

Application Paper

Cross-Domain Offshore Wind Power Forecasting: Transfer Learning Through Meteorological Clusters

Dominic Weisser¹, Chloe Hashimoto-Cullen² and Benjamin Guedj^{1,3}

¹University College London, United Kingdom.

²Sorbonne Université, France.

³Inria, France.

Keywords: transfer learning, wind power forecasting, Gaussian processes

Abstract

The global offshore wind sector is growing rapidly to meet decarbonisation targets. The increasing number of newly commissioned wind farms require accurate power forecasts from the outset. This allows grid operators to limit reserve requirements, maintain system stability and support effective energy trading. Although machine learning algorithms have delivered strong performances in offshore wind power forecasting, these methods typically depend on detailed, voluminous and site-specific data: data that new sites do not yet have. To address this, we introduce a transfer learning framework built around weather-driven forecasting models. Rather than training a single model on all weather conditions, we train several specialised models, each trained on historical data of a distinct weather pattern. When pre-trained on sufficiently large and varied datasets, these models adapt to new offshore sites by leveraging encoded cross-site knowledge. Because the models are already calibrated to seasonal variations, they eliminate the industry standard dependency of at least a full year of local measurements. Our framework is empirically verified with results showing that cross-domain knowledge transfer can meaningfully reduce the data requirement for new offshore wind farms, achieving an average MAE of 3.83% across eight target sites using just over seven months of data, demonstrating that reliable forecasting is achievable without a complete annual cycle of site-specific measurements. These results open broader opportunities elsewhere in the offshore wind sector. One such area is wind resource assessment, which evaluates the wind climate at proposed offshore wind farm locations. It typically requires extensive site-specific measurements before early-stage decisions can be made, which makes early assessment slower and more expensive. Thus, our weather pattern-based transfer learning framework offers a way to tap into the substantial body of historical data from wind farms with similar climates already in operation.

1. Introduction

The sustainable energy transition positions offshore wind as a cornerstone of future power systems. While the global offshore installed wind capacity is projected to nearly triple from 83 GW in 2024 to 238 GW by 2030 (Energy for Generations 2025), the growth is geographically uneven. As of 2025, offshore wind installations are concentrated in a small number of developed nations (Statista 2025), leaving unexploited potential in emerging markets. Countries including Brazil, India, Morocco, the Philippines, South Africa and Vietnam collectively possess an estimated 3.1 TW of offshore wind potential (ESMAP 2019), however according to Global Wind Energy Council 2025, Vietnam is the only one of these nations with operational offshore wind capacity. Exploiting these potential locations requires efficient wind resource assessment (WRA) to identify optimal sites, as well as accurate power forecasts from the onset of operation. Reliable prediction is the prerequisite for integrating intermittent wind power into local grids. This ensures stability and secures financial viability in new markets.

The task of Wind Power Forecasting (WPF) has been extensively studied, with approaches ranging from physical methods based on numerical weather prediction to statistical approaches like Autoregressive Integrated Moving

Average (ARIMA) or Support Vector Machines (SVM). Whilst physical models are robust, they tend to be computationally intensive and exhibit systematic biases from their limited ability to resolve sub-grid-scale atmospheric phenomena. Consequently, in recent years, Machine Learning (ML) and specifically Deep Learning (DL) models have gained massive popularity within the geophysical and energy communities. Thanks to their efficacy in modelling high-dimensional dynamic systems and capturing non-linear dependencies, DL architectures have achieved impressive accuracy for offshore WPF (Ally et al. 2025). However, the existing literature predominantly focuses on mature, data-rich environments. State-of-the-art models are typically single-site architectures trained on highly granular, farm-specific SCADA data accumulated over years of operation.

This reliance on deep, site-specific historical datasets creates a critical barrier for the "cold-start" of newly commissioned farms and WRA campaigns, especially in emerging markets. Without sufficient operational data, traditional DL models struggle to generalise, and training from scratch is statistically inefficient given the universal physical laws governing wind power generation. To address this, transfer learning (TL) has emerged as a promising solution, leveraging knowledge from a data-rich source domain (\mathcal{D}_S) to improve predictive performance in a target domain (\mathcal{D}_T) (Pan and Yang 2010). However, existing applications of TL in large-scale offshore wind have largely focused on transferring farm-level models (Li et al. 2024; Sajol et al. 2024), treating the meteorological data as a single dataset.

Parallel research in renewable energy systems has shown that splitting training data into meteorological clusters and training weather-specific sub-models can enhance forecasting accuracy in both offshore wind and photovoltaic applications (Rana et al. 2016). Yet, these clustering techniques remain isolated to single-site optimisation and have not been leveraged to facilitate knowledge transfer across different geographic locations. To the best of our knowledge, no study has yet synthesised these two paradigms to investigate the potential of weather-pattern specialised models as a mechanism for transfer learning - specifically, transferring knowledge across wind farms only between time-periods that share similar meteorological dynamics.

In this paper, we propose a novel TL framework designed to bridge this gap. We investigate whether target wind farms across Europe can achieve high forecasting accuracy with limited historical data, effectively replicating the challenge of newly commissioned sites. By capturing a comprehensive representation of the meteorological patterns prevalent offshore across Northern Europe we design a domain-agnostic feature space where operational wind power data can be clustered based solely on meteorological similarity rather than geographic location, therefore allowing specialised weather-pattern specific forecasting models to be trained only on the observations from a unique weather pattern. This architecture enables a highly efficient transfer mechanism: incoming data from a new target farm is mapped to the most similar meteorological cluster, allowing the corresponding pre-trained forecasting model to be fine-tuned with minimal local data.

The rest of this paper is structured as follows. Section 2 gives an overview of the dataset used, including the source and target wind farms as well as breaking down each methodological step. Section 3 presents and discusses the experimental results and Section 4 provides further discussion.

2. Methodology

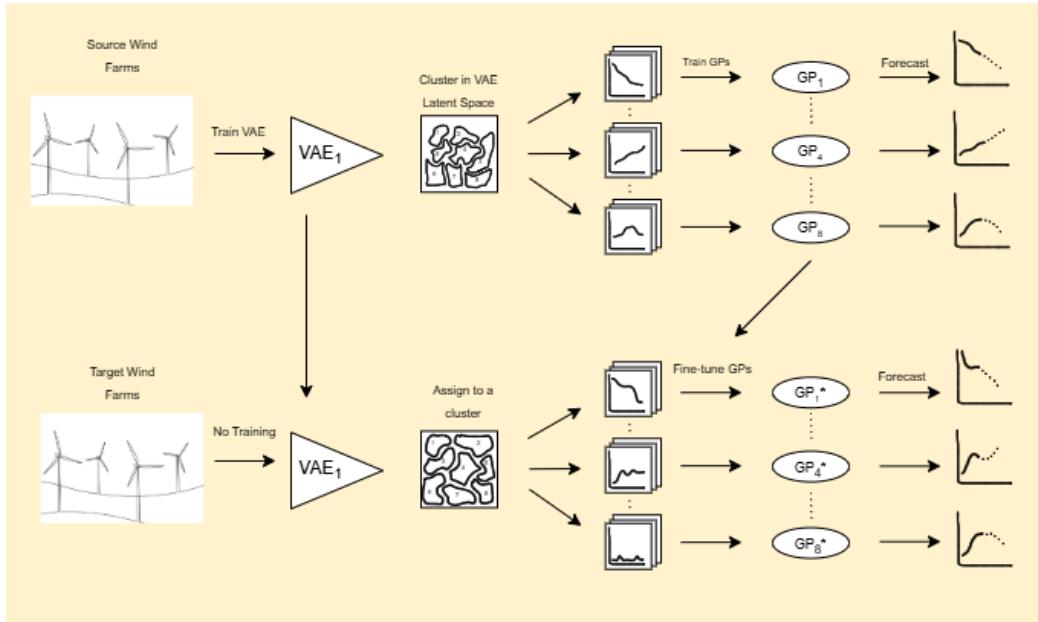


Figure 1: Overview of the Methodology

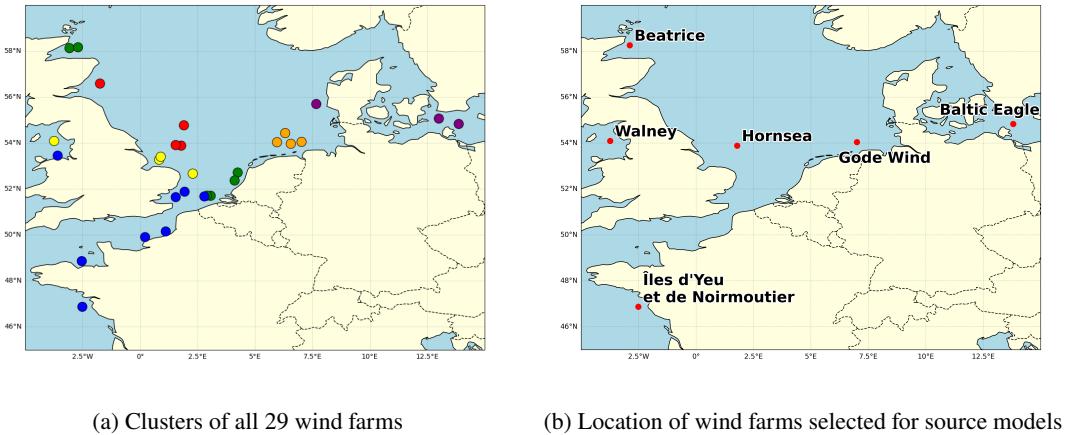
2.1 Dataset

We utilise the offshore wind farm dataset introduced by Grothe et al. (2022), which provides 40 years of hourly meteorological and synthetic power production data for 29 major European offshore wind farms up to and including 2019. The meteorological data is derived from ERA5, a global reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) that provides consistent hourly estimates of certain atmospheric variables. This data is combined with turbine-specific power curves to generate realistic hourly power production time series designed for research purposes.

2.2 Selecting source Wind Farms

To capture a comprehensive representation of meteorological patterns across Northern Europe, source wind farms were selected to maximise meteorological and operational diversity. Exposure to such diversity allows the models to learn from a broader range of atmospheric patterns, enhancing their ability to generalise to new sites that have distinct meteorological characteristics.

For each of the 29 available wind farms, a vector of 38 features was extracted from data recorded between January 1, 2018, and December 31, 2019. The feature set included wind speed statistics (mean, variance, percentiles), sea surface roughness distributions, wind directional consistency, power output metrics (including capacity factor and efficiency) and the frequency of extreme weather events. Dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2020), configured to consider 15 nearest neighbours for each data point and with a minimum distance of 0.1. This was followed by agglomerative hierarchical clustering (Ward 1963), which partitioned the farms into six distinct clusters based on both their meteorological and operational profiles. From each cluster, one farm was selected to represent the cluster as a source wind farm.



2.3 Identifying and Clustering Weather Patterns

To identify distinct weather patterns across the six source wind farms, hourly meteorological data from all farms is segmented into continuous, non-overlapping time periods of length p . Each period i is represented as a multivariate time-series $\mathbf{X}_i \in \mathbf{R}^{p \times d}$, with $d = 6$ meteorological features: wind speed, sea surface roughness, wind direction components ($\sin \theta$ and $\cos \theta$ for a wind direction θ) and horizontal and vertical wind velocity components at 100m above sea level (u_{100}, v_{100}). To ensure the forecasting models are trained on the most distinct weather patterns, the time-period length p and cluster count k were optimised via grid search, maximising the separation between identified weather patterns (see [Section 2.4](#)).

For a fixed time-period length p , we trained a Variational Autoencoder (VAE) (Kingma and Welling 2014) to learn a compressed representation of each weather pattern, with the encoder $\mathbf{q}_\phi(\mathbf{z} | \mathbf{X})$ mapping each input time-series to an 8-dimensional latent vector $\mathbf{z}_i \in \mathbf{R}^8$. The encoder architecture employed contains two 1D convolutional layers with kernel size 5 and 3 for local feature extraction, followed by a bidirectional Long Short-Term Memory (BiLSTM) network to capture temporal dependencies (Karim et al. 2018). The BiLSTM output is projected via fully connected layers to the parameters of the variational posterior, $\mu_\phi(\mathbf{X}) \in \mathbf{R}^8$ and $\log \sigma_\phi^2(\mathbf{X}) \in \mathbf{R}^8$. Latent samples are generated via the reparametrisation trick $\mathbf{z} = \mu_\phi(\mathbf{X}) + \sigma_\phi(\mathbf{X}) \odot \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and \odot denotes element-wise multiplication. The decoder $p_\theta(\mathbf{X} | \mathbf{z})$ mirrors this architecture. It expands the latent vectors through fully connected layers, processes the sequence via an LSTM, and employs transposed convolutional layers to reconstruct the original multivariate time-series $\hat{\mathbf{X}}$.

The model parameters θ and ϕ are optimised by minimising the β -VAE loss function (Higgins et al. 2017) with mean squared error reconstruction loss and KL divergence regularisation:

$$\mathcal{L}(\theta, \phi; \mathbf{X}) = \mathbb{E}_{q_\phi(\mathbf{z} | \mathbf{X})} \left(\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2 \right) + \beta D_{\text{KL}} \left(q_\phi(\mathbf{z} | \mathbf{X}) \parallel \mathcal{N}(0, I) \right),$$

Training is run for up to 200 epochs using AdamW optimisation. To mitigate posterior collapse, we employ a monotonic annealing schedule for β , linearly increasing from 0.01 to 1.0.

For inference, each period is deterministically encoded as the mean of the variational posterior, $\mathbf{z}_i = \mu_\phi(\mathbf{X}_i)$ providing a concise representation of the meteorological temporal dynamics. Prior to clustering, these latent vectors are L2-normalised. This ensures that the subsequent clustering is driven by the directional similarity of the weather patterns rather than the magnitude. Clustering is then performed on these normalised embeddings from all source wind farm periods using Hierarchical Agglomerative Clustering with Ward's linkage (Ward 1963) and Euclidean distance, grouping periods across the six farms into K distinct weather clusters.

2.4 Optimal Configuration Selection

The optimal configuration (p, K) is selected via grid search by maximising a composite quality score Q . Drawing upon the multi-objective evaluation framework of Rana et al. (2016), the composite score aggregates six metrics weighted to prioritise statistical robustness while ensuring physical interpretability and temporal coherence.

$$Q = 0.2S + 0.2\tilde{D} + 0.2\tilde{C} + 0.25\tilde{M} + 0.1T + 0.05H, \quad (2.1)$$

The structural quality of the latent clusters was evaluated using three complementary internal metrics: the silhouette score (S) (Rousseeuw 1987), which measures cluster cohesion and separation, the normalised Davies-Bouldin index (\tilde{D}) (Davies and Bouldin 1979) which evaluates the average similarity between clusters based on the ratio of within-cluster scatter to between cluster separation and the Calinski-Harabasz index (\tilde{C}) (Calinski and Harabasz 1974) which quantifies the ratio of between cluster to within-cluster dispersion. Beyond latent structure, we validate physical interpretability via meteorological separability (\tilde{M}), calculated as the log-normalised mean ANOVA F-statistic across wind speed, wind variability and power output, where high values ensure that the latent clusters correspond to statistically distinct physical weather clusters. Robustness is assessed through temporal coherence (T), defined as the average within-cluster autocorrelation (Warren Liao 2005), and distributional consistency (H = 1 - JSD), which uses the Jensen-Shannon Divergence (Lin 1991) to penalise shifts in cluster proportions between training and testing sets.

2.5 Gaussian Process Models

Following the assignment of meteorological periods to the K identified weather patterns, we model the relationship between atmospheric conditions and power generation using cluster-specific Gaussian Process regression (Williams and Rasmussen 2006). For each cluster k , a separate GP $f_k(\mathbf{x}) \sim \mathcal{GP}(m_k(\mathbf{x}), k_k(\mathbf{x}, \mathbf{x}'))$ is trained to learn the cluster-dependent power curve, enabling each model to specialise on the distinct dynamics characterising that specific weather type.

To capture both the overall period trajectory and immediate local conditions, we construct a 20-dimensional input vector $\mathbf{x} \in \mathbb{R}^{20}$ for each hourly forecast. This vector concatenates the 8-dimensional latent representation \mathbf{z} with 12 physical predictors: current wind speed and sea surface roughness; autoregressive lags (t-1, t-2) for both power and wind speed; and cyclical encodings (sin, cos) of the hour, month, and wind direction.

The covariance function (kernel) is designed to capture the multi-scale nature of wind power generation across the wind farms. We employ a composite kernel consisting of a Squared Exponential (RBF) component to model smooth power curve trends, summed with a Matérn- $\nu = 3/2$ component to capture the rougher, non-differentiable stochastic fluctuations typical of high-frequency atmospheric turbulence:

$$k_k(\mathbf{x}, \mathbf{x}') = \sigma_{f,k}^2 \left(k_{\text{RBF}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\ell}_k^{\text{RBF}}) + k_{\text{Matérn}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\ell}_k^{\text{Mat}}, \nu = 1.5) \right), \quad (2.2)$$

where $\sigma_{f,k}^2$ controls the output variance. Both components utilise Automatic Relevance Determination (ARD), where $\boldsymbol{\ell}_k \in \mathbb{R}^{20}$ represents a vector of length-scales. This allows the GP to automatically determine the relevance of each input feature for each specific cluster, for example weighting lags more heavily in stable conditions versus turbulence features during storms.

Observation noise is modelled via a Gaussian likelihood with variance $\sigma_{n,k}^2$, assuming $y = f_k(\mathbf{x}) + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma_{n,k}^2)$. Consequently, the covariance of the noisy observations is given by $\mathbf{K}_{y,k} = k_k(\mathbf{x}, \mathbf{x}') + \sigma_{n,k}^2 I$. The complete hyperparameter set $\theta_k = \{\sigma_{f,k}^2, \sigma_{n,k}^2, \boldsymbol{\ell}_k^{\text{RBF}}, \boldsymbol{\ell}_k^{\text{Mat}}, c_k\}$ is optimised by maximising the Exact Marginal Log-Likelihood (MLL) of the observations. Using the noisy covariance $\mathbf{K}_{y,k}$, the MLL is defined as:

$$\log p(\mathbf{y} | X, \theta_k) = -\frac{1}{2} (\mathbf{y}^\top \mathbf{K}_{y,k}^{-1} \mathbf{y}) - \frac{1}{2} \log |\mathbf{K}_{y,k}| - \frac{N}{2} \log(2\pi). \quad (2.3)$$

2.6 Transfer Learning

To evaluate the cross-site generalisation capability of the proposed framework, we apply a transfer learning strategy where the library of GP models trained on the six source farms is adapted to six unseen target farms. Leveraging the pre-trained library, specifically the fixed VAE encoder ϕ_S and the set of cluster-specific GP priors, the adaptation process proceeds in three stages: latent projection, cluster alignment and GP fine-tuning. First, the

target meteorological time series X_T are projected into the source latent space using the fixed source encoder $\mathbf{z}_T = \mu_{\phi_S}(\mathbf{X}_T)$. This enforces that target weather patterns are interpreted through the learned of the source domain. Each target period is then assigned to one of the K source weather clusters via a nearest-neighbourhood classification in the latent space, utilising the centroid from the source training set.

For each identified weather pattern $c \in \{1, \dots, k\}$, we initialise a target GP $f_T^{(c)}$ using the hyperparameters of the corresponding source GP, $f_S^{(c)}$. This initialisation provides a strong prior on the temporal covariance structure of the power output. The kernel hyperparameters θ_k are then optimised by maximising the exact marginal log-likelihood on the available target training data $\mathcal{D}_{\text{train}}$:

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmax}} \log p(\mathbf{y}_{\text{train}} \mid \mathbf{X}_{\text{train}}, \theta, \sigma_{n, \text{fixed}}^2).$$

To assess data efficiency, we conducted experiments varying the data availability of the two-year target datasets. Each target dataset is split into training fractions $\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, simulating a 'cold start' scenario where a newly commissioned wind farm has been operational for only a short period (ranging from approximately 2.4 to 12 months). The remaining $1 - \gamma$ is reserved for testing, with performance evaluated per cluster.

3. Results

3.1 Metrics Used

For performance evaluation, three primary metrics are employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the coefficient of Determination (R^2). To allow for comparison across wind farms with varying capacities, MAE is expressed as a percentage of each farm's total capacity. When aggregating calculations across multiple clusters or farms, the average is weighted by the number of time-periods in each cluster to account for different sample sizes.

3.2 Benchmark

We evaluate our approach in two settings. For source models, we compare K cluster-specific GPs with a single global GP trained on pooled data to assess the benefits of clustering. For transfer learning, we compare metrics after transferring source GPs under varying data capacities with GPs trained from scratch per target wind farm. As context, we reference the M3STIN model (Wang et al. 2025), the only study providing quantitative results on this dataset. Although its methodology differs, combining an Informer with a Graph Attention Network, we use its reported performance as an informal reference: trained on five years of hourly data from eight nearby wind farms, it achieves a best one-hour-ahead MAE of 3.60% ($R^2 = 97.36\%$) and 3.70% ($R^2 = 97.23\%$) for the M3STIN¹ variant.

3.3 Clustering of Meteorological Data

The grid search evaluation of the composite quality score identified the 6-hour time period window, partitioned into $K=8$ clusters as the optimal configuration, achieving a maximal metric of $Q = 0.576$. This parameter combination provided the most robust separation of meteorological patterns and serves as the structural foundation for the forecasting framework. The separation is visualised in Figure 3(a), where the 8-dimensional latent representations for each target period is embedded into a two-dimensional plane using t-SNE (Maaten and Hinton 2008). Figure 3(b) validates the physical distinctiveness of these clusters through the distribution of mean wind speeds. Clear differences are evident: specific clusters isolate stable extremes, such as the high wind-speed conditions in cluster 1 and 3 and the calm, low-speed periods in clusters 0 and 5. In contrast, cluster 2 exhibits greater internal variance, likely characterising transitional atmospheric states associated with the onset or decay of frontal systems.

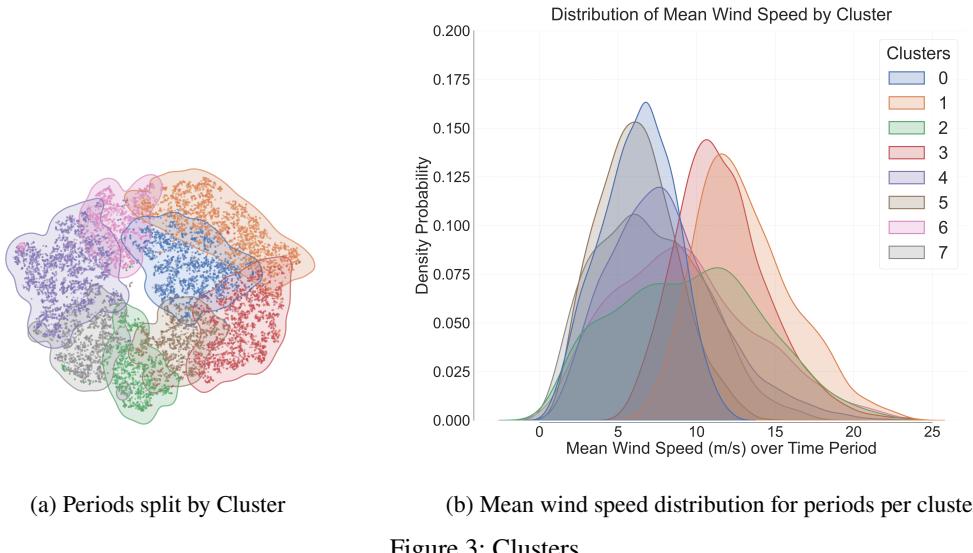


Figure 3: Clusters

3.4 Source GP Results

The forecasting performance of the proposed library of GP models was evaluated using two years of operational data across the six source wind farms, employing a 4:1 train-test split. Results were averaged across three independent random seeds to ensure statistical validity. Across all sites and weather clusters, the models delivered consistently high accuracy, achieving an overall MAE of 3.51% relative to installed capacity. In absolute terms, this corresponds to an average MAE of 23.4 MW, with a RMSE of 40.7 MW and a R^2 of 0.982.

The distinct variation in forecasting performance across clusters, as seen in Table 1, validates the underlying premise of our framework: that weather conditions possess inherent heterogeneity, with some clusters being significantly more predictable than others. For instance, cluster 1, characterised by consistent high wind speeds, yields high predictive accuracy. In contrast, cluster 5 - the least frequent cluster - exhibits the lowest performance, reflecting the difficulty of modelling rare and unique atmospheric conditions. Despite these variations, the framework remains robust. Furthermore, the GP models also effectively quantify the uncertainty, as illustrated in Figures 5.a and 5.b.

Cluster	Test samples	MAE (%)	
		GP	RF
0	2238	3.36	4.01
1	2425	3.39	3.71
2	2028	3.72	4.21
3	2315	3.74	4.25
4	2842	3.39	3.87
5	1622	4.02	4.55
6	2275	3.50	3.92
7	1740	3.06	3.32
Overall	17,485	3.51	3.94

Table 1: Cluster-level metrics for GP and RF models averaged across 3 seeds.

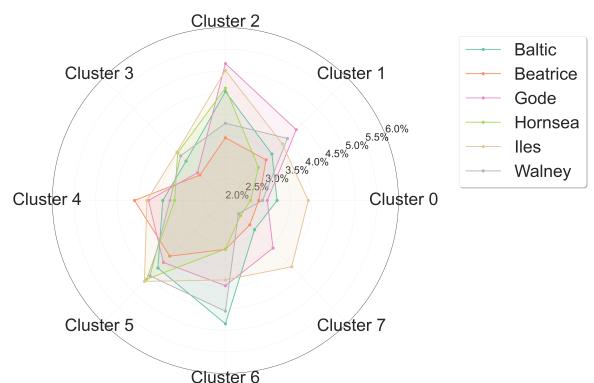


Figure 4: Metrics for GP forecasts, disaggregated by wind farm and cluster.

Farm specific analysis, illustrated in Figure 4, reveals further performance heterogeneity conditioned on the interaction between location and weather cluster. For instance, the Iles de Noirmoutier farm notably underperforms in clusters 0 and 7, whereas the Beatrice farm exhibits the lowest accuracy in cluster 4, despite performing well elsewhere. This variability validates our framework's strategy of concatenating source data across multiple sites. By aggregating weather patterns from various locations into each clusters source set, we ensure that our GP models are trained on a full distribution of wind behaviours associated with that weather pattern. Consequently, the source models possess a highly generalised 'memory' of potential intra-cluster variations, allowing them to adapt accurately and rapidly when transferred to the specific conditions of a target wind farm.

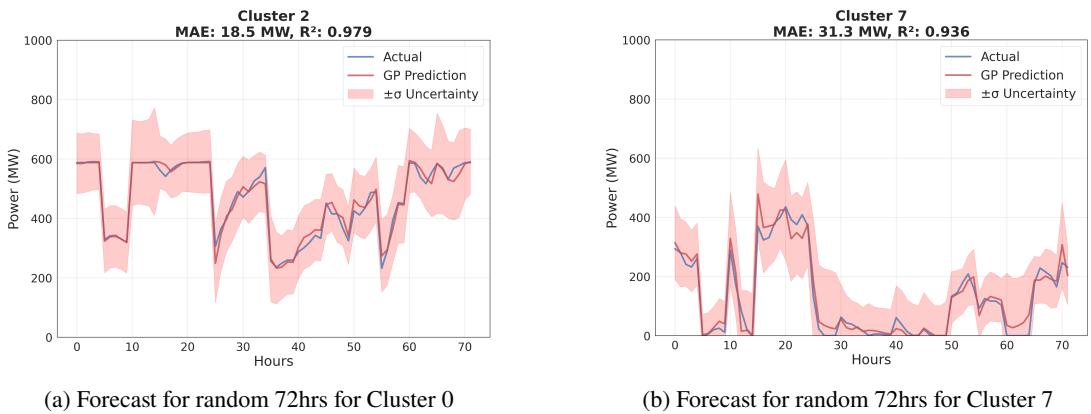


Figure 5

3.5 Transfer Learning Performance

Transfer learning experiments were conducted across eight target offshore wind farms using fine-tuning datasets ranging from 10% to 50% of the available two-year operational history. To ensure robustness, each experiment was evaluated using three random seeds, with the results averaged. It is important to note that lower data fractions correspond to increasingly stringent evaluation conditions; at 10% availability, the remaining 90% of the dataset is reserved for testing.

Table 2 presents the results averaged across all 8 sites, aggregated by cluster and data capacity. While increasing the amount of fine-tuning data consistently reduces error - achieving an overall aggregated MAE of 3.83% with just 30% (approx. seven months) of site-specific data - transferability varies significantly between clusters. Some weather clusters generalise well with limited data, while others require more site-specific inputs. For instance, cluster 1 demonstrates high transfer efficiency, adapting rapidly to achieve an MAE of 4.0% with only 10% (approx. three months) of site data. In contrast, cluster 6 proves more challenging, yielding a higher MAE of 7.1% with 10% local data, which decreases to 4.1% after increasing data to 50%. This suggests that the weather patterns associated with cluster 6 are more unique to individual farms, requiring more site-specific operational data to fine-tune the associated GP for the target farms.

Table 2: Model Performance by Cluster and Data Capacity

Clusters	Data Capacity									
	10%		20%		30%		40%		50%	
	MAE (%)	R ²	MAE (%)	R ²	MAE (%)	R ²	MAE (%)	R ²	MAE (%)	R ²
0	5.0%	0.83	4.2%	0.88	3.9%	0.91	3.8%	0.91	3.9%	0.91
1	4.0%	0.89	3.4%	0.91	3.2%	0.91	3.1%	0.90	3.1%	0.90
2	6.3%	0.93	4.4%	0.96	4.0%	0.97	3.7%	0.97	3.6%	0.97
3	5.0%	0.91	4.3%	0.93	4.1%	0.94	3.9%	0.94	3.9%	0.94
4	5.2%	0.95	4.2%	0.97	3.8%	0.97	3.7%	0.97	3.7%	0.97
5	5.9%	0.86	4.4%	0.92	4.1%	0.93	3.8%	0.94	3.7%	0.95
6	7.1%	0.92	5.2%	0.96	4.6%	0.97	4.3%	0.97	4.1%	0.97
7	5.6%	0.93	4.3%	0.96	3.9%	0.97	3.6%	0.97	3.5%	0.97
Average	5.21%	0.90	4.13%	0.94	3.83%	0.94	3.67%	0.94	3.60%	0.95

When we break down the forecasting performance of a single target site, as seen in Table 4 for the Gemini wind farm, we further see how transferability at the individual farm level is highly dependent on the weather cluster. For Gemini, transferability in clusters 1 and 3 is particularly strong; the pre-trained GP models align well with local weather dynamics, yielding MAEs of 2.2% and 3.8% with only 10% fine-tuning data. Furthermore, these two clusters represent 34.5% of the evaluated time periods, highlighting the substantial forecasting benefits this framework offers. Conversely, clusters 5 and 6 show lower transferability. In these cases, the source GP models are less representative of local conditions, resulting in higher initial MAEs of 7.3% and 7.0% with 10% data. Reducing these errors to 4.2% and 3.9% requires increasing local data usage to 50%. This highlights how the weather patterns for clusters 5 and 6 are more distinct to Gemini, necessitating a larger volume of local data to refine the learned representation and highlighting the inherent challenges of applying TL to offshore wind farms.

Table 3: Model Performance by Cluster and Data Capacity for the Gemini Wind Farm

Clusters	Data Capacity				
	10%	20%	30%	40%	50%
	MAE (%)	MAE (%)	MAE (%)	MAE (%)	MAE (%)
0	6.9%	5.6%	5.2%	5.0%	5.0%
1	2.2%	2.0%	1.9%	1.8%	1.8%
2	5.6%	3.4%	3.1%	2.8%	2.7%
3	3.7%	3.5%	3.2%	2.9%	2.8%
4	4.8%	3.9%	3.8%	3.6%	3.3%
5	7.3%	5.3%	4.8%	4.2%	4.2%
6	7.0%	5.0%	4.3%	4.0%	3.9%
7	7.0%	5.3%	4.8%	4.4%	4.2%
Average	5.04%	3.92%	3.60%	3.36%	3.27%

These findings demonstrate that forecasting performance is highly weather pattern dependent. Newly commissioned farms can achieve high accuracy with minimal data in some conditions, while complex meteorological conditions still require significant local adaptation. This variability suggests that the framework can be further optimised by selectively exploiting the clusters where knowledge transfer is most efficient, and directing additional modelling resources specifically toward the site-sensitive patterns that lag in performance.

4. Discussion

The demonstrated transferability of knowledge across European offshore sites through weather-specific models leads to two natural progressions for this framework. First, validation across more diverse climate zones, such as other continental regions, would establish the geographic boundaries of effective regime transfer and determine whether globally representative regime models could be developed. Second, incorporating turbine specifications (type, rated capacity, hub height, rotor diameter, etc...) to the framework would enable the models to learn how different turbine types respond to different weather patterns. This would allow the system to personalise forecasts for new offshore wind farms based on their specifications, (potentially) allowing for forecasts using only meteorological data. Subsequent integration of wake effect modelling for specific turbine layouts and validation across diverse operational sites would be necessary steps toward the ultimate vision: a dynamic rapid offshore wind potential assessment tool that provides configuration-specific forecasts tailored to location, turbine specifications, and farm layout without months of investment.

The weather regime-specific GP library developed in this work acts as an emulator of offshore power conversion, conditioned on the offshore meteorological state. Trained on data from six offshore wind farms across Europe, the library demonstrates strong generalisation across a range of capacities and turbine specifications. A natural extension is to incorporate turbine level engineering parameters, such as rotor diameter, hub height and manufacturer, as additional parameters during the training stage. Including these specifications would allow the GP models to more systematically capture cross-site variations in turbine performance, further reducing reliance on historical power output data and enabling the models to generalise primarily on meteorological inputs. When combined with established wake-modelling tools (e.g. FLORIS, PyWake, or Gaussian wake superposition) that account for farm layout and turbine interactions, this approach could form the basis of a virtual WRA workflow. Such a system has the potential to be extended into a virtual WRA platform capable of evaluating offshore wind farm performance across a variety of turbine types, locations and layouts. By generating power forecasts based on meteorological conditions, the platform would enable investors and planners to systematically assess the expected yield of different configurations prior to construction, reducing reliance on costly site-specific measurement campaigns and supporting more informed, data-driven decision-making in offshore wind development.

References

- Ally S, Verstraeten T, Daems PJ, Nowé A and Helsen J** (2025) Modular deep learning approach for wind farm power forecasting and wake loss prediction. *Wind Energy Science* **10**(4), 779–812. doi: [10.5194/wes-10-779-2025](https://doi.org/10.5194/wes-10-779-2025). <https://doi.org/10.5194/wes-10-779-2025>.
- Caliński T and JA H** (1974) A Dendrite Method for Cluster Analysis. *Communications in Statistics - Theory and Methods* **3**, 1–27. doi: [10.1080/03610927408827101](https://doi.org/10.1080/03610927408827101).
- DI D** (1979) A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* **1**, 224–227.
- Energy for Generations** (2025) Offshore wind targets underpin acceleration to 2030 and beyond. Available at <https://ember-energy.org/latest-insights/offshore-wind-targets-underpin-acceleration-to-2030-and-beyond/>.
- ESMAP** (2019) tech. rep. Going Global: Expanding Offshore Wind to Emerging Markets. World Bank Group and Energy Sector Management Assistance Program (ESMAP). Available at <https://documents.worldbank.org/curated/en/716891572457609829/pdf/Going-Global-Expanding-Offshore-Wind-To-Emerging-Markets.pdf>.
- Global Wind Energy Council** (2025) tech. rep. Global Wind Report 2025. Global Wind Energy Council. Available at <https://www.gwec.net/reports/globalwindreport>.
- Grothe O, Kächele F and Watermeyer M** (2022) Analyzing Europe's biggest offshore wind farms: a data set with 40 years of hourly wind speeds and electricity production. *Energies* **15**(5), 1700.
- Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S and Lerchner A** (2017) beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In *International Conference on Learning Representations*. Available at <https://openreview.net/forum?id=Sy2fzU9gl>.
- Karim F, Majumdar S, Darabi H and Chen S** (2018) LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* **6**, 1662–1669. ISSN: 2169-3536. doi: [10.1109/ACCESS.2017.2779939](https://doi.org/10.1109/ACCESS.2017.2779939). [http://dx.doi.org/10.1109/ACCESS.2017.2779939](https://doi.org/10.1109/ACCESS.2017.2779939).
- Kingma DP and Welling M** (2014) Auto-Encoding Variational Bayes. *ICLR*.
- Li D, Hu Y, Yang B, Fang Z, Liang Y and He S** (2024) A novel transfer learning strategy for wind power prediction based on TimesNet-GRU architecture. *Journal of Renewable and Sustainable Energy* **16**. doi: [10.1063/5.0200518](https://doi.org/10.1063/5.0200518).
- Lin J** (1991) Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* **37**(1), 145–151. doi: [10.1109/18.61115](https://doi.org/10.1109/18.61115).

- Maaten L van der and Hinton G** (2008) Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**(86), 2579–2605. <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- McInnes L, Healy J and Melville J** (2020) *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv: [1802.03426 \[stat.ML\]](https://arxiv.org/abs/1802.03426). Available at <https://arxiv.org/abs/1802.03426>.
- Pan SJ and Yang Q** (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* **22**(10), 1345–1359. doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191).
- Rana M, Koprinska I and Agelidis V** (2016) Solar Power Forecasting Using Weather Type Clustering and Ensembles of Neural Networks. In doi: [10.1109/IJCNN.2016.7727853](https://doi.org/10.1109/IJCNN.2016.7727853).
- Rousseeuw PJ** (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*.
- Sajol MSI, Islam MS, Hasan AJ, Rahman MS and Yusuf J** (2024) Wind Power Prediction across Different Locations using Deep Domain Adaptive Learning. In *2024 6th Global Power, Energy and Communication Conference (GPECOM)*. IEEE, 518–523.
- Statista** (2025) Number of offshore wind farms worldwide as of April 2025, by country. Available at <https://www.statista.com/statistics/264257/number-of-offshore-wind-farms-worldwide-by-country/>.
- Wang Z, Wang C, Chen L, Yu M and Yuan W** (2025) Short-term offshore wind power multi-location multi-modal multi-step prediction model based on Informer (M3STIN). *Energy* **322**, 135616. ISSN: 0360-5442. doi: <https://doi.org/10.1016/j.energy.2025.135616>. <https://www.sciencedirect.com/science/article/pii/S0360544225012587>.
- Ward JHJ** (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **58**.
- Warren Liao T** (2005) Clustering of time series data—a survey. *Pattern Recognition* **38**(11). ISSN: 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2005.01.025>. <https://www.sciencedirect.com/science/article/pii/S0031320305001305>.
- Williams CK and Rasmussen CE** (2006) *Gaussian Processes for Machine Learning*. vol 2. 3. MIT press Cambridge, MA.