



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

Received xx xxx xxxx

**Keywords:** transfer learning, wind power forecasting, Gaussian processes

## Abstract

The global offshore wind sector is expanding rapidly to meet decarbonisation targets, leading to an increasing number of newly commissioned wind farms that require accurate power forecasts from the outset. Reliable early-stage forecasting is essential for grid stability, reserve management, and efficient energy trading. While machine learning methods have demonstrated strong performance in offshore wind power forecasting, they typically rely on large volumes of site-specific data, which are unavailable for new installations. We propose a transfer learning framework based on meteorological clustering to address this data scarcity. Instead of training a single model across all atmospheric conditions, we learn a set of specialised forecasting models, each associated with a distinct weather pattern. Pre-trained on large and diverse offshore datasets, these models capture transferable, climate-dependent dynamics and can be adapted efficiently to new sites. By leveraging their built-in calibration to seasonal and meteorological variability, our approach removes the industry-standard requirement of a full year of local measurements. We evaluate the framework across eight offshore wind farms, demonstrating that accurate cross-domain forecasting can be achieved using just over seven months of site-specific data. The proposed method attains an average mean absolute error of 3.83%, confirming that reliable power forecasts are possible without observing a complete annual cycle. Beyond power forecasting, this meteorology-aware transfer learning approach now opens new opportunities for offshore wind applications such as early-stage wind resource assessment, where reducing data requirements can significantly accelerate and de-risk project development.

## Impact Statement

This work demonstrates that with transfer learning, accurate offshore wind power forecasting is achievable for newly commissioned wind farms without a full year of local measurements, significantly reducing early-stage uncertainty, costs, and delays in grid integration and project development.

## 1. Introduction

The sustainable energy transition positions offshore wind as a cornerstone of future power systems. Global offshore installed wind capacity is projected to nearly triple from 83 GW in 2024 to 238 GW by 2030 (Energy for Generations, 2025), yet this growth remains geographically uneven. As of 2025, offshore wind installations are concentrated in a small number of developed nations (Statista, 2025), leaving substantial untapped 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.

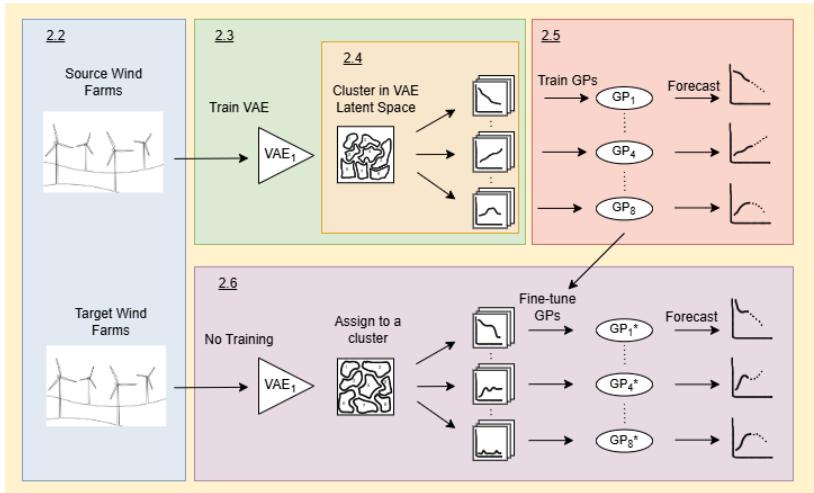
Unlocking this potential requires both efficient wind resource assessment (WRA) to identify optimal sites and accurate power forecasting from the onset of operation. Reliable wind power prediction is essential for integrating intermittent generation into local grids, ensuring system stability, limiting reserve requirements, and supporting the financial viability of offshore projects—particularly in nascent markets where uncertainty is highest.

Wind Power Forecasting (WPF) has been widely studied, with approaches ranging from physics-based numerical weather prediction to statistical methods such as Autoregressive Integrated Moving Average (ARIMA) models and Support Vector Machines (SVM). While physical models are robust, they are computationally intensive and often exhibit systematic biases due to limited resolution of sub-grid-scale atmospheric processes. Consequently, Machine Learning (ML), and especially Deep Learning (DL), methods have gained prominence in recent years. Their ability to model high-dimensional dynamical systems and capture non-linear dependencies has led to state-of-the-art performance in offshore WPF (Ally et al., 2025).

Due to the nature of these models, advances in the literature largely focus on mature, data-rich environments. Most state-of-the-art forecasting models are single-site architectures trained on highly granular, farm-specific SCADA data accumulated over several years. This strong reliance on extensive local historical data limits their applicability to newly commissioned wind farms. This dependency creates a critical barrier for the “cold-start” problem faced by new wind farms and early-stage WRA campaigns, particularly in emerging markets. In the absence of sufficient operational data, DL models trained from scratch struggle to generalise, despite the shared physical principles underlying wind power generation. Transfer learning (TL) has therefore emerged as a promising approach, enabling knowledge from a data-rich source domain  $\mathcal{D}_S$  to improve performance in a data-scarce target domain  $\mathcal{D}_T$  (Pan and Yang, 2010). However, existing TL applications in large-scale offshore wind forecasting primarily transfer farm-level models (Sajol et al., 2024; Li et al., 2024), treating meteorological inputs as a single homogeneous dataset.

In parallel, research has shown that partitioning data into meteorological clusters and training weather-specific sub-models can improve forecasting accuracy in offshore wind and photovoltaic systems (Rana, Koprinska, and Agelidis, 2016). Yet, these approaches remain confined to single-site optimisation and have not been leveraged to enable cross-site knowledge transfer. To the best of our knowledge, no study has combined meteorological clustering with transfer learning to selectively transfer knowledge across wind farms based on shared atmospheric dynamics.

In this paper, we propose a transfer learning framework that bridges this gap. We investigate whether offshore wind farms across Europe can achieve high forecasting accuracy using limited historical data, emulating the constraints of newly commissioned sites. By learning a domain-agnostic representation of offshore meteorological patterns across Northern Europe, we cluster observations solely based on meteorological similarity rather than geographic location. This enables the training of specialised forecasting models for distinct weather regimes. Data from a new target farm are mapped to the most similar meteorological cluster, allowing the corresponding pre-trained model to be efficiently fine-tuned with minimal site-specific data. The remainder of this paper is organised as follows. Section 2 describes the datasets and methodology. Section 3 presents and analyses the experimental results, and Section 4 provides further discussion.



**Figure 1.** Overview of the Methodology.

## 2. Methodology

### 2.1. Dataset

We utilise the offshore wind farm dataset introduced by Grothe, Kächele, and Watermeyer (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.

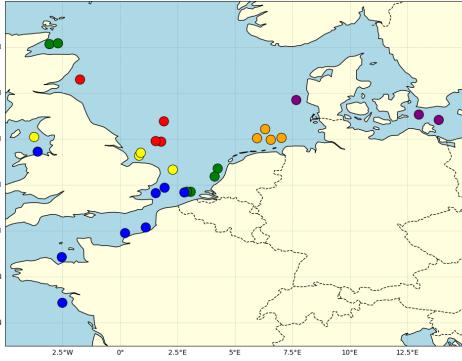
### 2.2. Selecting Source Wind Farms

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, Healy, and Melville, 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  $p$ .



(a) Distribution of wind farm clusters



(b) Wind farms selected for source models

**Figure 2.** Location and cluster assignment for available wind farms.

Each period  $i$  is represented as a multivariate time-series  $\mathbf{X}_i \in \mathbb{R}^{p \times d}$ , with  $d = 6$  meteorological features: wind speed, sea surface roughness, wind direction components ( $\sin \theta$  and  $\cos \theta$  for a wind direction  $\theta$ ) 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  $q_\phi(\mathbf{z} \mid \mathbf{X})$  mapping each input time-series to an 8-dimensional latent vector  $\mathbf{z}_i \in \mathbb{R}^8$ . 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,  $\mu_\phi(\mathbf{X}) \in \mathbb{R}^8$  and  $\log \sigma_\phi^2(\mathbf{X}) \in \mathbb{R}^8$ . Latent samples are generated via the reparametrisation trick  $\mathbf{z} = \mu_\phi(\mathbf{X}) + \sigma_\phi(\mathbf{X}) \odot \epsilon$ , where  $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  and  $\odot$  denotes element-wise multiplication. The decoder  $p_\theta(\mathbf{X} \mid \mathbf{z})$  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  $\theta$  and  $\phi$  are optimised by minimising the  $\beta$ -VAE loss function (Higgins et al., 2017) with mean squared error reconstruction loss and KL divergence regularisation:

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

Training is run for up to 200 epochs using AdamW optimisation. To mitigate posterior collapse, we employ a monotonic annealing schedule for  $\beta$ , 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, Koprinska, and Agelidis (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 measure cluster cohesion and separation, the normalised Davies-Bouldin index ( $\tilde{D}$ ) (Dl, 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 JA, 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 Jennen-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:

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

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:

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

## 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  $\phi_S$  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  $\mathbf{X}_T$  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  $\theta_k$  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

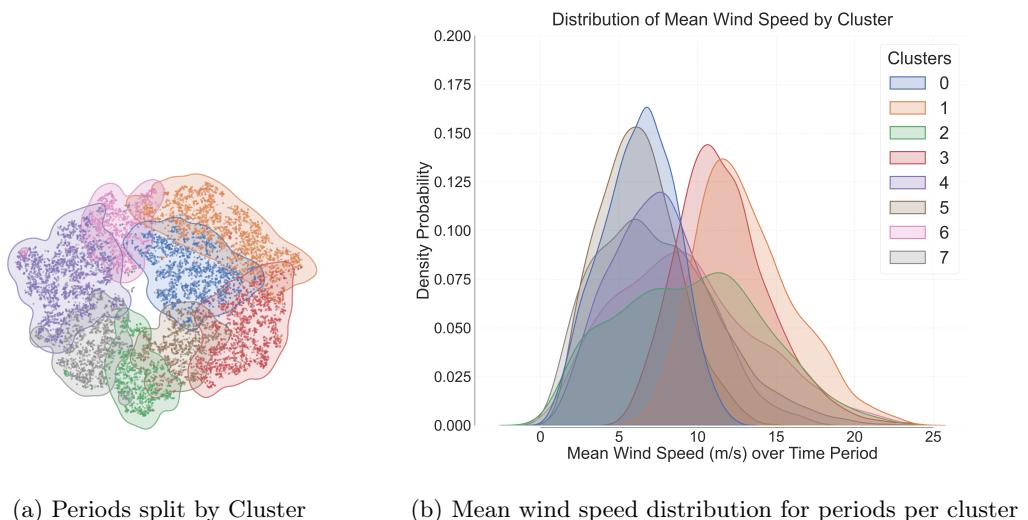
For performance evaluation, three primary metrics are employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the coefficient of Determination ( $R^2$ ). To allow for comparison across wind farms with varying capacities, MAE is expressed as a percentage of each farm’s total capacity. When 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 M3STIN<sup>1</sup> 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.

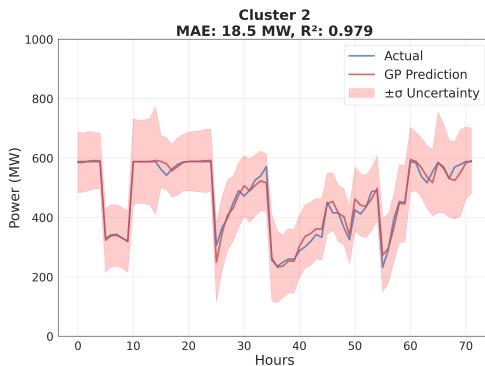
Cluster	Test samples	MAE (%)	
		GP	RF
0	2238	3.36	4.01
1	2425	3.39	3.71
2	2028	3.72	4.21
3	2315	3.74	4.25
4	2842	3.39	3.87
5	1622	4.02	4.55
6	2275	3.50	3.92
7	1740	3.06	3.32
<b>Overall</b>	<b>17,485</b>	<b>3.51</b>	<b>3.94</b>

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

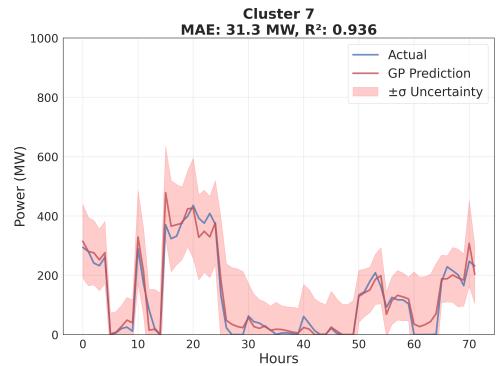


**Figure 4.** Metrics for GP forecasts, disaggregated by wind farm and cluster..

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.



(a) Forecast for random 72hrs for Cluster 0



(b) Forecast for random 72hrs for Cluster 7

**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 <sup>2</sup>	MAE (%)	R <sup>2</sup>	MAE (%)	R <sup>2</sup>	MAE (%)	R <sup>2</sup>	MAE (%)	R <sup>2</sup>
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.

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

**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%

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, further reducing the operational data dependency for new target farms. 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.

## References

- Ally, Stijn et al. (2025). "Modular deep learning approach for wind farm power forecasting and wake loss prediction". In: *Wind Energy Science* 10.4, pp. 779–812. doi: 10.5194/wes-10-779-2025. URL: <https://doi.org/10.5194/wes-10-779-2025>.
- Caliński, Tadeusz and Harabasz JA (Jan. 1974). "A Dendrite Method for Cluster Analysis". In: *Communications in Statistics - Theory and Methods* 3, pp. 1–27. doi: 10.1080/03610927408827101.
- Dl, Davies (1979). "A cluster separation measure". In: *IEEE Trans Pattern Anal Mach Intell* 1, pp. 224–227.
- Energy for Generations (2025). *Offshore wind targets underpin acceleration to 2030 and beyond*. URL: <https://ember-energy.org/latest-insights/offshore-wind-targets-underpin-acceleration-to-2030-and-beyond/>.
- ESMAP (2019). *Going Global: Expanding Offshore Wind to Emerging Markets*. Tech. rep. World Bank Group and Energy Sector Management Assistance Program (ESMAP). URL: <https://documents.worldbank.org/curated/en/716891572457609829/pdf/Going-Global-Expanding-Offshore-Wind-To-Emerging-Markets.pdf>.
- Global Wind Energy Council (2025). *Global Wind Report 2025*. Tech. rep. Global Wind Energy Council. URL: <https://www.gwec.net/reports/globalwindreport>.
- Grothe, Oliver, Fabian Kächele, and Mira Watermeyer (2022). "Analyzing Europe's biggest offshore wind farms: a data set with 40 years of hourly wind speeds and electricity production". In: *Energies* 15.5, p. 1700.
- Higgins, Irina et al. (2017). "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *International Conference on Learning Representations*. URL: <https://openreview.net/forum?id=Sy2fzU9gl>.
- Karim, Fazole et al. (2018). "LSTM Fully Convolutional Networks for Time Series Classification". In: *IEEE Access* 6, pp. 1662–1669. issn: 2169-3536. doi: 10.1109/access.2017.2779939. URL: <http://dx.doi.org/10.1109/ACCESS.2017.2779939>.
- Kingma, Diederik P. and Max Welling (2014). "Auto-Encoding Variational Bayes". In: *ICLR*.
- Li, Dan et al. (May 2024). "A novel transfer learning strategy for wind power prediction based on TimesNet-GRU architecture". In: *Journal of Renewable and Sustainable Energy* 16. doi: 10.1063/5.0200518.
- Lin, J. (1991). "Divergence measures based on the Shannon entropy". In: *IEEE Transactions on Information Theory* 37.1, pp. 145–151. doi: 10.1109/18.61115.
- Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605. URL: <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- McInnes, Leland, John Healy, and James Melville (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv: 1802.03426 [stat.ML]. URL: <https://arxiv.org/abs/1802.03426>.
- Pan, Simo Jialin and Qiang Yang (2010). "A Survey on Transfer Learning". In: *IEEE Transactions on Knowledge and Data Engineering* 22.10, pp. 1345–1359. doi: 10.1109/TKDE.2009.191.
- Rana, Md, Irena Koprinska, and Vassilios Agelidis (July 2016). "Solar Power Forecasting Using Weather Type Clustering and Ensembles of Neural Networks". In: doi: 10.1109/IJCNN.2016.7727853.
- Rousseeuw, Peter J. (1987). "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics*.
- Sajol, Md Saiful Islam et al. (2024). "Wind Power Prediction across Different Locations using Deep Domain Adaptive Learning". In: *2024 6th Global Power, Energy and Communication Conference (GPECOM)*. IEEE, pp. 518–523.
- Statista (2025). *Number of offshore wind farms worldwide as of April 2025, by country*. URL: <https://www.statista.com/statistics/264257/number-of-offshore-wind-farms-worldwide-by-country/>.
- Wang, Zhongrui et al. (2025). "Short-term offshore wind power multi-location multi-modal multi-step prediction model based on Informer (M3STIN)". In: *Energy* 322, p. 135616. issn: 0360-5442. doi: <https://doi.org/10.1016/j.energy.2025.135616>. URL: <https://www.sciencedirect.com/science/article/pii/S0360544225012587>.
- Ward, J. H. Jr. (1963). "Hierarchical Grouping to Optimize an Objective Function". In: *Journal of the American Statistical Association* 58.
- Warren Liao, T. (2005). "Clustering of time series data—a survey". In: *Pattern Recognition* 38.11. issn: 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2005.01.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0031320305001305>.
- Williams, Christopher KI and Carl Edward Rasmussen (2006). *Gaussian Processes for Machine Learning*. Vol. 2. 3. MIT press Cambridge, MA.