PROBABILISTIC WIND POWER FORECASTING VIA NON-STATIONARY GAUSSIAN PROCESSES

Domniki Ladopoulou[1] * **Dat Minh Hong**[1] **Petros Dellaportas**[1,2]
[1] Department of Statistical Science, University College London, London, United Kingdom
[2] Department of Statistics, Athens University of Economics and Business, Athens, Greece

ABSTRACT

Accurate probabilistic forecasting of wind power is essential for maintaining grid stability and enabling efficient integration of renewable energy sources. Gaussian Process (GP) models offer a principled framework for quantifying uncertainty; however, conventional approaches rely on stationary kernels, which are inadequate for modeling the inherently non-stationary nature of wind speed and power output. We propose a non-stationary GP framework that incorporates the generalized spectral mixture (GSM) kernel, enabling the model to capture time-varying patterns and heteroscedastic behaviors in wind speed and wind power data. We evaluate the performance of the proposed model on real-world SCADA data across short-, medium-, and long-term forecasting horizons. Compared to standard radial basis function and spectral mixture kernels, the GSM-based model outperforms, particularly in short-term forecasts. These results highlight the necessity of modeling non-stationarity in wind power forecasting and demonstrate the practical value of non-stationary GP models in operational settings.

Keywords Probabilistic forecasting · wind energy · Gaussian processes · non-stationary kernels · generalized spectral mixture kernel · renewable energy forecasting · SCADA data · uncertainty quantification

1 Introduction

Wind power plays a pivotal role in the global transition toward renewable energy. By the end of 2024, the total installed wind power capacity worldwide had reached approximately 1,136 gigawatts (GW), representing an annual increase of 117 GW compared to 2023 [1]. However, the variable and intermittent nature of wind generation introduces significant operational challenges to power systems. Therefore, accurate probabilistic wind power forecasting is essential to ensure grid stability and efficient resource planning. Recent reviews have highlighted the growing importance of hybrid models, probabilistic methods, and deep learning to advance wind power forecasting techniques [2]. In particular, Lagos *et al.* [3] performed a literature review demonstrating the increasing scientific research toward probabilistic forecasting methods, hybrid modeling approaches, and uncertainty quantification techniques in the wind energy sector. We focus exclusively on probabilistic forecasting approaches, which generate full predictive distributions rather than point estimates. This allows for uncertainty quantification, which is essential for risk-aware decision-making in power system operations.

Gaussian Process (GP) methods have been extensively utilized for wind power forecasting. Rogers *et al.* [4] proposed a heteroscedastic GP model that captures the input-dependent variance of wind turbine outputs, while Pandit *et al.* [5] incorporated turbine operational variables to enhance GP-based forecasts. Chen *et al.* [6] integrated GP with numerical weather prediction outputs, achieving improved forecast accuracy. However, standard GPs rely on default stationary kernels, which are inadequate for capturing the complex dynamics of wind data. Wind speed and power output often exhibit pronounced non-stationary behavior due to seasonal variations, atmospheric turbulence and local weather dynamics. As such, stationary kernels are not appropriate to model these fluctuations effectively.

To address challenges related to non-Gaussian uncertainties and time-varying characteristics, Kou *et al.* [7] proposed a sparse online warped GP model capable of adaptively learning non-Gaussian probabilistic distributions with reduced

* *Corresponding author*: **domna.ladopoulou.22@ucl.ac.uk**

computational cost. Nevertheless, their approach primarily focuses on output warping and online adaptation, while maintaining a stationary covariance structure.

Quantile regression-based methods have seen wide application. Yu *et al.* [8] proposed spatiotemporal quantile regression for regional wind power probabilistic forecasting. Later, Yu *et al.* [9] introduced a deep quantile regression model to better capture non-linear dependencies. Zhou *et al.* [10] proposed composite conditional non-linear quantile regression to improve the accuracy of prediction intervals for very short-term regional wind power forecasting. Other probabilistic modeling approaches have also emerged: Liao *et al.* [11] applied generative moment matching network to create realistic probabilistic scenarios without assuming specific distribution forms, while Zhang *et al.* [12] proposed an improved deep mixture density network to represent multimodal output distributions for wind power.

Several classical statistical approaches have also been introduced. Carpinone *et al.* [13] adopted Markov chain models for very short-term probabilistic forecasting. Dong *et al.* [14] utilized multiclass autoregressive moving average modeling for wind power prediction. Ma *et al.* [15] explored empirical dynamic modeling to reconstruct dynamic system behaviors directly from observed data.

Hybrid machine learning approaches are increasingly popular. Huang *et al.* [16] proposed an a priori-guided and data-driven hybrid framework to combine physical insights with data-driven learning. Gu *et al.* [17] proposed a hybrid framework combining fuzzy C-means clustering, whale optimization algorithm-optimized extreme learning machine for wind power forecasting, and Gaussian mixture models for uncertainty quantification. Wang *et al.* [18] developed a hybrid deep neural network framework focused on improving probabilistic forecasts. Transfer learning for wind forecasting has been explored by Liu and Wang [19], who applied multi-layer extreme learning machines to handle situations with limited training data. In addition, Fiocchi *et al.* [20] proposed a probabilistic multilayer perceptron trained on SCADA data with heteroscedastic outputs and transfer learning across turbines for condition monitoring, demonstrating improved performance over other probabilistic models.

Dong *et al.* [21] used spatiotemporal convolutional networks for forecasting the outputs of multiple wind farms. Krannichfeldt *et al.* [22] proposed an online ensemble approach for probabilistic wind power forecasting, capable of dynamically adapting to data changes. Eikeland *et al.* [23] developed a deep learning-based probabilistic forecasting model suitable for Arctic regions characterized by complex topography. Zhang *et al.* [24] proposed a multi-source temporal attention network that integrates heterogeneous numerical weather prediction data and historical measurements using temporal attention mechanisms to improve regional wind power probabilistic forecasting.

Che *et al.* [25] proposed a spatial-temporal probabilistic forecasting method based on multi-scale feature extraction and dynamic feature weighting, achieving improved performance in ultra-short-term wind power prediction settings. Meanwhile, Zhang *et al.* [26] developed a hybrid intelligent framework that integrates deterministic and probabilistic prediction approaches to enhance the accuracy and reliability of wind power forecasting.

It is important to note that while all the studies discussed adopt probabilistic forecasting methodologies, they do not focus on the same forecasting horizons. Wind power forecasting is typically categorized into different horizons according to application requirements. Long-term forecasting, spanning several months to years, supports strategic planning and wind farm development. Medium-term forecasting, covering several days up to a week, is useful for maintenance scheduling and operational planning. Short-term and ultra-short-term forecasting, ranging from a few hours to minutes ahead, plays a critical role in real-time grid balancing and dispatch.

While the aforementioned approaches have advanced probabilistic and deep learning methods for wind forecasting, the challenge of modeling non-stationarity within the GP framework remains largely unaddressed. Most existing GP-based methods rely on stationary kernels or limited modifications that fail to capture time-varying input–output relationships. In contrast, our proposed model explicitly incorporates a non-stationary covariance function, enabling appropriate modeling of the dynamic changes in the input–output relationship across varying wind conditions and timescales.

This work presents a probabilistic wind power forecasting framework that operates across multiple timescales, including long-term, medium-term, and short-term horizons. Using real-world supervisory control and data acquisition (SCADA) data, we address fundamental modeling limitations in prior work by directly modeling the dynamic, non-stationary nature of wind energy systems. The main contributions of the paper are as follows:

- A non-stationary GP framework is proposed for probabilistic wind power forecasting, explicitly modeling non-stationarity at the covariance level for wind speed and power time series.
- The necessity of modeling non-stationarity is demonstrated through improved point prediction accuracy and probabilistic scores, outperforming stationary GP baselines.
- We benchmark the model under realistic SCADA constraints and pre-processing steps, reflecting deployment conditions in operational wind farms.

- We demonstrate consistent performance improvements across long-, medium-, and short-term horizons on real SCADA data.

The remainder of this paper is structured as follows. Section 2 describes the proposed non-stationary GP framework and details the baseline models used for comparison. Section 3 outlines the experimental setup, including data sources and forecasting horizons, and presents the results of our comparative evaluation. Section 4 summarizes the main findings.

2 Methodology

We propose a non-stationary GP framework for probabilistic wind power forecasting, designed to model the non-stationary dynamics of wind speed and power output. Unlike conventional GP approaches that rely on stationary kernels, our method incorporates a non-stationary covariance structure, allowing the model to adapt to the peculiarities of the problem.

In this section, we first review the foundations of GP regression, followed by a description of the baseline kernels and the proposed non-stationary kernel. We then present the evaluation metrics used to assess model performance across both short-term and long-term forecasting horizons. In Section 3, we conduct a comparative evaluation using real-world SCADA data.

2.1 Gaussian process (GP) regression

In GP regression, we are given a data vector $\mathbf{y} = \{y_i\}_{i=1}^N \in \mathbb{R}^N$ whose entries are noisy evaluations of some function $f(\cdot)$ on a collection of D-dimensional vectors $X = \{\mathbf{x}_i\}_{i=1}^N \in \mathbb{R}^{N\times D}$, i.e. y_i is a noisy observation of $f(\mathbf{x}_i)$. We further assume that the noise $y_i - f(\mathbf{x}_i)$ is independent Gaussian with mean 0 and variance σ^2. Moreover, we place a GP prior over $f(\cdot)$, with mean function $\mu(\cdot)$ and covariance kernel $k_\theta(\cdot,\cdot)$, so that the collection of function values $f(X) := [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)]^T$ has a joint Gaussian distribution

$$f(X) \sim \mathcal{N}(\mu(X), K(X,X)),$$

where $\mu(X) = [\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_N)]^T$, and $K(X,X)_{ij} = k_\theta(\mathbf{x}_i, \mathbf{x}_j)$, for all i, j. Setting $A = K(X,X) + \sigma^2 I_N$, the log-marginal likelihood of the data becomes

$$\log p(\mathbf{y}|X) = -\frac{1}{2}(\mathbf{y} - \mu(X))^T A^{-1}(\mathbf{y} - \mu(X)) - \frac{1}{2}\log|A| - \frac{N}{2}\log(2\pi)$$

and the future observations y^* with covariates X^* have a normal conditional distribution with mean and variance given by

$$\begin{aligned} E(y^*|\mathbf{y}) &= \mu(X^*) + K(X^*,X)A^{-1}(\mathbf{y} - \mu(X)), \\ V(y^*|\mathbf{y}) &= K(X^*,X^*) - K(X^*,X)A^{-1}K(X,X^*). \end{aligned}$$

2.2 Kernel design and selection

The choice of kernel function is crucial for the performance of GP regression, influencing function smoothness, periodicity and generalization. In the context of wind power forecasting, where complex and non-stationary patterns are present, selecting an appropriate kernel becomes even more critical. In this section, we describe the baseline kernel functions used in our experiments, followed by the proposed non-stationary kernel.

2.2.1 Radial basis function kernel

We first used the radial basis function (RBF) kernel, also known as the squared exponential kernel, as a baseline. The RBF kernel is defined as:

$$k_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j|\theta) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d} - x_{j,d})^2}{\ell_d^2}\right),$$

where $\theta = \{\sigma_f^2, \ell_1, \ell_2, \ldots, \ell_D\}$ represents the kernel hyperparameters with ℓ_d denoting the length-scale parameters and σ_f^2 representing the output variance.

2.2.2 Spectral mixture kernel

To introduce greater flexibility, we considered the Spectral Mixture (SM) kernel [27], which decomposes a stationary kernel into spectral densities using Bochner's theorem [28]. To express the kernel in terms of the relative location between inputs, we define $\boldsymbol{\tau}$ as the difference between two input vectors, $\boldsymbol{\tau} = \mathbf{x}_i - \mathbf{x}_j \in \mathbb{R}^D$. This ensures that the kernel is stationary, which means that it depends only on the relative distance between the inputs rather than their absolute locations. With this definition, the SM kernel is given by:

$$k_{\text{SM}}(\boldsymbol{\tau}|\theta) = \sum_{q=1}^{Q} w_q \cos(2\pi \boldsymbol{\mu}_q^T \boldsymbol{\tau}) \prod_{d=1}^{D} \exp(-2\pi^2 \tau_d^2 v_q^{(d)}),$$

where Q is the number of mixture components in the spectral mixture, w_q represents the relative contribution of the q-th mixture component, $\boldsymbol{\mu}_q$ is a D-dimensional vector of spectral means, τ_d is the difference along the d-th dimension between the two inputs and $v_q^{(d)}$ represents the spectral variance of the q-th component along dimension d, controlling how quickly correlations decay along each input dimension. In addition, $\boldsymbol{\mu}_q^{-1}$ represents the periodicity of each component and $\nu_q^{(d)^{-1/2}}$ represents the length scale of the q-th component along dimension d, describing how rapidly function values change with respect to inputs $\mathbf{x}$.

A key trade-off of the increased flexibility provided by the SM kernel is the larger number of hyperparameters. Specifically, each mixture component introduces three hyperparameters, w_q, $\boldsymbol{\mu}_q$ and $v_q^{(d)}$. Since the kernel is defined as a sum over Q mixture components, this results in a total of $Q \times 3$ hyperparameters, in addition to the noise variance parameter. This high-dimensional parameter space increases the risk of overfitting in regression models and makes the optimization process more sensitive to initialization compared to standard kernels with fewer hyperparameters. An approach to alleviate this challenge is to train the GP model with multiple initializations and select the best-performing result, thereby exploring a broader range of local optima.

While the SM kernel offers greater flexibility than standard kernels, it still assumes stationarity—that is, it models spectral properties as constant across the input space. This can be a limiting assumption in real-world scenarios where the underlying function exhibits non-stationary behavior. In such settings, patterns in wind speed and power output can vary significantly over time due to changing environmental conditions, making a stationary model insufficient.

2.2.3 Generalized spectral mixture kernel

To address this limitation, we incorporate the Generalized Spectral Mixture (GSM) kernel [29], a non-stationary extension of the SM kernel that allows input-dependent hyperparameters. The GSM kernel extends the SM kernel by allowing the hyperparameters w_q, v_q, and μ_q to be a function of an input index $x \in \mathbb{R}$. Each hyperparameter can be written as $w_q(x)$, $v_q(x)$ and $\mu_q(x)$. We set $l_q(x) = v_q(x)^{-1/2}$ for notation convenience.

These functions are assigned independent GP priors with a zero mean function and a covariance matrix given by RBF kernels as follows:

$$\begin{aligned}
\log w_q(x) &\sim \mathcal{GP}(0, k_w(x, x')), \\
\log l_q(x) &\sim \mathcal{GP}(0, k_l(x, x')), \\
\text{logit}\mu_q(x) &\sim \mathcal{GP}(0, k_\mu(x, x')).
\end{aligned}$$

The resulting GSM kernel is given by:

$$k_{\text{GSM}}(x_i, x_j) = \sum_{q=1}^{Q} w_q(x_i) w_q(x_j) k_{\text{Gibbs},q}(x_i, x_j) \times \cos\left(2\pi(\mu_q(x_i)x_i - \mu_q(x_j)x_j)\right),$$

where the Gibbs kernel [30] is defined as:

$$k_{\text{Gibbs},q}(x_i, x_j) = \sqrt{\frac{2 l_q(x_i) l_q(x_j)}{l_q(x_i)^2 + l_q(x_j)^2}} \exp\left\{-\frac{(x_i - x_j)^2}{l_q(x_i)^2 + l_q(x_j)^2}\right\}.$$

The Gibbs kernel acts as a non-stationary counterpart to the squared exponential kernel, incorporating input-dependent length scales. The GSM kernel above is written for inputs in $\mathbb{R}$, but we can generalize it to multi-dimensional inputs

$\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^D$ using a product decomposition across dimensions:

$$k_{\text{GSM}}(\mathbf{x}_i, \mathbf{x}_j|\theta) = \prod_{d=1}^{D} k_{\text{GSM},d}(x_{i,d}, x_{j,d}|\theta_d),$$

where $\theta = (\theta_1, \ldots, \theta_D)$ represents the input dependent hyperparameters for each mixture $\theta_d = (w_{qd}, l_{qd}, \mu_{qd})_{i=1}^{Q}$ of the N-dimensional realizations $w_{qd}, l_{qd}, \mu_{qd} \in \mathbb{R}^N$ per dimension d.

Due to the complex hierarchical structure of GPs with the GSM kernel, we can no longer perform maximum marginal likelihood inference directly. This is because computing the marginal likelihood requires integrating out not only the latent function f but also all input-dependent hyperparameter functions, making the integral intractable. As an alternative, we employ maximum a posteriori (MAP) inference, where we maximize the log-posterior $\log p(\theta|\mathbf{y}) \propto \log p(\mathbf{y}|\theta) + \log p(\theta)$, where $p(\mathbf{y}|\theta)$ is the marginal likelihood, integrating out only the function values f. This allows us to perform gradient-based optimization, using numerical optimizers such as ADAM.

The final loss function is given by:

$$\mathcal{L}(\theta) = \log\left(\mathcal{N}_{\mathbf{y}} \prod_{q,d=1}^{Q,D} \mathcal{N}_{w_{qd}} \mathcal{N}_{\mu_{qd}} \mathcal{N}_{l_{qd}}\right),$$

where each component is

$$\begin{aligned}
\mathcal{N}_{\mathbf{y}} &= \mathcal{N}(\mathbf{y} \mid 0, K_\theta + \sigma_n^2 I) \\
\mathcal{N}_{w_{qd}} &= \mathcal{N}(w_{qd} \mid 0, K_{w_d}) \\
\mathcal{N}_{\mu_{qd}} &= \mathcal{N}(\mu_{qd} \mid 0, K_{\mu_d}) \\
\mathcal{N}_{l_{qd}} &= \mathcal{N}(l_{qd} \mid 0, K_{l_d}).
\end{aligned}$$

Similar to the SM kernel, the GSM kernel introduces trade-offs, including the risk of overfitting and sensitivity to local optima due to its many parameters. To mitigate this, we experiment with different numbers of mixtures and optimize with multiple initializations to improve convergence.

2.3 Probabilistic performance evaluation

A proper metric to evaluate the performance of our proposed modeling framework is the negative log-predictive density (NLPD), as it assesses both the mean and variance of the predictions by leveraging the probabilistic nature of GPs. Lower NLPD means a more appropriately confident prediction capability of the GP model. The NLPD is calculated by taking the negative logarithm of the predictive probability density of the future observations given the model:

$$\text{NLPD} = -\frac{1}{N} \sum_{i=1}^{N} \log p(y_i|\mathbf{x}_i, \mu(\mathbf{x}_i), \sigma_i^2).$$

We evaluated the performance of RBF, SM and GSM kernels across three different training set sizes, using ten different initializations to mitigate the risk of falling into a local optimum and ensure fair comparisons.

2.4 Non-probabilistic performance evaluation

For the best-performing out-of-sample scenario presented in Table 1, several criteria were used to demonstrate the superiority of the proposed approach in forecasting wind power across multiple time steps both short-term and long-term. Non-probabilistic performance metrics are also reported for completeness and comparability with conventional benchmarks.

We base our comparison on the commonly used root mean square error (RMSE) and mean absolute error (MAE), defined as $\text{RMSE} = (n^{-1} \sum_{i=1}^{n} (y_i - \mu(\mathbf{x}_i))^2)^{1/2}$ and $\text{MAE} = n^{-1} \sum_{i=1}^{n} |y_i - \mu(\mathbf{x}_i)|$, respectively. In addition, we present the normalized mean absolute percentage error (NMAPE), defined as $\text{NMAPE} = n^{-1} \sum_{i=1}^{n} |[(y_i - \mu(\mathbf{x}_i))/C] \times 100|$ where C is the rated power of the wind turbine equal to 2050 kW. Although these metrics offer valuable insights into how well our model forecasts output power, they rely solely on the predicted means $\mu(\mathbf{x}_i)$ and are therefore beyond the primary focus of this work.

A more critical aspect of the problem under investigation is leveraging the probabilistic nature of the modeling perspective. Therefore, the out-of-sample negative log predictive density score, discussed in 2.3, plays a more significant role in evaluating the superiority of the proposed methodology.

Some of the metrics were computed cumulatively over time to evaluate long-term wind power forecasting, as shown in Fig. 3. However, an important challenge was to assess the performance of our method in short-term wind power forecasting. To evaluate short-term forecast performance, fixed one-hour windows were used to analyse how the metrics evolve across different lead times.

3 Experiments

3.1 SCADA Data Overview and Filtering

The study utilizes SCADA and event data recorded at 10-minute intervals from a Senvion MM92 wind turbine located at Kelmarsh wind farm in the UK [31]. The dataset spans from January 3, 2016, to July 1, 2021, comprising over 1.7 million records across 110 variables, including date-time, wind speed and power output.

An operational status and events file was used to filter the data in order to improve the accuracy and reliability of the wind power forecasting model. These files contain detailed information about the operational state of each wind turbine, including events such as technical failures, standbys, and warnings triggered by either operational or environmental factors. Data recorded during such events can distort the true relationship between wind speed and power output, leading to inaccurate model training. To mitigate this, periods corresponding to standbys, warnings, and operational stops were excluded. Additionally, data from the week leading up to each forced outage was removed to minimize the inclusion of unstable turbine behavior before failure.

The impact of this filtering process is illustrated in Fig. 1, where the two power curves, i.e. output power versus wind speed, are shown before and after data filtering.

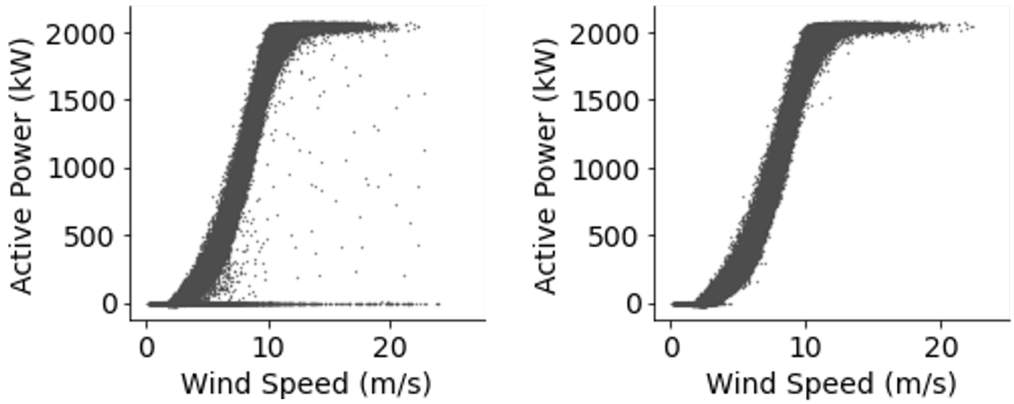

Figure 1: SCADA data power curve before and after removing standbys and warnings based on the operational status and event files. The left panel shows the raw data, while the right panel displays the filtered data used for analysis.

We used wind speed as the input for all our analyses, with the goal of expanding the training set and better assessing the performance of the GSM compared to the other two kernel types as the amount of training data increased.

3.2 Effectiveness of proposed model

In this section, we present the results of our proposed GP methodology using the GSM kernel, alongside the benchmark kernels introduced in Sections 2.2.1, 2.2.2: the RBF and SM kernels.

Three filtered SCADA data subsets were used to evaluate the performance of each model. The first scenario consists of $1,000$ training data points and $2,000$ test data points. The second scenario uses $5,000$ for training and $10,000$ for testing. To assess whether the proposed model benefits from increased training data, a third scenario was created with $7,000$ training points and $10,000$ test points.

For each model in every scenario, we perform hyperparameter optimization using 10 random initializations and report the average NLPD on the out-of-sample test set. This multiple-initialization strategy mitigates sensitivity to local

optima, particularly for the high-dimensional GSM kernel, which has a more complex parameter space than RBF or SM. The results are summarized in Table 1. Boxplots showing the average NLPD scores across all 10 initializations are presented in Fig. 2. As described in Section 2.2.3, we use marginal likelihood inference for the RBF and SM kernels, and MAP inference for the GSM kernel. The configuration of each GP prior is also shown in Table 1.

Table 1: Training and test set sizes, along with out-of-sample average negative log-predictive density (NLPD) scores for each scenario, averaged over 10 initializations. The best-performing model in each case is indicated in bold. The configuration of each GP prior is also included.

Training and Testing Sets			*Average NLPD*		
Scenario	*Training size*	*Testing size*	RBF	SM	GSM
1	1,000	2,000	7.95	7.95	**7.47**
2	5,000	10,000	7.59	7.59	**7.32**
3	7,000	10,000	7.20	7.23	**6.94**
GP Prior configuration			RBF	SM	GSM
No. of Mixtures			-	3	2
No. of Parameters			3	10	13

GSM: generalized spectral mixture; NLPD: negative log predictive density;
RBF: radial basis function; SM: spectral mixture.

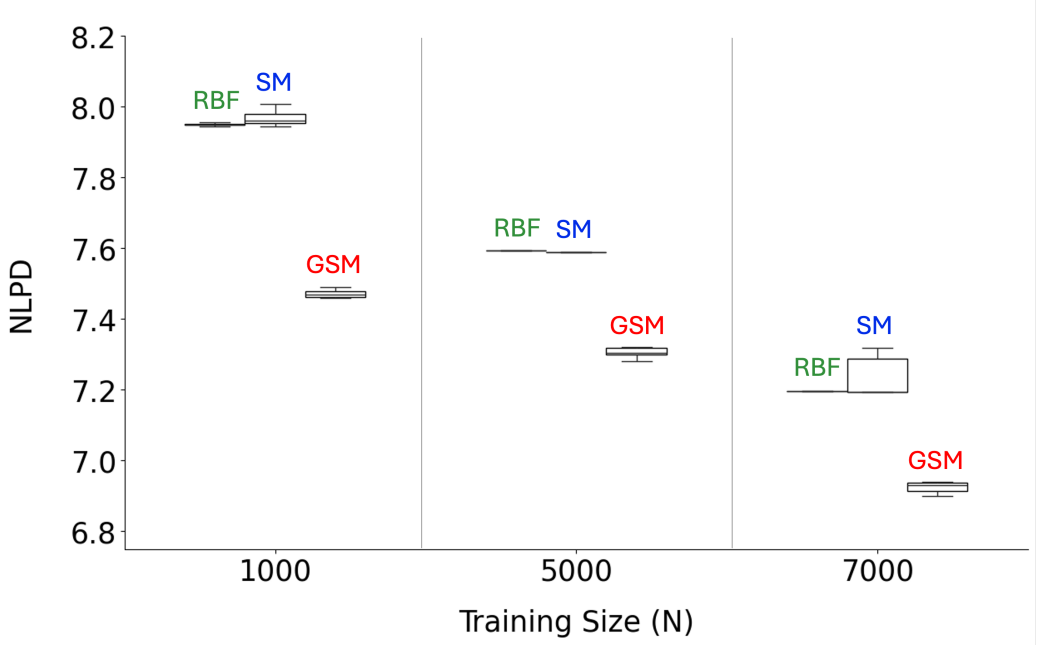

Figure 2: Comparison of out-of-sample average NLPD performance for the GSM, SM, and RBF kernels across three training and test set sizes over 10 initializations.

The results in Table 1 and Fig. 2 demonstrate that the proposed GSM kernel consistently outperforms the RBF and SM kernels across all training set sizes in terms of out-of-sample average NLPD. This indicates that the non-stationary properties captured by GSM are beneficial for modeling wind power data.

As the training set size increases, the average NLPD scores improve for all kernels, reflecting better generalization with more training data. However, the GSM kernel shows the largest improvement between scenarios 2 and 3, suggesting that it benefits more from additional training data than its stationary counterparts.

3.3 Forecasting comparative results

We conduct a comparative analysis of the three kernel functions using the models trained on $7,000$ observations, evaluating their extrapolation performance across multiple forecasting horizons: short-term, medium-term, and long-term. Results for the medium- and long-term horizons are shown in Fig.3, while short-term results are summarized in Table 2. In terms of RMSE, the RBF, SM, and GSM kernels exhibit comparable performance across medium- and long-term horizons, with minimal variation. However, based on the NLPD metric, the GSM kernel consistently outperforms the others, indicating improved uncertainty quantification.

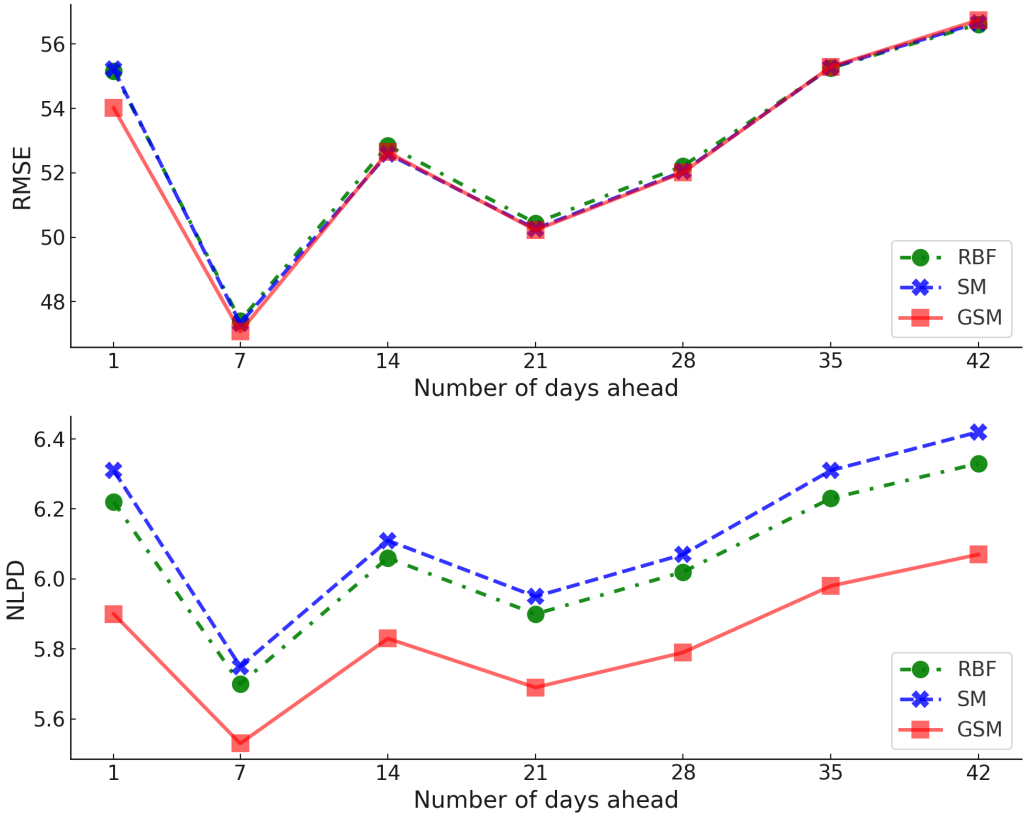

Figure 3: Average RMSE and NLPD performance of RBF, SM and GSM kernels over increasing time horizons of predictions for the best performing third scenario.

Of particular interest is the short-term forecasting performance, which is especially relevant for real-time grid stability and operational applications. Table 2 presents the out-of-sample results over a 24-hour horizon, with forecasts evaluated in hourly intervals beyond the last training point. The table reports multiple performance metrics, including 10-minute MAE, RMSE, NLPD, and NMAPE averaged hourly.

From a probabilistic perspective, the GSM kernel consistently achieves lower NLPD scores across most of the hourly intervals, confirming its superiority in capturing predictive uncertainty. This advantage becomes particularly pronounced during periods of higher energy output, where more accurate uncertainty quantification is essential for reliable grid integration. In the few intervals where GSM does not outperform in terms of NLPD, the total energy produced per hour is comparatively low, indicating that low wind speed values were observed.

Moreover, the GSM kernel performs competitively or better across all non-probabilistic metrics (MAE, RMSE, and NMAPE), further highlighting the benefits of modeling non-stationarity in wind power time series. This demonstrates the necessity of the non-stationary GSM kernel within GP frameworks, especially in short-term forecasting, where wind speed dynamics vary rapidly and predictive uncertainty plays a critical role in operational decision-making.

Table 2: Out-of-sample performance metrics for different kernels in the short-term forecasting of active power output (measured in kW). Units for each metric are indicated in parentheses beside the metric names. The total energy output per hour (kWh) is also provided for each time window. The best-performing model in each case is shown in bold. NMAPE is expressed as a percentage of the rated power (2050 kW).

		Probabilistic metric			**Non-probabilistic metrics**								
Hours	**Energy**	**Average NLPD**			**RMSE (kW)**			**MAE (kW)**			**NMAPE (%)**		
ahead	**(kWh)**	RBF	SM	GSM	RBF	SM	GSM	RBF	SM	GSM	RBF	SM	GSM
(1, 2]	62.81	4.89	5.01	**4.86**	31.64	34.10	**31.21**	27.96	29.87	**27.91**	**1.36**	1.46	**1.36**
(2, 3]	38.68	4.38	**4.37**	4.41	14.69	15.22	**11.45**	12.53	12.05	**9.03**	0.61	0.59	**0.44**
(3, 4]	46.76	**4.34**	4.35	4.43	**12.06**	13.96	13.02	**9.68**	10.86	12.06	**0.47**	0.53	0.59
(4, 5]	11.66	4.85	4.91	**4.69**	30.63	31.75	**25.47**	27.61	27.77	**20.77**	1.35	1.35	**1.01**
(5, 6]	43.19	4.71	4.67	**4.60**	26.81	25.68	**22.09**	21.49	19.98	**17.07**	1.05	0.97	**0.83**
(6, 7]	18.63	4.43	4.42	**4.39**	16.86	17.17	**10.03**	15.10	15.15	**8.34**	0.74	0.74	**0.41**
(7, 8]	70.80	**4.30**	4.31	4.39	**9.45**	11.66	10.06	**8.11**	10.27	9.02	**0.40**	0.50	0.44
(8, 9]	50.94	4.28	**4.27**	4.40	**7.70**	8.69	10.69	**6.18**	6.75	9.34	**0.30**	0.33	0.46
(9, 10]	53.65	**4.62**	4.63	4.68	**24.24**	24.53	25.44	18.64	**18.37**	21.16	0.91	**0.90**	1.03
(10, 11]	81.40	4.33	**4.30**	4.46	12.01	**11.33**	14.84	10.14	**8.97**	13.66	0.49	**0.44**	0.67
(11, 12]	33.62	4.55	4.59	**4.50**	21.80	23.50	**17.21**	20.21	21.72	**13.23**	0.99	1.06	**0.65**
(12, 13]	209.36	7.66	7.72	**6.99**	72.44	71.43	**70.34**	58.38	57.05	**55.85**	2.85	2.78	**2.72**
(13, 14]	125.56	5.42	5.41	**5.15**	42.55	41.77	**38.84**	40.92	40.33	**37.02**	2.00	1.97	**1.81**
(14, 15]	50.97	4.65	4.73	**4.60**	25.09	27.51	**21.97**	23.86	26.26	**19.58**	1.16	1.28	**0.96**
(15, 16]	149.26	5.20	5.16	**5.03**	38.33	37.21	**35.84**	34.49	33.47	**31.50**	1.68	1.63	**1.54**
(16, 17]	265.84	5.95	5.94	**5.62**	51.27	50.09	**48.83**	43.61	41.92	**40.88**	2.13	2.04	**1.99**
(17, 18]	98.85	6.70	6.65	**6.23**	61.41	59.57	**59.39**	46.72	45.97	**44.15**	2.28	2.24	**2.15**
(18, 19]	138.56	9.10	9.11	**8.17**	86.34	84.49	**84.43**	75.87	74.85	**73.35**	3.70	3.65	**3.58**
(19, 20]	281.45	13.72	14.21	**11.97**	120.61	120.67	**119.19**	117.48	117.20	**116.09**	5.73	5.72	**5.66**
(20, 21]	500.31	9.44	9.81	**8.68**	**89.37**	90.32	89.89	77.67	77.53	**76.86**	3.79	3.78	**3.75**
(21, 22]	352.23	8.31	8.76	**7.72**	**79.08**	81.39	79.33	**69.51**	71.63	69.55	**3.39**	3.49	**3.39**
(22, 23]	627.13	11.92	12.49	**10.76**	**108.58**	109.84	109.38	**95.40**	97.12	97.09	**4.65**	4.74	4.74
(23, 24]	1032.00	7.15	7.11	**6.59**	66.85	64.99	**64.80**	50.07	48.33	**48.25**	2.44	2.36	**2.35**

GSM: generalized spectral mixture; MAE: mean absolute error; NMAPE: normalized mean absolute percentage error; NLPD: negative log predictive density; RBF: radial basis function; RMSE: root mean square error; SM: spectral mixture.

4 Conclusions

This work introduced a nonstationary GP framework for probabilistic wind power forecasting, employing the GSM kernel to model time-varying dynamics in wind speed and power output. Using real-world SCADA data, the proposed model demonstrated consistent improvements in both point prediction accuracy and uncertainty quantification compared to standard stationary kernels across short-, medium-, and long-term forecasting horizons.

The results underscore the necessity of incorporating non-stationary modeling within GP frameworks for realistic wind power forecasting. In particular, the GSM kernel provided better performance in short-term settings, where the non-stationarity of wind patterns is most pronounced.

References

[1] Global Wind Energy Council (GWEC). Global Wind Report 2025. `https://www.gwec.net/reports/globalwindreport`. gwec.net, accessed Apr. 28, 2025.

[2] Wen-Chang Tsai, Chih-Ming Hong, Chia-Sheng Tu, Whei-Min Lin, and Chiung-Hsing Chen. A review of modern wind power generation forecasting technologies. *Sustainability*, 15(14):10757, 2023.

[3] Ana Lagos, Joaquín E Caicedo, Gustavo Coria, Andrés Romero Quete, Maximiliano Martínez, Gastón Suvire, and Jesús Riquelme. State-of-the-art using bibliometric analysis of wind-speed and-power forecasting methods applied in power systems. *Energies*, 15(18):6545, 2022.

[4] TJ Rogers, P Gardner, N Dervilis, K Worden, AE Maguire, E Papatheou, and EJ Cross. Probabilistic modelling of wind turbine power curves with application of heteroscedastic gaussian process regression. *Renewable Energy*, 148:1124–1136, 2020.

[5] Ravi Kumar Pandit, David Infield, and Athanasios Kolios. Gaussian process power curve models incorporating wind turbine operational variables. *Energy Reports*, 6:1658–1669, 2020.

[6] Niya Chen, Zheng Qian, Ian T. Nabney, and Xiaofeng Meng. Wind power forecasts using gaussian processes and numerical weather prediction. *IEEE Transactions on Power Systems*, 29(2):656–665, 2014.

[7] Peng Kou, Feng Gao, and Xiaohong Guan. Sparse online warped gaussian process for wind power probabilistic forecasting. *Applied energy*, 108:410–428, 2013.

[8] Yixiao Yu, Xueshan Han, Ming Yang, and Jiajun Yang. Probabilistic prediction of regional wind power based on spatiotemporal quantile regression. In *2019 IEEE industry applications society annual meeting*, pages 1–16. IEEE, 2019.

[9] Yixiao Yu, Ming Yang, Xueshan Han, Yumin Zhang, and Pingfeng Ye. A regional wind power probabilistic forecast method based on deep quantile regression. *IEEE Transactions on Industry Applications*, 57(5):4420–4427, 2021.

[10] Yan Zhou, Yonghui Sun, Sen Wang, Rabea Jamil Mahfoud, Hassan Haes Alhelou, Nikos Hatziargyriou, and Pierluigi Siano. Performance improvement of very short-term prediction intervals for regional wind power based on composite conditional nonlinear quantile regression. *Journal of Modern Power Systems and Clean Energy*, 10(1):60–70, 2021.

[11] Wenlong Liao, Zhe Yang, Xinxin Chen, and Yaqi Li. Windgmmn: Scenario forecasting for wind power using generative moment matching networks. *IEEE transactions on artificial intelligence*, 3(5):843–850, 2021.

[12] Hao Zhang, Yongqian Liu, Jie Yan, Shuang Han, Li Li, and Quan Long. Improved deep mixture density network for regional wind power probabilistic forecasting. *IEEE Transactions on Power Systems*, 35(4):2549–2560, 2020.

[13] A Carpinone, Massimiliano Giorgio, Roberto Langella, and Alfredo Testa. Markov chain modeling for very-short-term wind power forecasting. *Electric power systems research*, 122:152–158, 2015.

[14] Yunxuan Dong, Shaodan Ma, Hongcai Zhang, and Guanghua Yang. Wind power prediction based on multi-class autoregressive moving average model with logistic function. *Journal of Modern Power Systems and Clean Energy*, 10(5):1184–1193, 2022.

[15] Jiayi Ma, Ming Yang, and You Lin. Ultra-short-term probabilistic wind turbine power forecast based on empirical dynamic modeling. *IEEE Transactions on Sustainable Energy*, 11(2):906–915, 2019.

[16] Yi Huang, Guo-Ping Liu, and Wenshan Hu. Priori-guided and data-driven hybrid model for wind power forecasting. *ISA transactions*, 134:380–395, 2023.

[17] Bo Gu, Hao Hu, Jian Zhao, Hongtao Zhang, and Xinyu Liu. Short-term wind power forecasting and uncertainty analysis based on fcm–woa–elm–gmm. *Energy Reports*, 9:807–819, 2023.

[18] Qianchao Wang, Lei Pan, Haitao Wang, Xinchao Wang, and Ying Zhu. Short-term wind power probabilistic forecasting using a new neural computing approach: Gmc-deepnn-pf. *Applied Soft Computing*, 126:109247, 2022.

[19] Yanli Liu and Junyi Wang. Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting. *Applied Energy*, 312:118729, 2022.

[20] Filippo Fiocchi, Domniki Ladopoulou, and Petros Dellaportas. Probabilistic multilayer perceptrons for wind farm condition monitoring. *Wind Energy*, 28(4):e70012, 2025. e70012 we.70012.

[21] Xiaochong Dong, Yingyun Sun, Ye Li, Xinying Wang, and Tianjiao Pu. Spatio-temporal convolutional network based power forecasting of multiple wind farms. *Journal of Modern Power Systems and Clean Energy*, 10(2):388–398, 2021.

[22] Leandro Von Krannichfeldt, Yi Wang, Thierry Zufferey, and Gabriela Hug. Online ensemble approach for probabilistic wind power forecasting. *IEEE Transactions on Sustainable Energy*, 13(2):1221–1233, 2021.

[23] Odin Foldvik Eikeland, Finn Dag Hovem, Tom Eirik Olsen, Matteo Chiesa, and Filippo Maria Bianchi. Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An arctic case. *Energy Conversion and Management: X*, 15:100239, 2022.

[24] Hao Zhang, Jie Yan, Yongqian Liu, Yongqi Gao, Shuang Han, and Li Li. Multi-source and temporal attention network for probabilistic wind power prediction. *IEEE Transactions on Sustainable Energy*, 12(4):2205–2218, 2021.

[25] Jinxing Che, Fang Yuan, Dewen Deng, and Zheyong Jiang. Ultra-short-term probabilistic wind power forecasting with spatial-temporal multi-scale features and k-fsdw based weight. *Applied Energy*, 331:120479, 2023.

[26] Jiawei Zhang, Rongquan Zhang, Yanfeng Zhao, Jing Qiu, Siqi Bu, Yuxiang Zhu, and Gangqiang Li. Deterministic and probabilistic prediction of wind power based on a hybrid intelligent model. *Energies*, 16(10):4237, 2023.

[27] Andrew Wilson and Ryan Adams. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, pages 1067–1075. PMLR, 17–19 Jun 2013.

[28] Salomon Bochner. *Lectures on Fourier Integrals.(AM- 42)*, volume 42. Princeton University Press, 1959.

[29] Sami Remes, Markus Heinonen, and Samuel Kaski. Non-stationary spectral kernels. In *31st Conference on Neural Information Processing Systems*, 2017.

[30] M. Gibbs. *Bayesian Gaussian Processes for Regression and Classification*. Phd thesis, University of Cambridge, 1997.

[31] Charlie Plumley. Kelmarsh wind farm data (0.0.3). *Zenodo*, February 2022.