OffWindFM: A Conceptual Roadmap for Foundation Models in Offshore Wind Energy

Herrero Gómez-Tejedor Sofía Yaqin[†‡], Maurizio Vassallo[‡], Damien Ernst[‡]
[†]Universidad Pontificia de Comillas, Spain
[‡]University of Liège, Belgium
syherrero@student.uliege.be, {mvassallo, dernst}@uliege.be

Abstract—Offshore wind farms (OWFs) face complex, interdependent challenges that span power forecasting, maintenance, control, and grid integration. Current AI approaches address these challenges in isolation, requiring separate models for each task. We introduce OffWindFM, a unified multimodal foundation model (FM) that integrates heterogeneous data streams, such as SCADA measurements, meteorological forecasts, spatial farm layouts, and maintenance records, through specialized sub-encoders and a shared transformer-based fusion backbone. Pre-trained via a masked multi-horizon forecasting objective and fine-tuned with lightweight task heads, OffWindFM delivers end-to-end power-and-load forecasts and is readily extensible to maintenance, control optimization, market bidding, and provides a scalable foundation for comprehensive wind farm digital twins.

Index Terms—Offshore Wind Farms, Foundation Models, Multimodal Learning, Continual Learning, Time-series Forecasting

I. Introduction

Offshore wind energy (OWE) is rapidly gaining momentum as an important element in the transition towards renewable energy sources. In particular, the European Commission has set a target capacity of 300 GW of OWE by 2050 as a response to the expected growth in energy demand [1]. To achieve this goal, more and more focus is being placed on offshore wind farms (OWFs). Thanks to stronger and more consistent winds, OWFs can harness more power, making them a superior solution to onshore wind turbines. This advantage in energy yield stems from superior wind conditions, with studies showing offshore power density can reach 1522 W/m^2, almost three times the maximum power density found onshore at a 50-meter height [2]. Moreover, their deployment in isolated, non-urban sites makes them more socially accepted than onshore wind. However, the harsh marine environment, complex operational logistics, and high maintenance costs present significant challenges for their reliability and economic viability.

In this context, the need for smart, scalable models that are capable of supporting operations across forecasting, maintenance, and control is underscored. Foundation models (FMs) have appeared as an approach offering a promising solution to these challenges. These are large models trained on vast amounts of data. Once trained, FMs demonstrate strong generalization capabilities across a variety of tasks, enabling efficient adaptation to specific applications with minimal labeled data [3]. While FMs have successfully been applied across domains such as natural language processing (NLP) and computer vision (CV), their use in engineering systems, particularly in energy systems, such as OWE, remains underexplored. Recent studies that explore machine learning (ML) algorithms applied to OWE tend to focus on isolated components of the system, rather than exploring how they could possibly be leveraged across multiple aspects of WF operations.

In this paper, we propose a conceptual roadmap for developing a multimodal foundation model for OWFs. We identify key technical challenges, define the required data modalities and model architecture, and establish evaluation protocols for future empirical validation. Our contribution lies in unifying several OWF operational challenges under a single modeling paradigm and providing an implementation path for the research community.

The remainder of the paper can be summarized as follows. Section II provides an overview of the main challenges associated with OWFs, along with a review of related work. Section III introduces the FM approach. Section IV suggests a conceptual roadmap for implementing the proposed FM. Finally, Section V concludes the paper, highlighting the key limitations and outlining future directions.

II. Challenges

The effective operation and management of wind turbines (WTs) are characterized by a set of complex challenges, such as installation, planning, and environmental concerns. Moreover, OWE encounters additional complications from the harsh marine environment, where extreme weather conditions, saltwater corrosion, and complex accessibility lead to a higher levelized cost of energy (LCOE) compared to onshore WFs [2].

We will proceed by listing some of the main challenges linked to OWF. As outlined in Figure 1, these span forecasting, maintenance, control, and grid integration, and are discussed in the following subsections.

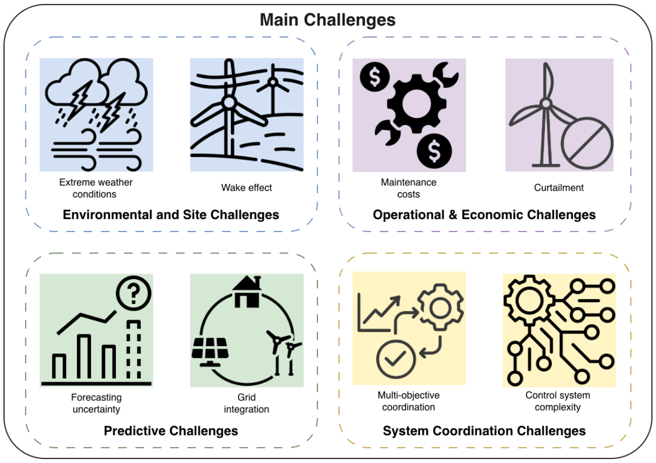

Fig. 1. Overview of the main OWE challenges.

A. Wind Variability and Forecasting Uncertainty

Wind speed is the fundamental variable of wind power generation, as shown in Eq. 1:

$$P_{generation} \propto v_{wind}^3. \tag{1}$$

The cubic dependence means that small errors in wind speed forecasts can lead to big errors in predicted power generation. Therefore, accurate forecasting is crucial for the commercial and grid integration of wind energy, enabling effective market bidding and stable grid operation. Furthermore, at the wind farm level, forecasts are essential for system-level optimization strategies, such as proactive wake steering to maximize total farm output and load mitigation to extend turbine lifespan.

Accurate power-output forecasts are challenging. Statistical methods try to infer future conditions from past data, yet offshore wind datasets are vast, heterogeneous, and highly nonlinear. Capturing the complex interactions across SCADA[1] measurements, weather forecasts, and wake effects requires rapid multimodal fusion. Even early machine-learning approaches demonstrate limited capacity for modeling these interdependencies and poor adaptation to unique SCADA patterns at different sites [4]. These limitations motivate the emergence of advanced AI-based models, capable of capturing the intricate interactions among the variables. Authors in [5] designed a hybrid model combining a convolutional neural network (CNN) cascaded with a radial basis function neural network (RBFNN) for 24-hour-ahead forecasting. Reference [3] uses a pre-trained BERT (Bidirectional Encoder Representations from Transformer) model, fine-tuned for spatio-temporal wind power forecasting, achieving high performance in tested scenarios.

Although these models achieve state-of-the-art performance, they still operate in silos, as they have no real understanding of the WFs as a complex, dynamic system. For instance, they are unable to capture dependencies across several time frames simultaneously. Models are trained to perform single tasks, for example, forecasting at a fixed time horizon (i.e., 30 minutes, 24 hours), rendering them inadaptable to different temporal contexts. This rigidity in the approach limits their utility in modeling the WF as a continuously evolving system. Moreover, most models fail to handle the wake effect produced by wind interaction with the WTs, which alters the incoming wind reaching neighbouring turbines. This is especially significant in a context where this phenomenon accounts for 10% to 20% of annual energy production (AEP) losses [6]. Current models fail to address this challenge, as they are typically trained for task-specific applications. In contrast, an FM could not only capture this behavior, but also adapt the turbine control methods to minimize losses.

B. Operation and Maintenance Costs

In OWE production, operation, and maintenance (O&M) costs represent a substantial expense, accounting for approximately 20% of the LCOE and representing 60% of operational expenditures (OpEx)[2]. These high costs derive from the marine environments, where the remote locations complicate logistics and obstruct access to the turbines. Historically, maintenance was reactive, where components were repaired or replaced after a malfunction was detected. However, the combination of unexpected faults and complex repair logistics leads to unacceptable downtimes, significantly lowering AEP and efficiency while elevating O&M costs. To address these limitations, preventive maintenance strategies were adopted, involving prescheduled inspections to preserve the state of WTs. While preventive maintenance approaches show better results than reactive methods, its fixed-time basis makes it overly reliant on failure rates and average failure interval estimates; the higher the inaccuracy of the estimates, the lower the effectiveness of the maintenance scheme [7]. For this reason, predictive maintenance has emerged as an enhanced approach.

[1]Supervisory Control and Data Acquisition (SCADA) is a system that enables remote monitoring, control, and automation.

[2]Recurring costs associated with the day-to-day operation and maintenance of turbines and related infrastructure.

This uses historical SCADA data to predict potential equipment failures before they occur, thereby minimizing downtime and maximizing turbine availability. Still, inaccurate modeling can lead to over-maintenance, which eventually increases OpEx and O&M costs. In this regard, authors in [8] showed that predictive maintenance can reduce costs by 2.5% in the short run only if the prediction model is highly accurate.

The opportunity cost of creating accurate predictive models has motivated the exploration of studies that leverage ML capabilities for this purpose. The work in [9] combines long short-term memory (LSTM) networks with fuzzy synthetic assessment to predict components' state in real-time. Nonetheless, this method is validated only on one turbine component, therefore lacking generalizability capabilities. Reference [10] introduces graph neural networks (GNNs) for unsupervised WT condition monitoring using SCADA data. Authors in [11] fine-tune a pre-trained large time series model for the prediction of WT SCADA data. Even though they find that the model's accuracy is not always consistent and depends on data abundance, they observe that for limited data availability, their proposed algorithm outperforms other baseline models.

A major gap persists in the integration of numerical weather prediction (NWP), component health, and logistics into a unified framework [4]. This integration is particularly important in environments where WTs accessibility is limited, weather conditions are unpredictable, and failure events can have cascading impacts. The lack of weather data integration is especially inefficient when planning maintenance operations for WTs [7]. For example, without weather forecasting integration, turbines at risk of failure may continue operating until breakdown. A predictive model that considers both equipment health and upcoming weather events could advise operators to shut down the turbine in advance, minimizing downtime by avoiding unserviceable failures during adverse conditions.

C. Control System Complexity

Two control methods can be distinguished at the structural level: yaw control, which adjusts nacelle orientation, and pitch control, which modifies the blade angle. While control methods differ on the turbine employed, most traditional controllers rely on simplified models manually tuned to follow maximum power point tracking (MPPT) [12]. This strategy adjusts rotor speed and blade pitch to maximize energy capture, using wind speed measurements and forecasts to continuously update nacelle orientation and blade pitch.

However, these approaches struggle with multi-objective coordination (maximizing power output, limiting structural loads, and ensuring grid compliance) while also neglecting wake interactions, component aging, and demand-driven curtailment requirements [13]. By not addressing these limitations, optimal WF performance is not optimal, and the scalability of control strategies under dynamic conditions is limited.

These challenges have prompted a wave of studies that explore AI-based techniques concerning OWE control strategies. The work in [14] reviews the role of AI across turbine control and wind farm wake control, highlighting the benefits of deploying neural networks, fuzzy logic control, and reinforcement learning (RL) in optimizing performance under marine conditions. These AI-driven approaches are a step closer to developing scalable solutions in OWE, offering robust performance under data-limited scenarios.

Although several advanced control frameworks support multimodal data ingestion and coordinate WTs across a farm, these methods remain task-specific, whether maximizing power generation [12], reducing mechanical load, or enhancing energy quality [13]. Moreover, these methods generally depend on fixed action spaces and specific farm layouts, for example authors [14] found that most models are trained on specific small to medium-scale farms. Consequently, any change in the site, data modality, or operational goal would require costly retraining [4]. Thus, a research gap prevails in developing a context-aware model capable of dynamically fusing heterogeneous inputs and addressing interdependencies to suggest an optimal control strategy. This gap could be bridged by a multi-agent FM trained on multimodal datasets that coordinates multiple turbines to output the most favorable operation.

D. Grid Integration and Curtailment

A complication that all renewable energy sources (RES) face is their correct integration into the grid. Unlike conventional power plants, RES generation is directly tied to its highly fluctuating energy source. For this reason, WTs' output cannot be dispatched on demand, making them vulnerable to mismatches between generation and real-time demand [12]. This mismatch is particularly adverse when considering the constrained and dynamic power system, where over-generation can lead to transmission congestion and grid instabilities that can trigger cascading effects.

Curtailment methods have appeared as a countermeasure, where a RES's generation is purposely limited by slowing it down or by shutting it off. Conventionally, this approach is implemented through static rule-based controls by capping a turbine's output at a fixed percentage or switching it off during low-demand periods [15]. The rigidity of these traditional approaches fails to capture grid dynamics, evolving economic trends, and fluctuating weather conditions, thus underestimating the power system environment. In addition to wasting energy, conventional curtailment approaches ignore turbine health despite the fact that constant switching can significantly accelerate mechanical wear, hence potentially increasing costs as well.

Recent research has arisen in which ML techniques are implemented to forecast congestion events and to optimize ramp-down strategies that reduce fatigue loads. Authors in [16] propose a DRL-based strategy for co-located wind farms and battery energy storage systems (BESS) to jointly manage wind curtailment and perform energy arbitrage in the spot market. Similarly, work in [17] considers three artificial neural networks (ANNs) to predict vertical power flow, wind power output, and grid component loading under N-1 contingency. Jointly, these models forecast congestion to determine the amount of renewable energy to be curtailed. Despite achieving desirable performance, it was noted that the model remained constrained to short-term horizons and specific topologies, lacking the generalization capability of extending across broader network conditions. This stresses the benefits that an adaptive, scalable FM could bear.

Furthermore, most current approaches neglect market trends, focusing only on operational parameters of the farm [16]. This detachment constrains the economic optimization of wind operations, especially when considering curtailment services. A model that took into account market trends could implement market bidding strategies so as to maximize profits while also strategically operating the WF.

III. Foundation Models

These challenges underscore the potential advantages of adopting an FM approach to effectively bridge the growing gaps between the industry's needs and current computational capabilities.

A. Foundation Model Approach

FMs represent a fundamental shift in AI development, moving from task-specific architectures to large-scale, general-purpose neural networks that learn rich hierarchical representations via self-supervised learning on vast, unlabeled data. This paradigm has been enabled primarily by the transformer architecture [18], whose attention mechanisms allow for effective modeling of long-range dependencies and parallel computation at scale. This architecture is considered a breakthrough in AI research as its self-attention mechanism allows effective long-range dependency modeling and parallel processing, enabling scalability. After pre-training, the model can adapt to new downstream tasks through fine-tuning with minimal task-specific data, avoiding the need to train from scratch for each application.

The original transformer architecture consists of two components: an encoder, which projects each input token into a high-dimensional vector space and uses self-attention to capture contextual relationships across the entire sequence, and a decoder, which generates the output autoregressively by attending both to these encoder representations and to its own previously generated tokens. The FM approach can be divided into three main steps, represented in Fig. 2.

1) Pre-Training. In this stage, the model is fed massive amounts of unlabeled data that have gone through sampling, filtering, and pre-processing to meet quality and diversity requirements. This data is then used in a self-supervised learning setup that typically involves the reconstruction of masked or transformed portions of the input. A general-purpose pre-trained FM is thus obtained in this step.
2) Fine-Tuning. The pre-trained model is then adapted to a set of downstream tasks using minimal labeled datasets. This supervised learning process updates the model's weights, allowing it to specialize in concrete applications displaying high performance.
3) Inference. The fine-tuned model is deployed to perform real-world tasks, in which user prompts are filtered and pre-processed to match the model's expected format. The model then generates outputs according to the prompts.

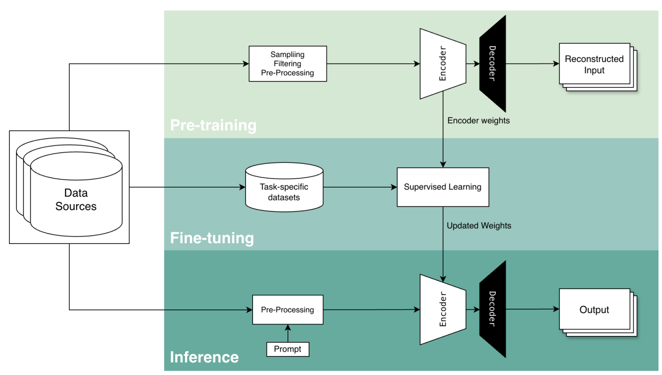

Fig. 2. Overview of the FM approach in three stages: pre-training, fine-tuning, and inference.

B. Advantages

Multimodal and spatiotemporal integration. FMs can ingest and reason across multiple data modalities, including SCADA readings, sequential data, weather forecasts, etc., while also accounting for spatiotemporal dynamics. In this way, system-level understanding of the environment is achieved.

Homogenization and generalization. Concerning the heterogeneous data digestion, FMs can establish cross-task adaptability, avoiding the need for siloed models for each distinct function. By reducing the need for individual specialized models, FMs reduce system fragmentation and increase system understanding. These unified frameworks enable to model complex interdependencies, allowing for the optimal multiple-objective coordination within a single inference pass.

Simulation capability. As opposed to traditional physics-ruled methods, which require expert-crafted rules and can incur significant computational overhead, FMs capture the underlying relationships directly from the datasets. Even though they require long training time, at inference time, they can generate high-fidelity predictions much faster than numerical simulators, enabling rapid analysis and real-time scenario testing without sacrificing accuracy.

Scalability. Once pre-trained, FMs require minimal labelled data to adapt to specific contexts, easily allowing for scalability. This is particularly beneficial in OWFs where data is scarce and expensive to obtain.

C. Hurdles to Implementation

Data accessibility. FMs need vast amounts of data for pre-training and the correct performance of downstream tasks. This is complicated by the scarcity of data, which tends to be proprietary and hard to access. This poses a major challenge, as public data tends to be site-specific, hindering the generalizability of upcoming models.

Lack of domain-specific fine-tuning. Downstream tasks rely heavily on the fine-tuning process applied to the pre-trained FM. In emerging applications such as power grid ones, there is minimal infrastructure and standardization on how to fine-tune large models.

Cybersecurity risks. The entrance of FMs into the highly interconnected landscape of energy systems also introduces some novel vulnerabilities. Their reliance on large-scale data as well as their automated decision-making process creates new routes for cyberattacks. Moreover, their black-box decision processes hinder threat detection, further delaying real-time responses.

Trust. While FMs have an outstanding performance, the fact that they do not follow physics or mathematics-based models has given rise to many trust-related concerns. Their black-box nature, which prevents interpretability, has placed them at the center of critical discussions regarding their opacity, especially in high-stakes domains such as energy systems.

IV. Proposed implementation

Building on the concepts introduced in Sections II–III, we outline a two-stage roadmap for deploying OffWindFM, designed to progressively tackle the unique challenges of offshore wind farms. Recent work by [19] demonstrated the viability of phased FM development for power grid applications.
We adopt this framework to OWE with the following modifications: i) masked multi-horizon forecasting as the pre-training objective, ii) cross-modal attention fusion to integrate weather forecasts with operational data, iii) physics-informed loss terms for wake effects and turbine aerodynamics, and iv) continual learning protocols to adapt to non-stationary failure modes.

A. Model Architecture

Given the heterogeneous data streams of OWFs, the proposed model deploys a multimodal encoder composed of specialized sub-encoders whose outputs are merged into a single latent space, following the example of models such as PerceiverIO [20]. A shared decoder ingests that unified representation to generate its output, which is then fed to modular task heads. These heads allow their independent fine-tuning without retraining the entire model [21, 22]. While this architecture enables adaptation to diverse OWF applications (Fig. 3), simultaneously targeting all tasks would be overly ambitious as an initial step.

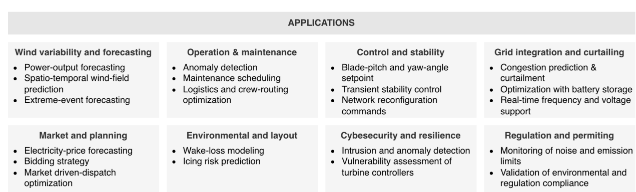

Fig. 3. Examples of FM applications in OWFs.

We therefore adopt a phased implementation strategy. In the near term, we pre-train OffWindFM-v0 on multi-horizon power-and-load forecasting (top right of Fig. 4), which forces the model to learn turbine aerodynamics, wake interactions, and weather coupling—the foundational physics driving both power output and mechanical stress. In the long term, we envision OffWindFM to address more OWF applications by including more sub-encoders, modular task heads, and possibly adapters, as discussed in Section IV-C.

B. Pilot deployment: OffWindFM for forecasting applications

OffWindFM-v0 is trained to master multi-horizon power-and-load forecasting by ingesting the different data modalities of an OWF through a four-phase implementation roadmap.

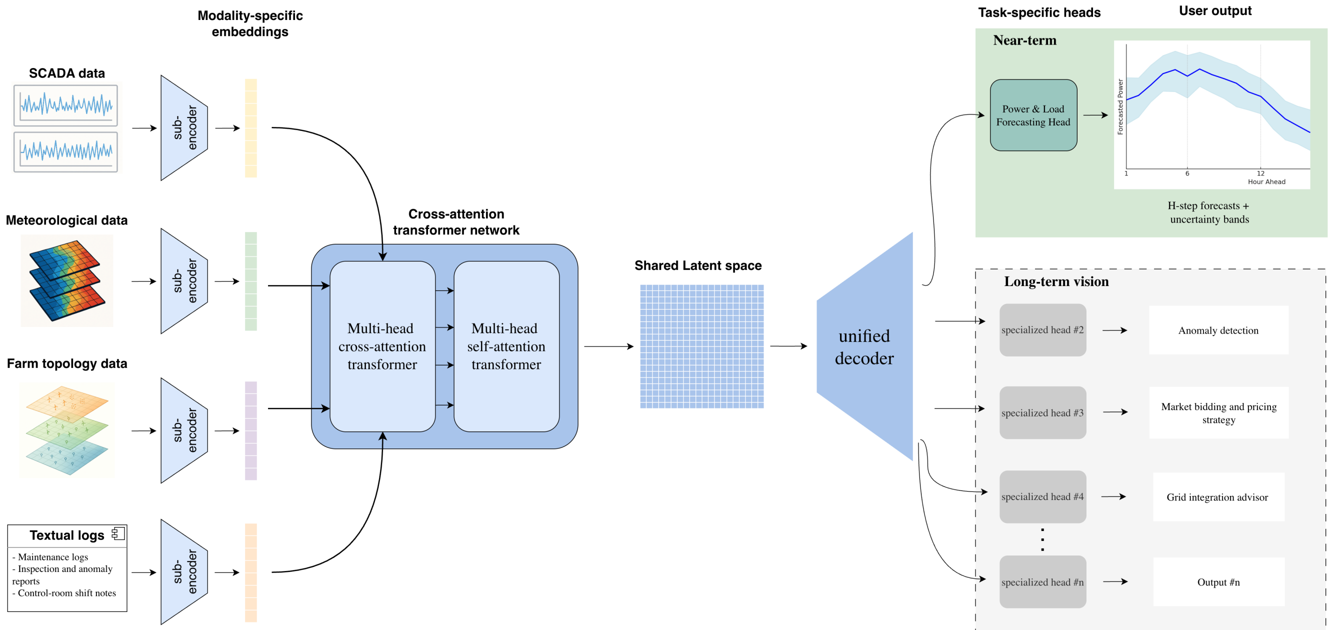

Fig. 4. OffWindFM architecture for multi-horizon forecasting. SCADA time-series, meteorological fields, farm topology, and textual logs are each processed by dedicated sub-encoders. Per-modality embeddings are merged through a cross-attention and then a self-attention transformer block to form a shared latent representation. A unified decoder funnels this latent space into a lightweight multi-horizon power & load forecasting head, outputting 24-hour ahead point predictions accompanied by their uncertainty bands. In the long term, additional task heads are integrated to target as many applications as possible.

1) Phase 1: identify, collect, and generate training data: Training OffWindFM-v0 requires a vast and comprehensive dataset, including:

- SCADA time-series: time-stamped WT operational data, including electrical measurements, mechanical states, temperatures, etc.
- Meteorological data: NWP from on-site met masts for forward-looking forecasts, LiDARs, etc.
- Farm topology: turbine coordinates, hub heights, rotor diameters.
- Maintenance logs: textual data from maintenance management systems such as inspection reports, technician notes, etc.

Existing sources include: NREL OpenOA for aggregated wind farm performance [23], Engie's La Haute Borne dataset (7 onshore turbines, 2013-2017), Kelmarsh dataset (6 offshore turbines, 3 years) [24], and FINO platforms for offshore meteorology [25]. Comprehensive SCADA with aligned NWP, topology, and maintenance logs is unavailable publicly. Failure events and extreme weather scenarios are underrepresented with respect to normal conditions.

Moreover, it is possible to generate synthetic data using physics-based simulators: SOWFA for large-eddy wake simulation [26], FLORIS for engineering wake models, and OpenFAST for turbine dynamics [27]. This extensive and diverse dataset—combining synthetic generation with strategic access to real operational data.

However, we emphasize that data availability, not model architecture, constitutes the primary bottleneck for offshore wind FMs.

2) Phase 2: model architecture development: Echoing recent works on multi-component and multimodal architectures, which leverage adapted specialized modules and keep the configurations that yield consistent gains on validation data [20], [21], [28], during phase 2, we cycle through alternative sub-encoder architectures and training regimes. We monitor performance across the relevant forecasting horizons on held-out validation subsets and freeze the best-performing configurations for downstream tasks [29].

All four input streams flow through different sub-encoders, which produce latent representations. These are then fused via cross-modal attention layers. In this way, a shared understanding of the farm state is collected in the embedded space, enabling the decoder to route that unified representation into its modular task heads. For the pilot OffWindFM version, we attach a multi-horizon forecasting head that projects the fused representation into H-step[3] power and mechanical-load forecasts, alongside parallel quantile branches for

[3]Forecasting horizon, number of time steps ahead the model is predicting.

uncertainty estimates.
During pre-training, a subset of the next 24 h of turbine-level power and mechanical-load targets are randomly masked, and the FM is trained to reconstruct these values from prior inputs. This masked-forecasting task ensures the unified latent space captures both deterministic dynamics and stochastic behaviors.

3) Phase 3: Validation on industry data: Since high-resolution SCADA, meteorological, and layout data are often proprietary, OffWindFM-v0 is pre-trained and evaluated on openly available datasets and synthetic wake-flow simulators, as mentioned in phase 1. The next step is to partner with OWF operators to fine-tune and validate the model with their proprietary data, either by processing data behind their firewall or via federated-learning protocols, which have gained traction in recent years [30]. This ensures access to real SCADA, NWP, and farm layout without exposing sensitive information.

4) Phase 4: implementation on use-cases: In the near term, OffWindFM would be deployed for multi-horizon power-and-load forecasting into existing farm operations: power-and-load predictions are served to SCADA dashboards and market-bidding tools. Operators can visualize point forecasts alongside uncertainty bands, enabling more informative reserve scheduling and curtailment decisions.

C. Ecosystem Expansion: OffWindFM for a larger set of applications

While Section IV-B explores the near-term pilot deployment of the proposed approach, targeting a single application, the long-term objective of the model is for it to encompass as many tasks as possible. Here, we discuss how to build new industry-relevant applications on top of OffWindFM.

1) Phase 5: Extending the FM to more applications: Extending OffWindFM to new applications builds directly on its modular, multimodal backbone.
When a novel task emerges, it is possible to introduce a new sub-encoder that ingests the relevant data stream. That sub-encoder plugs into the existing cross-modal fusion layers, enriching the shared latent space with information that complements SCADA, NWP, topology, and log inputs. Once the new sub-encoder is fused, the corresponding task head tailored to the downstream objective is attached. Training focuses solely on these additional modules: the FM's core encoders remain frozen, while the fusion-layer parameters, the new sub-encoder, and the task head parameters are updated. This add-and-freeze approach has been validated by recent multimodal FM studies such as ATLAS [22] and AdapterFusion [31].
To efficiently fine-tune the FM for new tasks or new data, we leverage parameter-efficient adaptation techniques. This parameter-efficient fine-tuning minimizes compute and data requirements, prevents interference with existing capabilities [22], and accelerates deployment of dozens or even hundreds of specialized applications on top of the same FM [20] [32].

V. Conclusion

Offshore wind farms (OWFs) face interdependent operational challenges (forecasting, maintenance, control, and grid integration) that current AI approaches address in isolation through task-specific models. We propose OffWindFM, a unified multimodal foundation model (FM) that integrates heterogeneous data streams (SCADA, meteorological forecasts, farm layouts, maintenance logs) through specialized sub-encoders and cross-modal fusion layers.
Compared to traditional physics-based simulators or univariate statistical models, the FM offers a single, end-to-end multimodal solution that requires no manual feature engineering, runs in real-time on farm-scale data, and continuously improves as new inputs are received.
Despite its promise, OffWindFM relies on large, often proprietary data streams that can vary significantly in quality and availability across sites, which limits transfer between sites. Moreover, its novel-adapter-only fine-tuning may limit responsiveness to unforeseen fault modes. Future directions include the development of federated and continual fine-tuning pipelines for on-site adaptation, and constructing an open, multimodal benchmark to standardize evolution across forecasting, maintenance, market, and control tasks.

References

[1] European Commission. Offshore Renewable Energy. 2024. url: https://energy.ec.europa.eu/topics/renewable-energy/offshore-renewable-energy_en.

[2] Belachew Desalegn et al. "Onshore versus offshore wind power trends and recent study practices in modeling of wind turbines' life-cycle impact assessments". In: Cleaner Engineering and Technology (2023). issn: 2666-7908. doi: 10.1016/j.clet.2023.100691.

[3] Zefeng Lai et al. "BERT4ST: Fine-tuning pre-trained large language model for wind power forecasting". In: Energy Conversion and Management 307 (2024), p. 118331. issn: 0196-8904. doi: 10.1016/j.enconman.2024.118331.

[4] Masoud Masoumi. "Machine Learning Solutions for Offshore Wind Farms: A Review of Applications and Impacts". In: Journal of Marine Science and Engineering 11.10 (2023). issn: 2077-1312. doi: 10.3390/jmse11101855.

[5] Ying-Yi Hong and Christian Lian Paulo P. Rioflorido. "A hybrid deep learning-based neural network for 24-h ahead wind power forecasting". In: Applied Energy 250 (2019), pp. 530–539. issn: 0306-2619. doi: 10.1016/j.apenergy.2019.05.044.

[6] Kuichao Ma et al. “Research on Evaluation Method of Wind Farm Wake Energy Efficiency Loss Based on SCADA Data Analysis”. In: Sustainability 16.5 (2024). issn: 2071-1050. doi: 10.3390/su16051813.
[7] Harriet Fox et al. “A Review of Predictive and Prescriptive Offshore Wind Farm Operation and Maintenance”. In: Energies 15.2 (2022), p. 504. doi: 10.3390/en15020504.
[8] Rasmus Dovnborg Frederiksen et al. “Cost-Effectiveness of Predictive Maintenance for Offshore Wind Farms: A Case Study”. In: Energies 17 (2024), p. 3147. doi: 10.3390/en17133147.
[9] Yongchao Zhu et al. “Operational state assessment of wind turbine gearbox based on long short-term memory networks and fuzzy synthesis”. In: Renewable Energy 181 (2022), pp. 1167–1176. issn: 0960-1481. doi: 10.1016/j.renene.2021.09.070.
[10] Chenlong Feng, Chao Liu, and Dongxiang Jiang. “Unsupervised anomaly detection using graph neural networks integrated with physical-statistical feature fusion and local-global learning”. In: Renewable Energy 206 (2023), pp. 309–323. issn: 0960-1481. doi: 10.1016/j.renene.2023.02.053.
[11] Yuwei Fan et al. Fine-Tuning Pre-trained Large Time Series Models for Prediction of Wind Turbine SCADA Data. 2024. arXiv: 2412.00403 [cs.LG].
[12] Eduardo José Novaes Menezes, Alex Maurício Araújo, and Nadège Sophie Bouchonneau da Silva. “A review on wind turbine control and its associated methods”. In: Journal of Cleaner Production 174 (2018), pp. 945–953. doi: 10.1016/j.jclepro.2017.10.297.
[13] Hooman Ghaffarzadeh and Ali Mehrizi-Sani. “Review of Control Techniques for Wind Energy Systems”. In: Energies 13.24 (2020). issn: 1996-1073. doi: 10.3390/en13246666.
[14] Dongran Song et al. “Review on the Application of Artificial Intelligence Methods in the Control and Design of Offshore Wind Power Systems”. In: Journal of Marine Science and Engineering 12.3 (2024). issn: 2077-1312. doi: 10.3390/jmse12030424.
[15] Nathan Oaks Farrar, Mohd Hasan Ali, and Dipankar Dasgupta. “Artificial Intelligence and Machine Learning in Grid Connected Wind Turbine Control Systems: A Comprehensive Review”. In: Energies 16.3 (2023). issn: 1996-1073. doi: 10.3390/en16031530.
[16] Jinhao Li, Changlong Wang, and Hao Wang. “Optimal Energy Storage Scheduling for Wind Curtailment Reduction and Energy Arbitrage: A Deep Reinforcement Learning Approach”. In: arXiv preprint arXiv:2304.02239 (2023).
[17] Elena Memmel et al. “Predicting Renewable Curtailment in Distribution Grids Using Neural Networks”. In: IEEE Access 11 (2023), pp. 20319–20336. doi: 10.1109/ACCESS.2023.3249459.
[18] Ashish Vaswani et al. Attention Is All You Need. 2017. arXiv: 1706.03762 [cs.CL].
[19] Hendrik F. Hamann et al. “Foundation models for the electric power grid”. In: Joule 8.12 (2024), pp. 3245–3258. issn: 2542-4351. doi: 10.1016/j.joule.2024.11.002.
[20] Andreas Jaegle et al. “Perceiver IO: A General Architecture for Structured Inputs & Outputs”. In: (2021). url: https://arxiv.org/abs/2107.14795.
[21] Tianyi Zhang et al. “A Stronger Mixture of Low-Rank Experts for Fine-Tuning Foundation Models”. In: (2023). url: https://arxiv.org/abs/2306.07617.
[22] Yaqing Wang et al. “ATLAS: Adapter-Based Multi-Modal Continual Learning With a Two-Stage Learning Strategy”. In: (2023). url: https://arxiv.org/abs/2306.09961.
[23] Jordan Perr-Sauer et al. “OpenOA: An Open-Source Codebase For Operational Analysis of Wind Farms”. In: Journal of Open Source Software 6.58 (2021), p. 2171. doi: 10.21105/joss.02171.
[24] Charlie Plumley. Kelmarsh wind farm data. Version 0.0.3. Zenodo, Feb. 2022. doi: 10.5281/zenodo.5841834.
[25] Beatriz Cañadillas et al. “Offshore wind farm wake recovery: Airborne measurements and its representation in engineering models”. In: Wind Energy 23.5 (2020), pp. 1249–1265. doi: 10.1002/we.2484.
[26] Matthew J. Churchfield et al. “A numerical study of the effects of atmospheric and wake turbulence on wind turbine dynamics”. In: Journal of Turbulence 13 (2012), N14. doi: 10.1080/14685248.2012.668191.
[27] J Jonkman et al. Definition of a 5-MW Reference Wind Turbine for Offshore System Development. Tech. rep. National Renewable Energy Laboratory (NREL), Golden, CO., Jan. 2009. doi: 10.2172/947422.
[28] V. Abad-Alcaraz et al. “Multimodal deep learning for solar radiation forecasting”. In: Applied Energy 393 (2025), p. 126061. issn: 0306-2619. doi: 10.1016/j.apenergy.2025.126061.
[29] Roman Bachmann et al. MultiMAE: Multi-modal Multi-task Masked Autoencoders. 2022. arXiv: 2204.01678 [cs.CV].
[30] Peter Kairouz et al. “Advances and Open Problems in Federated Learning”. In: Foundations and Trends in Machine Learning 12.1–2 (2019), pp. 1–210. url: https://arxiv.org/abs/1912.04977.
[31] Jonas Pfeiffer et al. AdapterFusion: Non-Destructive Task Composition for Transfer Learning. 2021. arXiv: 2005.00247 [cs.CL].
[32] Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019. url: https://arxiv.org/abs/1810.04805.