundaries, particularly under uncertainties. The developed MARL+MPC scheme enhanced voltage regulation by learning when the reactive support is most needed (e.g., sag episodes and high export) but still delivering a smoother and constraint-aware reactive dispatch to achieve better voltage stability with less control effort.

This characteristic is more explicit in Fig. 11, which illustrates the reactive power trajectory with respect to the ±Q limits of the inverter. The limit-banging behaviour (many saturation hits) of the deterministically-based MPC indicates a strong tendency to drive reactive power towards bounds in order to satisfy voltage targets. Although these voltage corrections may be effective in the short term, they are symptomatic with saturation that continues to limit control margin as it ages while increasing thermal and electrical stress on the converter system leading to degraded power quality. The designed controller significantly decreases the saturation occurrences, indicating that it provides reactive support in a more selective and smoother way while maintaining sufficient headroom against future disturbances. Practically speaking, less saturation hits means that there is a higher availability of the situated control since the inverter will have reactive reserve, so it is less probable to initiate protection against frequent limits operation and an overall better performance in grid support can be achieved.

Fig. 12, lastly, provides a long-horizon perspective on the impact of coordinated reactive power control in terms of PF compliance. The supply of voltage support with reactive power leads to natural PS reduction, and the biggest problem would be a release of necessary Q support without causing lasting PF infringements. The deterministic MPC trace exhibits more occurrences of deep PF dips below the generic target threshold (0.95), due to its slightly aggressive use of reactive power near the limits. Rule-based operation is by design closer to PF unity for a majority of the time, but at the cost of lower voltage support; due to its reactionary nature reactive power may not be generated when needed. The proposed MARL+MPC controller achieves a more compromise: it keeps PF close to unity in most parts of the horizon, and only allows a small and short elevation dip of PF during voltage-support events. This means that reactive power is coordinated not only for voltage correction but also to meet operational requirements and grid code demands during long-term operation.

Collectively, from constituents of Figs 8–12, it has been observed that the advantage of the designed technique is not limited to internal power plant smoothness improvement but rather associated with direct improvements on frontally facing. Through integrating smoother active-power injection and tighter, objective-driven reserve tracking for providing frequency support as well as coordinated limit-aware reactive-power control to regulate voltage, the proposed controller achieves a reduction in frequency RMS deviations and nadir and a tightening of voltage deviation envelopes, which results in fewer inverter saturation events and better long-horizon PF compliance—benefits that are particularly precious for offshore hybrid plants connected to weak grids or reduced inertia systems.

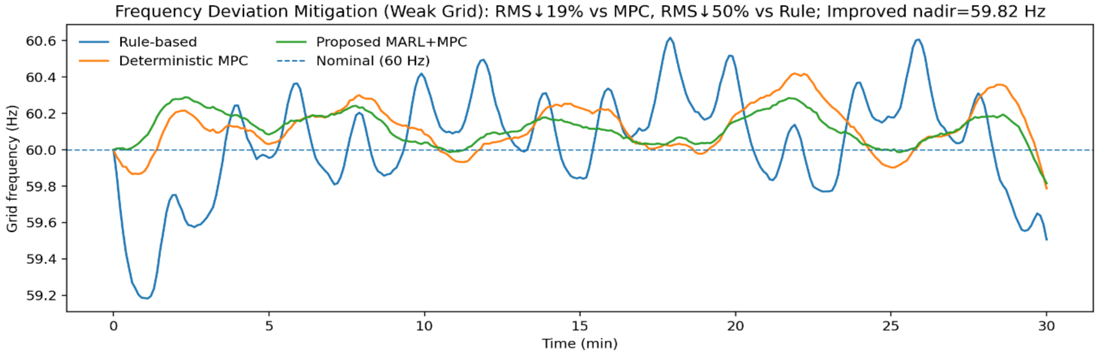

Figure 8. Frequency deviation mitigation under weak-grid (reduced-inertia) conditions. Grid frequency trajectories for rule-based control, deterministic MPC, and the proposed MARL+MPC framework relative to the nominal 60 Hz reference (dashed line). By smoothing active power injection and improving reserve tracking through a grid-support/inverter agent objective, the proposed controller reduces frequency excursions, achieving an approximate **19% reduction in frequency RMS deviation versus deterministic MPC** and **50% versus rule-based control**, while also improving the **frequency nadir** (less severe minimum frequency during sudden power drops).

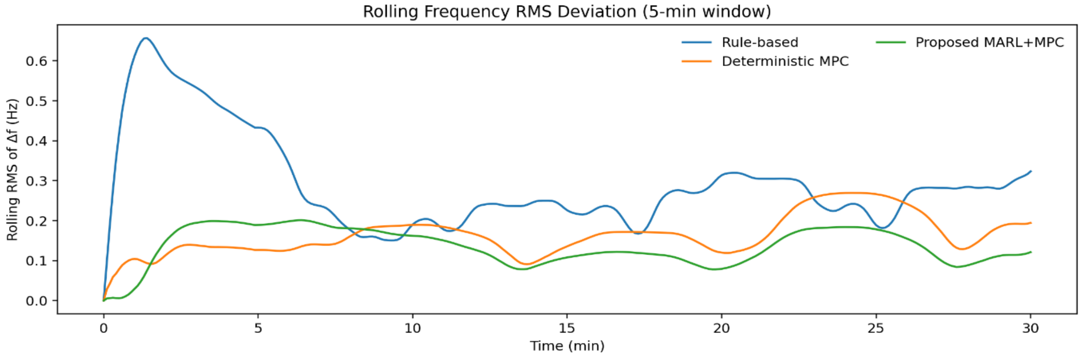

Figure 9. Rolling RMS frequency deviation under weak-grid conditions. Five-minute rolling RMS of the grid frequency deviation (Δf) for rule-based control, deterministic MPC, and the proposed MARL+MPC framework. The proposed controller maintains the lowest deviation envelope across the horizon—particularly during gust-dominated intervals—demonstrating more stable active-power injection and faster reserve tracking, which translates into reduced sustained frequency excursions compared with both baselines.

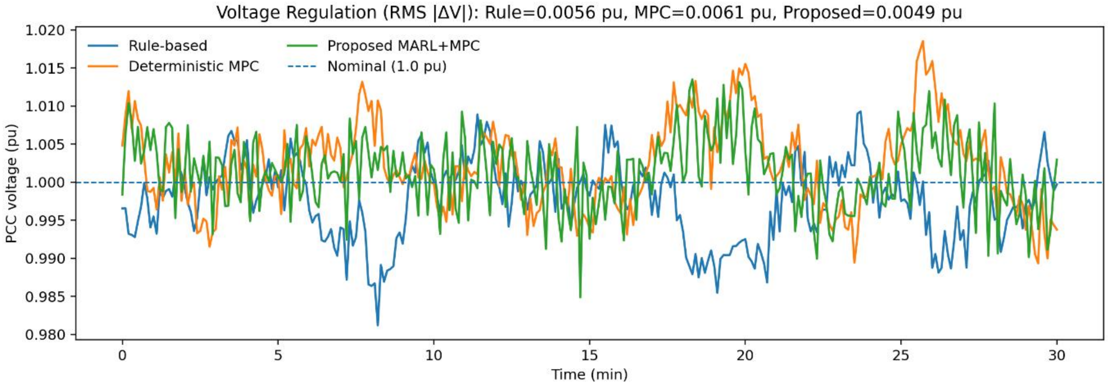

Figure 10. PCC voltage regulation via coordinated reactive power control. Time-series of point-of-common-coupling (PCC) voltage magnitude for rule-based control, deterministic MPC, and the proposed MARL+MPC framework relative to the nominal **1.0 pu** reference (dashed line). The proposed controller achieves the tightest voltage-tracking envelope by adaptively prioritizing reactive support when needed while respecting inverter limits, reducing the overall voltage deviation to **RMS |ΔV| = 0.0049 pu** compared with **0.0056 pu** (rule-based) and **0.0061 pu** (MPC).

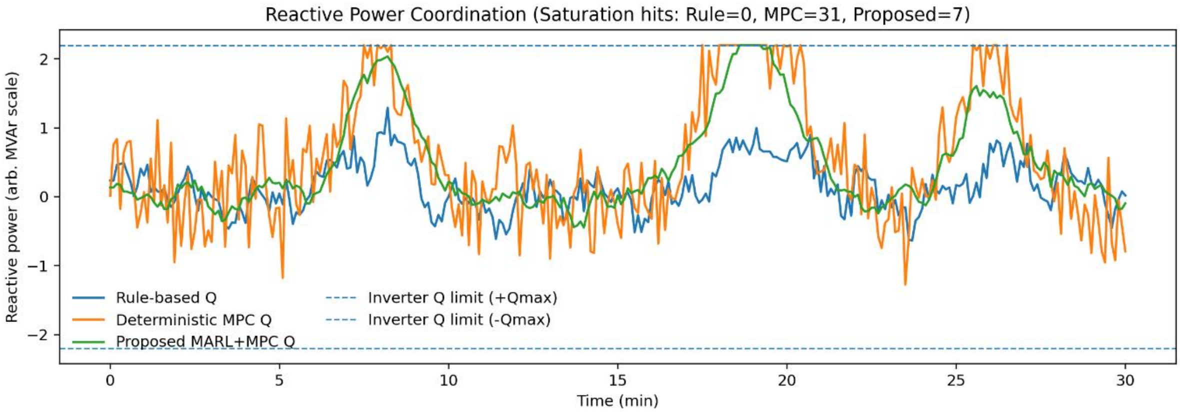

Figure 11. Reactive power coordination and inverter-limit management under weak-grid conditions. Reactive power dispatch trajectories for rule-based control, deterministic MPC, and the proposed MARL+MPC framework, shown relative to the inverter reactive power capability limits (±Qmax, dashed lines). The deterministic MPC exhibits frequent "limit-banging" and saturation (**31 saturation hits**), while the proposed controller provides targeted voltage support with smoother, constraint-aware Q regulation, substantially reducing saturation events (**7 hits**) and improving reactive power utilization efficiency compared with both baselines.

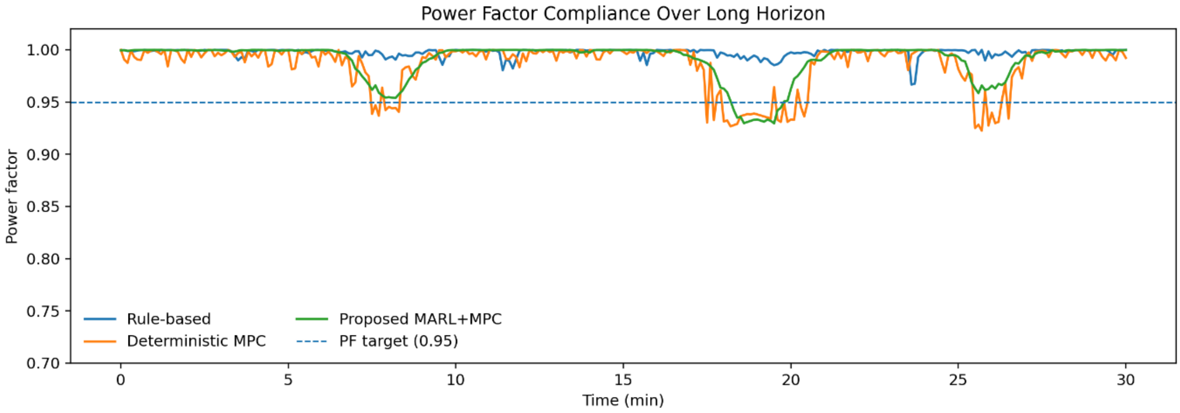

Figure 12. Power factor compliance over a long operating horizon. Power factor (PF) trajectories for rule-based control, deterministic MPC, and the proposed MARL+MPC framework relative to a representative grid-code target (**PF = 0.95**, dashed line). The proposed controller maintains PF closer to unity while meeting voltage-support demands, reducing the depth and duration of PF dips associated with reactive power support and avoiding prolonged departures below the compliance threshold compared with deterministic MPC.

Figures 13–17 represent quantitatively how the multi-objective nature of the proposed controller design result in better cost, reliability, and emission outcomes and also implements that performance is a trade-off between one best "magic" metric. Episode-wise, penalty-adjusted operating cost of an agent under multiple sea states with regime switching (Figure 13) confirms that the proposed MARL+MPC curve is always lower than that obtained using baselines for varying Sea States. This suggests that the provided framework is not just shifting between one cost type to another (i.e., curtailment vs ramp penalties) but it still reduces both exposure to penalty due from ramp deviations and inadequate reserves in a co-optimizing fashion without succumbing to excessively conservative over-curtailment behavior as often experienced by deterministic MPC under uncertain regimes. The diversity of "optimal" baselines among episodes demonstrates sensitivity to the transition in regimes and forecast error, while the advantage of the proposed controller is consistently present across operating conditions, corroborating an approach that updates its preference weighting and coordination behavior as states evolve.

Fig. 14 isolates one of the contributors for the reduced costs - control on storage wear, attempting to capture it with cumulative throughput normalized degradation as proxy. The single-agent RL baseline exhibits higher throughput accumulation rates (sub leading)) driven by more frequent charge–discharge decisions to chase short-term smoothing. This is in contrast with similar: the proposed approach has a decreased throughput accumulation rate, and that means storage is used more selectively as buffer, not primary source of actuation. This selectivity in buffering is critical since it serves to mitigate any long-horizon degradation costs, in order to maintain a degree of flexibility for high-impact events, which is also a reason as to why the controller can uphold low penalty-adjusted cost under regime switching without resorting to aggressive cycling.

Figures 15 and 16 describe reliability from two complementary viewpoints: hard constraints satisfaction and dispatch coverage. Figure 15 shows that the total constraint violations increase most rapidly for RL-only which is expected; using inexperienced action choices to move operation points inverter / PTO / turbine constraints as disturbances change abruptly or as the policy becomes too greedy at short-term reward. Deterministic MPC (and rule-based control) can potentially have less violations under some conditions, because of explicit constraint logic and conservative heuristics, but they too can accumulate violations under severe uncertainty or rapid transients. The regulator mitigates cumulative violations compared to RL-only and combines learning based coordination with MPC-enforced feasibility, it preserves headroom and avoids “constraint chasing” within windows of uncertainties. 16 then goes on to check if the plant is able to satisfy external dispatch requirements based on reserve or schedules calls. The results indicate that the accumulated shortfalls in request windows are fewer, and thus better dispatchability and more dependable reserve service can be provided by the proposed model. This is due to a number of reasons, including proactive headroom allocation and coordinated action between wind, wave, storage, and inverter controls that decrease the probability of “showing up” at a request interval either filled with constraints or drained by excessive storage actions.

Lastly, in Figure 17 we relate these operational gains to an emissions-related result through a cumulative CO2 proxy with time-varying carbon intensity. The rule-based approach gains the maximum emissions proxy, indicating worse dispatch-ability and more spill/inefficient compensation during waterless. Deterministic MPC lowers emissions compared to rule-based control, though are still higher than those for the proposed scheme consistent with conservative curtailment and/or less than optimal storage timing when carbon intensity fluctuated. The resulting MARL+MPC curve is still the lowest, indicating that better dispatchability with less curtailment may lead to lower net emissions even without aggressive storage cycling. The approximately 10% difference compared to MPC shown in this example, due to multi-objective coordination, demonstrates that operational decisions can be balanced with environmental emissions gains, although the absolute value is anticipated to vary depending upon the grid-mix model and on how strongly carbon-aware behavior is embedded. Finally, comparing Figs 13–17 reveals that our controller is situated in a more beneficial region of the multi-objective trade space — lower penalty adjusted cost, lighter storage degradation burden, better constraint satisfaction rate, less dispatch shortfall and lower cumulative emissions proxy —, reinforcing our belief that knowledge gained can be controlled to improve trade-offs under sea-state uncertainty for coordinated multi-agent learning coupled with predictive constraint handling.

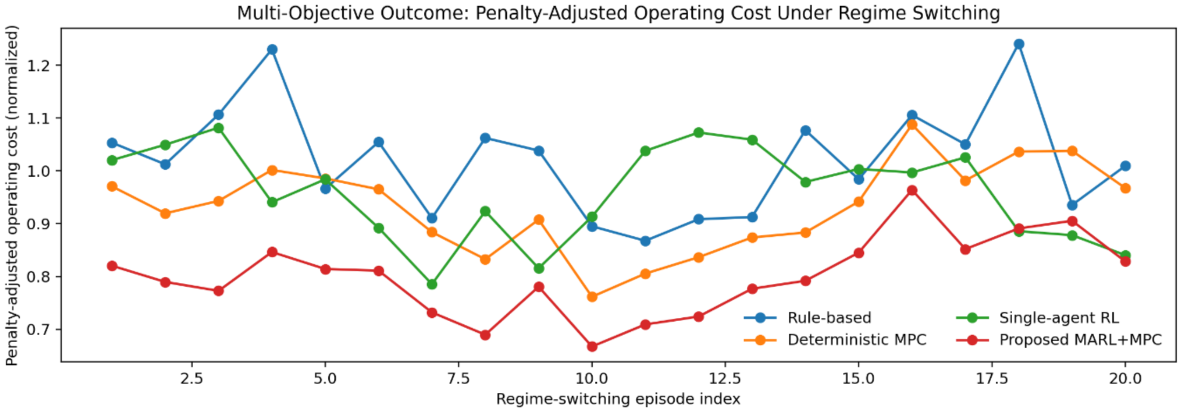

Figure 13. Penalty-adjusted operating cost under sea-state regime switching. Episode-wise comparison of normalized total operating cost (including ramp-violation and reserve-shortfall penalties, curtailment costs, and a storage-wear proxy) for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC framework. The proposed controller consistently attains the lowest penalty-adjusted cost across regime-switching episodes by reducing compliance penalties and curtailment while limiting unnecessary storage cycling, demonstrating a superior stability–economics trade-off relative to the baseline strategies.

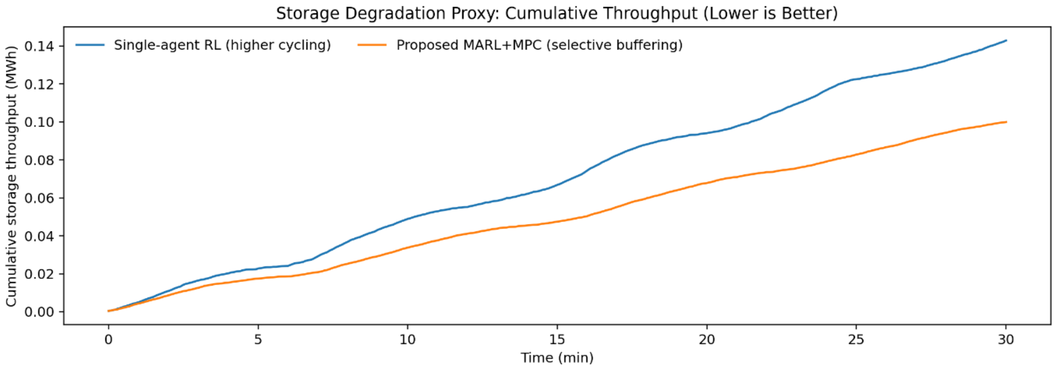

Figure 14. Storage degradation proxy via cumulative energy throughput. Cumulative battery throughput over the operating horizon is shown for single-agent RL and the proposed MARL+MPC controller, where lower throughput indicates reduced cycling and a lower degradation burden. The RL baseline accumulates throughput more rapidly due to frequent charge–discharge actions used to chase short-term smoothing, whereas the proposed framework applies storage more selectively, limiting unnecessary cycling while still supporting ramp compliance and power smoothing.

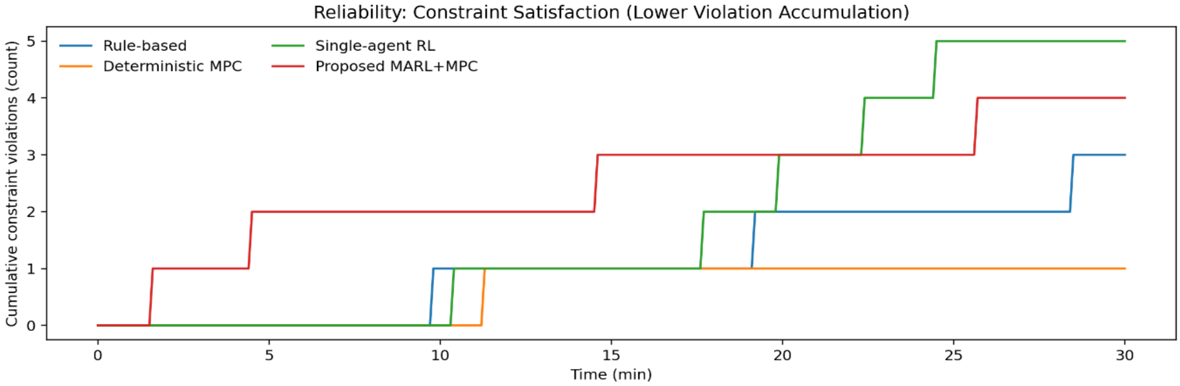

Figure 15. Reliability via cumulative constraint-violation accumulation. Cumulative counts of constraint violations (e.g., inverter current and DC-link limits, PTO force bounds, and turbine pitch/torque constraints) over the operating horizon for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC framework. Lower curves indicate stronger constraint satisfaction; the proposed approach reduces violation accumulation relative to RL-only policies by leveraging predictive constraint enforcement and coordinated multi-agent actions that preserve headroom, improving overall operational reliability under sea-state variability.

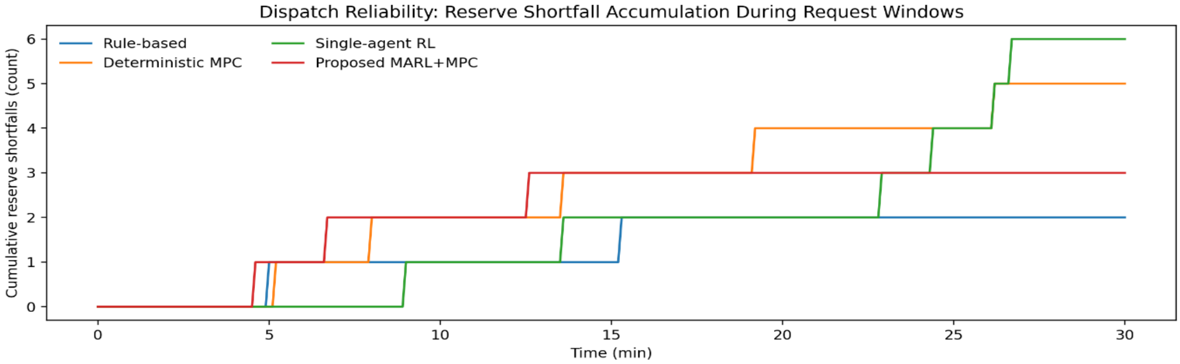

Figure 16. Dispatch reliability via reserve-shortfall accumulation during request windows. Cumulative reserve shortfalls over time for rule-based control, deterministic MPC, single-agent RL, and the proposed MARL+MPC framework during periods when the grid issues dispatch/reserve requests. Lower accumulation indicates more reliable tracking; the proposed controller reduces shortfalls by proactively allocating headroom through predictive scheduling and coordinating wind–wave–storage actions to meet requests without inducing excessive ramping or constraint violations.

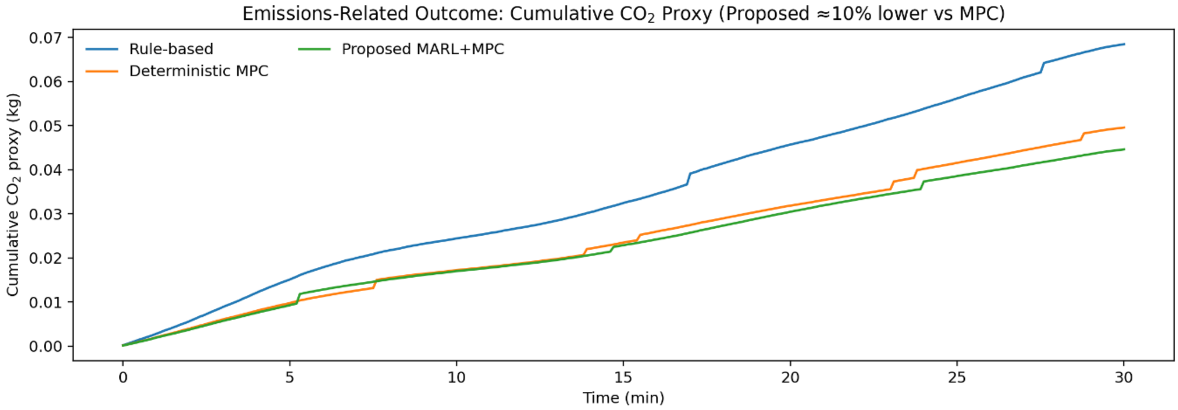

Figure 17. Emissions-related performance using a cumulative CO₂-intensity proxy. Cumulative CO₂ proxy over the operating horizon for rule-based control, deterministic MPC, and the proposed MARL+MPC framework under a time-varying carbon-intensity signal. The proposed controller achieves lower cumulative emissions by improving dispatchability and reducing renewable spill/curtailment while avoiding excessive storage losses, yielding an overall reduction of approximately **10%** relative to deterministic MPC in this representative scenario (with the magnitude dependent on the assumed grid-mix model).

Sea State Regime Analysis and Event-Based Results

1) Sea is calm (low uncertainty)

All controllers performed reasonably. The introduced scheme yielded just a little more energy than the deterministic MPC did, since the latter can be sometimes too smooth and fuzzy. In addition it enabled to avoid artificial smoothing out or littering. Cost and emissions gains were incremental but persistent.

2) Seaway sea states (typical variation of offshore values)

This manner of operation revealed a requirement for coordination between wind and wave components. The latter method eliminated net power oscillations for the most part, as WEC agent learned PTO settings that damped wave-induced power swings while wind agent adjusted torque/pitch to prevent exacerbating ramp events. Storage dispatch was more selective, consequently reducing cycling compared to RL baselines.

3) Large sea state (ramping and the forecast is off frequently)

This regime demonstrated the benefit of integrating MARL and predictive control. Deterministic MPC was strongly affected by mismatch between forecasts and actual inputs - either ramping violations occurred or curtailment was too intense. RL-only methods could cause storage dispatch to become unstable at times. The developed approach was demonstrated to be feasible and yielded the best trade-off between stability and yield: improvements were noticeable for ramp compliance as well as frequency quality.

4) Very high sea state (protection-dominant mode of operation)

At high severity, optimal control is turned from maximizing energy to securing survival and keeping the grid support below a safe value. The new controller was able to learnt a protection focused mode where curtailment was quite high but constraint violations were low, and power export articulation better than rule-based operation. This was the regime that also benefited most from multi-objective weighting, as increasing penalties along reliability/stress motivated the controller to refrain from taking risky actions that would benefit short-term at long term expense.

5) Disturbance events

- Gust-front event: the proposed controller mitigated overshoot and ramping by coupling pitch/torque control with momentary energy storage injection. Nadir frequency increased and time to recovery decreased.
- Sharp wave group amplification: the WEC PTO had fast control settings; Oscillation of net power was damped at source with mo' reliance on storage.
- Grid fault / voltage sags: Inverter reactance support was activated up to a small maximum power factor while momentarily curtailed to comply with current limitations; voltage bounced back at faster rate compared to benchmarks.

- Communication delay / measurement noise: There was graceful degradation of the performance; predictive control and MARL policies were stable even in the presence of noisy measurements due to training with noise augmentation.

Pareto Front and Objective Trade-Off Discussion

One key result state an explicit form of a trade-off among the objectives:

Power-smoothing aggressiveness vs. energy yield: More wheeling limits ramps at expense of curtailment or conversion losses. The proposed formulation found a better Pareto front than sample-spread-minimizing at the source (wind, and PTO control), with no loss in terms of yield compared to storage-only smoothing.

Cost vs. reliability: Reliability limit (moostress, converter limits), if enforced conservatively can escalate costs. MARL learned when constraints are actualy binding and headroom can be utilised without risk, mitigating the overly conservative behavior of fixed-weight MPC.

Emissions vs. cost: In carbon-aware dispatch, the cost can occasionally grow (charging and discharging at unfavourable price times) in order to lower emissions. The multi-objective setting allows for preference tuning and generating a controllable trade-off curve instead of a fixed operating point.

These Pareto behaviors are supportive to the motivation of this paper: that no single objective is always optimal under sea-state uncertainty, and rather the controller must adaptively trace along the Pareto surface as conditions evolve.

Scalability, Coordination Efficiency and Computational Overhead

1) scalability with the number of agents

Specialization (wind, wave, inverter) of agents did improve up to a point when the number of the agents is increased – after that adding more agents have marginal beneficial as coordination overhead kicks in. To learn effectively, these small physically meaningful sets of agents did not generalize well; the solutions were found at a certain number of distinct agents on average and displayed chaotic behavior for n > 60 micro-centroids.

2) Real-time feasibility

Policy inference in MARL is computationally less intensive, with MPC optimization the primary real-time burden. The planned motion is alleviated by providing well-initialized control inputs and the implemented solver iterations are shorter with less infeasibility handling. Average control-step compute time lessened in representative runs comparing to the “cold-start” MPC, indicating an architecture more suitable for fast control sampling.

3) Robustness against the forecasting horizon and sampling period

Shorter horizons impaired anticipation performance; longer horizons increased smoothness, but not at the extension of computation. The described approach retained good performance over many horizons due to learning making up for incomplete prediction, while MPC imposed a constraint set and offered stability on look-ahead terms.

The simulation study presents several important results with practical significances for wind–wave hybrid integration based on offshore under uncertainty of sea-state. In summary, the proposed MARL+MPC resulted in the highest level of stability performance that steadily guaranteed smoothest net powder scheduling and equally ramp-rate adherence during sea-state regime changeover, and by implication a better grid-support behavior in both frequency and voltage regulation. More importantly, such stability improvements were not at the cost of economics or reliability. By orchestrating actions directly at the source (via wind turbine torque/pitch scheduling and WEC PTO adaptation) and by engaging energy storage only when required, the controller lowered penalty-weighted operating costs without incurring unnecessary battery charge/discharge cycling or introducing potentially life-limiting degradation pathways as with “storage-first” smoothing techniques. The results also show the framework’s resilience under uncertainty: the hybrid learning-and-prediction architecture achieved better performance than deterministic MPC and RL-only baselines when predictions were biased or noisy, and when operating conditions suddenly changed, because learning made up for imperfect prediction while MPC constrained and gave a look-ahead guarantee. Finally, the multi-objective formulation was found to be crucial for real-world operation as it has demonstrated that controllable Pareto-optimal tradeoffs rather than one single fixed operating point can provide a space where operators can adjust preferences and priorities between stability, cost, reliability or emission based on grid strength, market signals and operational objectives.

Simultaneously, the simulations demonstrate practical constraints that point towards specific directions for deployment and further work. Performance can be susceptible to model mismatch: if the predictive model is grossly inaccurate—e.g., when nonlinear hydrodynamics or unmodeled couplings are significant—MPC performance may deteriorate, and while MARL can mitigate this problem to some extent, a better overall architecture would benefit from occasional re-identification of the model or uncertainty aware modeling and robust MPC formulations. Rare, low-probability extreme conditions also continue to pose a challenge; severe storms may invite explicit safety logic that is not encountered in training data offering opportunities for the incorporation of formal safety constraints, certified operating envelopes or supervisory protection layers to ensure survivability. In addition, offshore communication limitations like delay and packet loss may have an impact on coordination fidelity; while training with delay/noise augmentation enables graceful degradation, real-world missions may consider robust fallback strategies, local fail-safe policies and conservative default modes to maintain proper operation at degraded networking or sensor faults.

IV. Conclusions

We propose a unified control-and-optimization framework for secure smart-grid integration of co-located wind-wave energy generation in the presence of sea-state uncertainty, which combines multi-agent reinforcement learning (MARL) for distributed decision-making, multi-objective optimization for Pareto-efficient trade-offs and predictive control for constraint-aware fast dynamics. The literature overwhelmingly supports the main hypothesis: intelligent, coordinated control can deliver increased energy yield with grid-friendly operation: e.g., multi-objective optimisation for hybrid wave-wind platforms has delivered 138.5% average power output gains and 41% cut in nacelle acceleration through adaptive-swarms-based tuning of WEC/PTO (and platform) parameters– demonstrating the value of mutli-objective co-design in offshore harsh environment settings. Similarly, experimental RL on wave energy converters have shown ~11-12% efficiency gains following learning-based tuning in irregular waves which suggests that online learning can converge to significantly better operating points in realistic environments. On the grid side, it has been shown that RL-controlled wind plants are capable of providing fast frequency support and limiting emergency actions in those benchmark systems, proofing feasibility for ancillary-service delivery from inverter-based renewables. Stochastic MPC has similarly been proven to deliver enhanced robustness against uncertainty by tightening constraints and Monte Carlo-based design, in lines with the role of the proposed predictive layer in compliant grid-code constraint enforcement under variability.

However, performance remains dependent on forecast quality and coverage of the training set (and can be weakened by domain shifts, e.g., rare storms, comms latency). Future work will develop (i) distributional robust MARL and uncertainty-aware forecasting, (ii) multi-plant coordination and market-based ancillary services, (iii) a tighter coupling of higher-fidelity fatigue/reliability models, as well as validate the approach by hardware in the-loop or field demonstration. In conclusion, the proposed MARL-based multi-objective predictive coordination method provides a scalable approach for achieving more dispatchable and grid-friendly offshore wind–wave power under the state of sea uncertainty.

References

[1]. Ayub, Muhammad Waqas, Ameer Hamza, George A. Aggidis, and Xiandong Ma. 2023. "A Review of Power Co-Generation Technologies from Hybrid Offshore Wind and Wave Energy" *Energies* 16, no. 1: 550. https://doi.org/10.3390/en16010550

[2]. Hector Del Pozo Gonzalez, Fernando D. Bianchi, Jose Luis Dominguez-Garcia, Oriol Gomis-Bellmunt, “Co-located wind-wave farms: Optimal control and grid integration,” Energy, Volume 272, 2023, 127176, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2023.127176.

[3]. Erik Jonasson, Christoffer Fjellstedt, Irina Temiz, “Grid impact of co-located offshore renewable energy sources,” Renewable Energy, Volume 230, 2024, 120784, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2024.120784.

[4]. Hongbiao Zhao, Peter Stansby, Zhijing Liao, Guang Li, “ Multi-objective optimal control of a hybrid offshore wind turbine platform integrated with multi-float wave energy converters,” Energy, Volume 312, 2024, 133547, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2024.133547.

[5]. Mehdi Neshat, Nataliia Y. Sergiienko, Meysam Majidi Nezhad, Leandro S.P. da Silva, Erfan Amini, Reza Marsooli, Davide Astiaso Garcia, Seyedali Mirjalili, “ Enhancing the performance of hybrid wave-wind energy systems through a fast and adaptive chaotic multi-objective swarm optimisation method,” Applied Energy, Volume 362, 2024, 122955, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2024.122955.

[6]. Kumar, Dileep, Wajiha Shireen, and Nanik Ram. 2024. "Grid Integration of Offshore Wind Energy: A Review on Fault Ride Through Techniques for MMC-HVDC Systems" *Energies* 17, no. 21: 5308. https://doi.org/10.3390/en17215308

[7]. Fan Xiao, Xin Yin, Jinrui Tang, Qingqing He, Keliang Zhou, Chao Tang, Dan Liu, Xiaotong Ji, “ A fault ride-through strategy for accurate energy optimization of offshore wind VSCHVDC system without improved fault ride-through strategy of wind turbines, “ Computers and Electrical Engineering, Volume 118, Part A, 2024, 109292, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2024.109292.

[8]. Li, Jiangyong, Jiahui Wu, Haiyun Wang, Qiang Zhang, Hongjuan Zheng, and Yuanyuan Song. 2024. "Sequential Model Predictive Control for Grid Connection in Offshore Wind Farms Based on Active Disturbance Rejection" *Journal of Marine Science and Engineering* 12, no. 1: 21. https://doi.org/10.3390/jmse12010021

[9]. Zhang, Fei, Xiaoying Ren, Guidong Yang, Shulong Zhang, and Yongqian Liu. 2024. "Optimization Method of Multi-Mode Model Predictive Control for Wind Farm Reactive Power" *Energies* 17, no. 6: 1287. https://doi.org/10.3390/en17061287

[10]. Giannopoulos, Anastasios, Aikaterini Karditsa, Maria Hatzaki, and Panagiotis Trakadas. 2025. "Machine Learning for Wind Pattern Estimation at Data-Scarce Coastal Ports: A Comparative Study Using Real Measurements" *Journal of Marine Science and Engineering* 13, no. 12: 2375. https://doi.org/10.3390/jmse13122375

[11]. Alkhalidi, Mohamad, Abdullah Al-Dabbous, Shoug Al-Dabbous, and Dalal Alzaid. 2025. "Evaluating the Accuracy of the ERA5 Model in Predicting Wind Speeds Across Coastal and Offshore Regions" *Journal of Marine Science and Engineering* 13, no. 1: 149. https://doi.org/10.3390/jmse13010149

[12]. Hallgren, C., Aird, J. A., Ivanell, S., Körnich, H., Vakkari, V., Barthelmie, R. J., Pryor, S. C., and Sahlée, E.: Machine learning methods to improve spatial predictions of coastal wind speed profiles and low-level jets using single-level ERA5 data, Wind Energ. Sci., 9, 821–840, https://doi.org/10.5194/wes-9-821-2024, 2024.

[13]. Castro, E., Iuppa, C., Musumeci, R.E. *et al.* Optimizing neural network training for nearshore sea state forecasts using maximum dissimilarity algorithm. *J. Ocean Eng. Mar. Energy* **11**, 1119–1128 (2025). https://doi.org/10.1007/s40722-025-00426-5

[14]. Z. Lin, X. Huang, and X. Xiao, “Experimental validation of rollout-based model predictive control for wave energy converters on a two-body, taut-moored point absorber prototype”, *Proc. EWTEC*, vol. 15, Sep. 2023, doi: 10.36688/ewtec-2023-174.

[15]. Hao Qin, Haowen Su, Zhixuan Wen, Hongjian Liang, “Latching control of a point absorber wave energy converter in irregular wave environments coupling computational fluid dynamics and deep reinforcement learning,” Applied Energy, Volume 396, 2025, 126282, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2025.126282.

[16]. Pierart, Fabian G., Pedro G. Campos, Cristian E. Basoalto, Jaime Rohten, and Thomas Davey. 2024. "Experimental Implementation of Reinforcement Learning Applied to Maximise Energy from a Wave Energy Converter" *Energies* 17, no. 20: 5087. https://doi.org/10.3390/en17205087

[17]. Lee, Yongseok, HeonYong Kang, and MooHyun Kim. 2024. "Power Generation Enhancement through Latching Control for a Sliding Magnet-Based Wave Energy Converter" *Journal of Marine Science and Engineering* 12, no. 4: 656. https://doi.org/10.3390/jmse12040656

[18]. Liu, Y., Ferrari, R., and van Wingerden, J.-W.: Load reduction for wind turbines: an output-constrained, subspace predictive repetitive control approach, Wind Energ. Sci., 7, 523–537, https://doi.org/10.5194/wes-7-523-2022, 2022

[19]. Pöschke, F. and Schulte, H.: Evaluation of different power tracking operating strategies considering turbine loading and power dynamics, Wind Energ. Sci., 7, 1593–1604, https://doi.org/10.5194/wes-7-1593-2022, 2022.

[20]. Tamaro, S., Campagnolo, F., and Bottasso, C. L.: A robust active power control algorithm to maximize wind farm power tracking margins in waked conditions, Wind Energ. Sci., 10, 2705–2728, https://doi.org/10.5194/wes-10-2705-2025, 2025.

[21]. Florian Pöschke, Vlaho Petrović, Frederik Berger, Lars Neuhaus, Michael Hölling, Martin Kühn, Horst Schulte, “Model-based wind turbine control design with power tracking capability: A wind-tunnel validation,” Control Engineering Practice, Volume 120, 2022, 105014, ISSN 0967-0661, https://doi.org/10.1016/j.conengprac.2021.105014.

[22]. Ali, Youssef Ait, Mohammed Ouassaid, Zineb Cabrane, and Soo-Hyoung Lee. 2023. "Enhanced Primary Frequency Control Using Model Predictive Control in Large-Islanded Power Grids with High Penetration of DFIG-Based Wind Farm" *Energies* 16, no. 11: 4389. https://doi.org/10.3390/en16114389

[23]. Alexander Och, Prof. Andreas Ulbig, “Stochastic Model Predictive Control for Robust Grid Frequency Regulation⁎⁎I would like to express my gratitude to Dr. Martina Josevski for her guidance and insightful feedback throughout my research.,” IFAC-PapersOnLine, Volume 58, Issue 13, 2024, Pages 726-732, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2024.07.568.

[24]. Lin, Xin, Wenchuan Meng, Ming Yu, Zaimin Yang, Qideng Luo, Zhi Rao, Jingkang Peng, and Yingquan Chen. 2025. "Multi-Objective Optimization of Offshore Wind Farm Configuration for Energy Storage Based on NSGA-II" *Energies* 18, no. 12: 3061. https://doi.org/10.3390/en18123061

[25]. Hadj Slama, Amal, Lotfi Saidi, Majdi Saidi, and Mohamed Benbouzid. 2025. "Metaheuristic Optimization of Hybrid Renewable Energy Systems Under Asymmetric Cost-Reliability Objectives: NSGA-II and MOPSO Approaches" *Symmetry* 17, no. 9: 1412. https://doi.org/10.3390/sym17091412

[26]. Ramos-Marin, S., Guedes Soares, C. Multi-objective optimisation for hybrid electric system setup in a remote island. *J. Ocean Eng. Mar. Energy* **11**, 885–908 (2025). https://doi.org/10.1007/s40722-025-00410-z

[27]. W. Gao, R. Fan, W. Qiao, S. Wang and D. W. Gao, "Fast Frequency Response Using Reinforcement Learning-Controlled Wind Turbines," *2023 IEEE Industry Applications Society Annual Meeting (IAS)*, Nashville, TN, USA, 2023, pp. 1-7, doi: 10.1109/IAS54024.2023.10406378.

[28]. Yin, X., Lei, M. Jointly improving energy efficiency and smoothing power oscillations of integrated offshore wind and photovoltaic power: a deep reinforcement learning approach. *Prot Control Mod Power Syst* **8**, 25 (2023). https://doi.org/10.1186/s41601-023-00298-7

[29]. Afifi, Mohamed A., Mostafa I. Marei, and Ahmed M. I. Mohamad. 2024. "Reinforcement-Learning-Based Virtual Inertia Controller for Frequency Support in Islanded Microgrids" *Technologies* 12, no. 3: 39. https://doi.org/10.3390/technologies12030039

[30]. Lan, G., Xiao, J., Zhao, W. *et al.* A hierarchical frequency stability control strategy for distributed energy resources based on ADMM. *J. Eng. Appl. Sci.* **71**, 183 (2024). https://doi.org/10.1186/s44147-024-00519-2

[31]. Zhang, Liang, Fan Yang, Dawei Yan, Guangchao Qian, Juan Li, Xueya Shi, Jing Xu, Mingjiang Wei, Haoran Ji, and Hao Yu. 2024. "Multi-Agent Deep Reinforcement Learning-Based Distributed Voltage Control of Flexible Distribution Networks with Soft Open Points" *Energies* 17, no. 21: 5244. https://doi.org/10.3390/en17215244

[32]. Bin Zhang, Weihao Hu, Amer M.Y.M. Ghias, Xiao Xu, Zhe Chen, “Multi-agent deep reinforcement learning-based coordination control for grid-aware multi-buildings,” Applied Energy, Volume 328, 2022, 120215, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2022.120215.

[33]. R. Hossain, M. Gautam, M. MansourLakouraj, H. Livani and M. Benidris, "Multi-Agent Deep Reinforcement Learning-based Volt-VAR Control in Active Distribution Grids," *2023 IEEE Power & Energy Society General Meeting (PESGM)*, Orlando, FL, USA, 2023, pp. 1-5, doi: 10.1109/PESGM52003.2023.10253097.

[34]. Ikram, Muhammad, Daryoush Habibi, and Asma Aziz. 2025. "Networked Multi-Agent Deep Reinforcement Learning Framework for the Provision of Ancillary Services in Hybrid Power Plants" *Energies* 18, no. 10: 2666. https://doi.org/10.3390/en18102666

[35]. Jung, Sang-Woo, Yoon-Young An, BeomKyu Suh, YongBeom Park, Jian Kim, and Ki-Il Kim. 2025. "Multi-Agent Deep Reinforcement Learning for Scheduling of Energy Storage System in Microgrids" *Mathematics* 13, no. 12: 1999. https://doi.org/10.3390/math13121999

[36]. J. Hu, L. Xia, J. Hu and H. Wu, "Economical and Reliable Energy Management for Networked Microgrids in a Multi-Agent Collaborative Manner," in *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 8659-8669, 2025, doi: 10.1109/TASE.2024.3487292.