

Ultra-Short-Term Probabilistic Wind Forecasting: Can Numerical Weather Predictions Help?

Feng Ye

*Department of Industrial & Systems Engineering
Rutgers University
Piscataway, USA
feng.ye@rutgers.edu*

Joseph Brodie

*AKRF Inc.
New York, USA
jbrodie@akrf.com*

Travis Miles

*Department of Marine & Coastal Sciences
Rutgers University
Piscataway, USA
tnmiles@marine.rutgers.edu*

Ahmed Aziz Ezzat[†]

*Department of Industrial & Systems Engineering
Rutgers University
Piscataway, USA
aziz.ezzat@rutgers.edu*

Abstract—Ultra-short-term wind forecasting (i.e. wind speed and power predictions issued for sub-hourly forecast horizons), are pivotal to the effective management and integration of wind farms into modern-day electricity systems. The dominant consensus in the forecasting literature and practice is that data-driven approaches may be best suited for such short-term horizons. This is in contrast to numerical weather predictions (NWP), or hybrid models thereof, for which the value is typically substantiated at relatively longer horizons (> 1 -3 hours). We propose a probabilistic data-driven model that actually makes use of NWP information (albeit indirectly) for ultra-short-term wind speed and power forecasting. Instead of directly using NWPs as input regressors (as in hybrid approaches), we indirectly invoke NWP information in selecting key parameters within the data science model, thereby guiding it to adhere to certain physical principles related to local wind field formation and propagation. We show that such indirect integration of NWPs within our data science model outperforms several prevalent forecasting methods, including but not limited to persistence forecasts, which are known to be highly competitive at ultra-short-term horizons. This work serves as an exemplar for leveraging the rich, yet coarser-resolution information of NWPs in benefiting data-science-based ultra-short-term wind forecasting models.

Index Terms—Spatio-temporal Learning, Ultra-short-term Wind Forecasting, Wind Energy.

I. INTRODUCTION

Ultra short-term wind forecasting refers to the prediction of wind speed and power for very short-term horizons (i.e., few minutes up to an hour ahead). The value of such ultra-short-term predictions stems from their high utility to a wide spectrum of critical wind farm and power system operations, including but not limited to economic dispatch and reserve planning [1], [2], asset management [3], [4], and control [5].

The methods for wind speed and power forecasting can be broadly classified based on whether they make use of numerical weather predictions (NWP) or not [6]. There is

an overall consensus in the wind forecasting literature and practice that invoking NWP information is valuable (and in fact, indispensable) for relatively longer forecast horizons, that is, for $h > 1$ -3 hours (h denotes the forecast horizon). For ultra-short-term horizons, however, data-driven models, primarily those based on statistical and machine learning (ML) are arguably regarded as the best approaches, mainly due to their ability to extrapolate patterns and correlations from sheer volumes of historical data into the near future [7]–[10].

Along this line, we propose a data science model for ultra-short-term wind speed and power forecasting. The key distinguishing feature of our model is its ability to leverage NWP information—which is typically available to the farm or power system operator—in order to benefit ultra-short-term wind forecasting. The proposed approach is based on a spatio-temporal Gaussian process (GP) [11], within which NWP information guides the selection of key kernel hyperparameters that encode information about the local wind flow in the region under study. This *indirect integration* of NWPs within a data-science-based forecasting model breaks away from the mainstream approach of hybrid forecasting, wherein NWPs are *directly integrated* as input regressors to data science models [12], [13], and are therefore likely to carry over the deficiencies and multi-type biases of NWPs when directly used for ultra-short-term forecast horizons [14].

We train and test our model using actual observations that have been recently collected in proximity to the offshore wind energy areas in the NY/NJ Bight, where several Gigawatt-scale projects are currently in-development [15]. We demonstrate that our approach achieves noticeable improvements, in terms of both wind speed and power forecasting, relative to several benchmarks in the forecasting literature and practice, including persistence forecasts, which are known to be highly competitive for very-short-term horizons. We therefore envision our work to serve as an exemplar for leveraging the rich, yet coarse-resolution information of NWPs within data-science-based ultra-short-term wind forecasting.

This work has been supported in part by the National Science Foundation (Award #: ECCS-2114422) and in part by an IIF-SAS research grant.

[†]Corresponding author; contact e-mail: aziz.ezzat@rutgers.edu

The remainder of this paper is organized as follows. Section II describes the real-world data used in this paper. Section III reviews the concept of spatio-temporal asymmetry for wind fields, which will then be used in deciphering the potential role of NWP in data-driven ultra-short-term forecasting. In Section IV, we introduce our proposed forecasting method, which is then followed by Section V where we present and discuss our results. Finally, Section VI concludes the paper.

II. DATA DESCRIPTION

Our dataset comprises 10-min wind speed observations, at 100-m altitude, obtained via two buoys (E05 and E06), that have been recently deployed by the New York State Energy Research & Development Authority (NYSERDA)—See the rose plot in Fig. 1(a). We also obtain a set of hourly wind velocity NWP from a meso-scale meteorological model operated by Rutgers University, called RU-WRF (short for the Rutgers University Weather Research & Forecasting model) [16], [17]. Both data and NWP outputs span the month of December 2019. Fig. 1(b) shows the histograms of the actual observations versus their correspondent NWP, while Fig. 1(c) shows a 12-day time series of the actual data versus statistically interpolated NWP for E05 (top) and E06 (bottom).

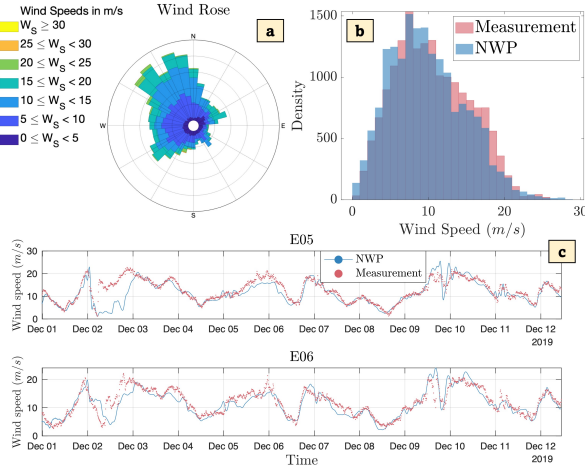


Fig. 1. (a) Wind rose plot for the actual wind speed measurements recorded in the NY/NJ Bight (Data from two buoys combined to produce this figure); (b) Histograms of actual wind speed measurements (mean = 10.52 m/s) versus NWP wind speed forecasts (mean = 9.87 m/s); (c) Actual wind speed observations (10-min) versus statistically interpolated NWP.

III. SPATIO-TEMPORAL DATA ANALYSIS

Let $Z(\mathbf{s}, t)$ denote a random process that varies over space-time, such that $\mathbf{s} \in \mathbb{R}^2$ denotes the location (in longitude and latitude) and t denotes time. A cornerstone of spatio-temporal models is to invoke a covariance function (often known as a kernel) that encodes the similarity between a pair of spatial-temporal observations and enables GP-based forecasting. Assuming (local) stationarity, this kernel is denoted as $C(\mathbf{h}, u) : \mathbb{R}^2 \times \mathbb{Z}^+ \rightarrow \mathbb{R}$, where $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = t_i - t_j$ are spatial and temporal lags, respectively. A prevalent way to model $C(\cdot, \cdot)$ in the data science literature is through the

so-called separable approach, wherein $C(\mathbf{h}, u)$ is expressed as $C(\mathbf{h}, u) = C^s(\mathbf{h}) \times C^t(u)$, such that $C^s(\mathbf{h})$ and $C^t(u)$ are two independent kernels for space and time, respectively [18]. Popular selections for $C^s(\mathbf{h})$ and $C^t(u)$ include the Gaussian, squared exponential, and Matérn kernels [11].

A key limitation of the separable approach is that it assumes, by design, that space-time correlations are symmetric, i.e. $\text{cor}\{Z(\mathbf{s}_i, t), Z(\mathbf{s}_{i'}, t + u)\} = \text{cor}\{Z(\mathbf{s}_{i'}, t), Z(\mathbf{s}_i, t + u)\}$. Processes that involve a flow over time (e.g., wind fields) typically violate that assumption, because the along-wind dependence (i.e., correlations in the direction of the flow) are typically stronger than opposite-wind dependence [18], [19].

To demonstrate this using our data, we use an estimator of asymmetry, expressed as in (1) [20], [21].

$$a(\mathbf{s}_i, \mathbf{s}_{i'}, u) := \delta(\mathbf{s}_i, \mathbf{s}_{i'}, u) - \delta(\mathbf{s}_{i'}, \mathbf{s}_i, u), \quad (1)$$

where $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ denote the coordinates of E05 and E06, respectively, and $\delta(\cdot, \cdot, \cdot)$ is the empirical spatio-temporal semi-variogram (a measure of dissimilarity between spatio-temporal observations), and is expressed as in (2).

$$\delta(\mathbf{s}_i, \mathbf{s}_{i'}, u) = \frac{1}{2(N - u - 1)} \sum_{k=1}^{N-u-1} \{y(\mathbf{s}_i, k + u) - y(\mathbf{s}_{i'}, k)\}^2. \quad (2)$$

In (2), N is the number of observations, $\delta(\mathbf{s}_i, \mathbf{s}_{i'}, u)$ means that the measurements taken at site $\mathbf{s}_{i'}$ are u time lag behind that at site $\mathbf{s}_i$, while $\delta(\mathbf{s}_{i'}, \mathbf{s}_i, u)$ means the measurements taken at site $\mathbf{s}_i$ are u time lag behind that at site $\mathbf{s}_{i'}$. Hence, if the wind is blowing from site $\mathbf{s}_{i'}$ towards site $\mathbf{s}_i$, then we should expect $\delta(\mathbf{s}_i, \mathbf{s}_{i'}, u) < \delta(\mathbf{s}_{i'}, \mathbf{s}_i, u)$, and therefore $a(\mathbf{s}_i, \mathbf{s}_{i'}, u) < 0$, indicating a lack of symmetry.

We then perform a t -test for each time lag, $u \in \{1, \dots, 36\}$ (in 10-min intervals), with $\mathcal{H}_0 : \bar{a}(\mathbf{s}_1, \mathbf{s}_2, u) = 0$, where $\bar{a}(\mathbf{s}_1, \mathbf{s}_2, u)$ is the average asymmetry at time lag u . Fig. 2 shows the values of $\bar{a}(\mathbf{s}_1, \mathbf{s}_2, u)$ versus the time lag, together with the test's 95% confidence intervals. The negative values shown in Fig. 2 suggests a noticeable asymmetry in along-wind versus opposite-wind dependence, as a result of the wind propagation across the prevailing westerly wind during this time of the year—Recall Fig. 1(a). We also find that the maximum asymmetry occurs at time lags of ~ 1 -3 hours, which is approximately the expected time for wind conditions to propagate from E06 towards E05.

The above analysis suggests the potential benefit of modeling asymmetry: When attempting to predict the wind conditions at a downstream location, then one may potentially assign higher weight to the observations recorded few hours ago at an upstream location, since those upstream (but past) measurements are expected to be highly correlated with their downstream counterparts at the current time. To enable this, we need an accurate representation of the prevailing wind flow (both magnitude and direction) at the time of the forecast. This is, in fact, where we plan to integrate NWP information. Details of this integration are discussed in Section IV.

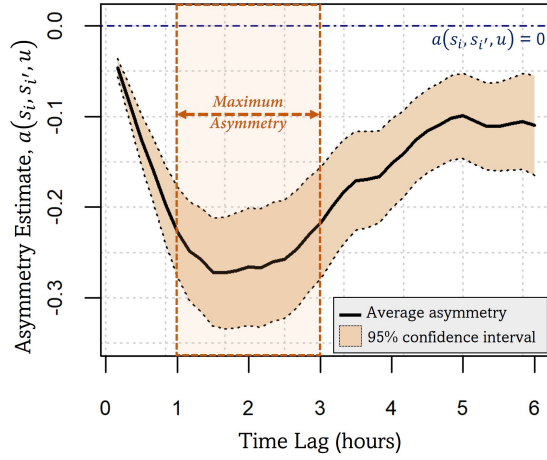


Fig. 2. Difference in empirical semi-variograms versus the time lag, along with 95% t -test confidence intervals. Noticeable asymmetry levels are observed, peaking at ~ 1 -3-hour time lag, which aligns with the expected duration for wind conditions to propagate across the wind field (the distance between E05 and E06 is 77 Km, while the average wind speed across both sites is 38 Km/hr).

IV. METHODOLOGY

We first introduce spatio-temporal Gaussian processes in Section IV-A, then discuss the role of NWP in Section IV-B.

A. Spatio-temporal Gaussian Processes (GPs)

Let $\mathbf{Z} = [z(\mathbf{s}_1, t_1), z(\mathbf{s}_1, t_2), \dots, z(\mathbf{s}_1, t_T), \dots, z(\mathbf{s}_n, t_T)]^T$ be a vector of spatio-temporal wind speeds, where $z(\mathbf{s}_i, t_j)$ is the wind speed at location $\mathbf{s}_i$ and time t_j . A GP model can be expressed as in (3).

$$z(\mathbf{s}_i, t_j) = m(\mathbf{s}_i, t_j) + \gamma(\mathbf{s}_i, t_j), \quad (3)$$

where $m(\mathbf{s}_i, t_j)$ is referred to as the GP mean function, which, for ultra-short-term forecasting, can be expressed as a constant, $m(\mathbf{s}_i, t_j) = \beta_0, \forall i, j$. The term $\gamma(\cdot, \cdot)$ is a zero-mean, spatio-temporal Gaussian random field, with an $nT \times nT$ covariance matrix denoted by $\Sigma + \delta \mathbf{I}$, where δ is the noise parameter, and $\mathbf{I}$ is the identity matrix. The entries of Σ are computed using the GP kernel, $C(\mathbf{h}, u)$ (details of which are to follow).

For a GP, the joint distribution of the training data $\mathbf{Z}$ and a set of testing data $\mathbf{Z}_*$ follows a multivariate Gaussian distribution, as shown in (4).

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{Z}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{bmatrix} \right), \quad (4)$$

where $\mathbf{m} = [m(\mathbf{s}_1, t_1), \dots, m(\mathbf{s}_n, t_T)]^T$ is the vector of mean function evaluations at the training data, and $\mathbf{m}_*$ is similarly defined for the testing data. The matrix Σ_* holds the covariance values between $\mathbf{Z}$ and $\mathbf{Z}_*$, while Σ_{**} denotes the covariance matrix of the testing data. The forecast distribution conditioning the joint Gaussian prior distribution on the observations can then be expressed as in (5).

$$\mathcal{P}(\mathbf{Z}_* | \mathbf{Z}, m(\cdot), \gamma(\cdot, \cdot)) \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\Sigma}), \quad (5)$$

such that $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are computed as in (6).

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \mathbf{m}_* + \Sigma_*^T (\Sigma + \delta \mathbf{I})^{-1} (\mathbf{Z} - \mathbf{m}) \\ \hat{\Sigma} &= \Sigma_{**} - \Sigma_*^T (\Sigma + \delta \mathbf{I})^{-1} \Sigma_* \end{aligned} \quad (6)$$

B. Modeling $C(\mathbf{h}, u)$ and The role of NWP

Defining the kernel $C(\mathbf{h}, u)$ is essential for spatio-temporal GPs. The analysis in Section III motivates the need for a nonseparable kernel that acknowledges the impact of the wind flow on the space-time correlations. Here, we adopt an under-explored class of nonseparable covariance models that can particularly capture asymmetric behavior in spatio-temporal data. Consider a spatial random field on $\mathbb{R}^2$ with a spatial, motion-invariant covariance function $C_s(\cdot)$. Now, let's assume this field moves over time with a random velocity vector $\Theta \in \mathbb{R}^2$, creating a spatial-temporal random process with an asymmetric covariance, $C_a(\mathbf{h}, u)$, expressed as in (7).

$$C_a(\mathbf{h}, u) = \mathbb{E}_{\Theta} \{ C_s(\mathbf{h} - \Theta u) \}. \quad (7)$$

Assuming $C_s(x) = \exp(-x^2)$ and letting $\Theta \sim \mathcal{N}(\boldsymbol{\tau}, \Psi)$ yields the closed-form expression in (8), which is referred to as Schlather's covariance model [22].

$$\begin{aligned} C_a(\mathbf{h}, u) &= \frac{1}{\sqrt{|\mathbf{I}_{2 \times 2} + 2\Psi u^2|}} \\ &\times \exp \left\{ -(\mathbf{h} - \boldsymbol{\tau} u)^T (\mathbf{I}_{2 \times 2} + 2\Psi u^2)^{-1} (\mathbf{h} - \boldsymbol{\tau} u) \right\}, \end{aligned} \quad (8)$$

where $|\cdot|$ denotes the matrix determinant.

The choice of $\boldsymbol{\tau}$ and Ψ (the parameters of the prevailing flow) is crucial for the effective use of $C_a(\mathbf{h}, u)$ in practice. Incorrect specifications of $\boldsymbol{\tau}$ and Ψ can severely limit (or even reverse) the benefits of an asymmetric approach. In prior works [21], [23], $\boldsymbol{\tau}$ and Ψ have been either pre-set, or estimated using historical measurements. We believe, however, that local measurements do not necessarily capture the prevailing flow characteristics, but rather are merely instantaneous representations of the wind velocity at a particular location and time. Instead, this work advocates the use of NWP wind velocity predictions—which are typically available to the farm or power system operator at the time of the forecast—as more meaningful representations of the prevailing flow. In particular, we estimate $\boldsymbol{\tau}$ and Ψ as in (9) and (10), respectively.

$$\boldsymbol{\tau} = [\tau_1, \tau_2]^T = [\bar{v}, \bar{w}]^T, \quad (9)$$

$$\Psi = \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} \\ \Psi_{2,1} & \Psi_{2,2} \end{bmatrix} = \begin{bmatrix} \text{cov}(\mathbf{v}, \mathbf{v}) & \text{cov}(\mathbf{v}, \mathbf{w}) \\ \text{cov}(\mathbf{w}, \mathbf{v}) & \text{cov}(\mathbf{w}, \mathbf{w}) \end{bmatrix}, \quad (10)$$

where $\mathbf{v} = [v_1, \dots, v_{T+h}]^T$ and $\mathbf{w} = [w_1, \dots, w_{T+h}]^T$ are the NWP outputs for the eastward and northward winds, respectively, during both the training and forecast horizon windows, whereas $\bar{v}$ and $\bar{w}$ are the sample means of $\mathbf{v}$ and $\mathbf{w}$, respectively, and $\text{cov}(\cdot, \cdot)$ denotes the sample covariance.

Putting all the pieces together, we can now make fully probabilistic forecasts by plugging in the estimated kernel in (5) and (6). Two key assumptions that enable our approach are local stationarity and co-location of inputs. In particular, we assume that the parameters of the prevailing flow, namely $\boldsymbol{\tau}$ and Ψ , remain unchanged across the training and forecast windows. We also require that the NWP inputs used to estimate $\boldsymbol{\tau}$ and Ψ are co-located, either exactly or approximately,

with the actual observations. This assumption is key to the accurate representation of the prevailing flow, and ultimately, the effective integration of NWP within our model. Research is currently ongoing to relax those assumptions.

V. REAL-WORLD CASE STUDY

We test our method using a rolling forecasting scheme, for forecast horizons, $h \in \{10, \dots, 60\}$ minutes. For each forecast roll, we train the model, obtain the forecasts, roll by six hours, and then repeat the training and forecasting procedures. This leads to a total of 100 rolls (considering we have 1 month of observations). Thus, we have a total of 6 forecasts/hour $\times$ 100 rolls $\times$ 2 spatial sites = 1,200 testing instances. For each forecast roll, five days of historical data and meteorological forecasts are used for model training. We find that a combination of an asymmetric kernel $C_a(\mathbf{h}, u)$ and a separable kernel yields better performance than solely using $C_a(\mathbf{h}, u)$, so we employ a convex combination of both and estimate the convex combination coefficient using the training data, along with the remaining GP hyperparameters.

A. Wind Speed Forecasting Results

We compare the wind speed forecasts obtained from our proposed approach against five prevalent forecasting benchmarks: (1) **GP**: This is a data-driven spatio-temporal GP, with a separable squared exponential kernel; (2) **ARMA**: The autoregressive moving average model, trained separately for each location (no spatial correlations); (3) **PER**: Persistence forecasts assume wind conditions persist in the forecast horizon; (4) **LSTM**: a deep learning model based on recurrent neural networks that is well-suited for time series data [24]. We fit a separate LSTM for each location and use grid search to optimize the hyperparameters; (5) **NWP**: Those are the hourly (physics-based) NWP model outputs, which we statistically interpolate (using cubic splines) to the target 10-min resolution; and (6) **HYB**: This is a hybrid model that calibrates NWP using local observations via a simple model output statistics (MOS) regression approach [25].

Table I (left) shows the mean absolute error (MAE) values for the wind speed forecasts for all models at various forecast horizons, $h \in \{10, \dots, 60\}$ minutes ahead. First, we clearly note how data-driven methods (including ours) are performing significantly better than physics-based approaches (NWP and HYB) at ultra-short term horizons, especially for the first 30 minutes. This agrees with the general consensus in the forecasting literature and practice regarding the superiority of data-driven approaches in ultra-short-term wind forecasting.

Second, we note how our approach performs noticeably better than all methods, including data-driven approaches (namely, GP, ARMA, LSTM, and PER), with average percentage improvements, ranging between 1.20 - 14.2%. Finally, we would like to stress how our method in particular outperforms the persistence forecast (PER), which is typically known to be highly competitive for such ultra-short-term horizons. Another major advantage of our proposed approach is its ability to naturally output probabilistic forecasts—Fig. 3 depicts the fore-

casts for five consecutive days, with 95% forecast intervals, suggesting notable agreement with the actual observations.

B. Wind Power Forecasting Results

To further demonstrate the value of our approach, we transform the wind speed forecasts into wind power predictions. Currently, there are no existing wind farms in the NY/NJ Bight (where the wind measurements are obtained), so we use actual power curves, constructed using the method of bins [26]–[28], on SCADA data obtained from an operational wind farm in the US [29]. We scale the power output to the $[0, 1]$ interval, such that a value of 1 represents the maximum rated capacity. We then use the constructed power curve to convert both the actual wind speed values, as well as the correspondent forecasts (from the six competing methods) into wind power predictions. Table I (right) shows the MAE values of the wind power predictions for the six models at different forecast horizons. Again, our model is able to outperform all of its competitors across all forecast horizons. We also notice that the improvements in the power domain are often higher than those in the wind speed domain, which aligns with the theoretical cubic speed-to-power functional relationship.

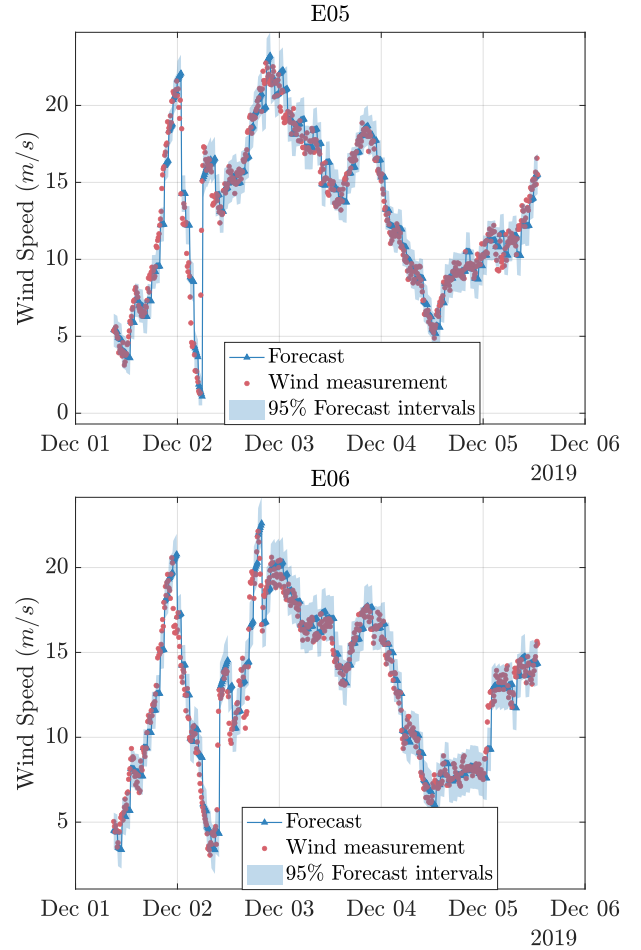


Fig. 3. Five-day forecasts from our proposed approach, along with 95th forecast intervals, on top of actual observations for E05 (top) and E06 (bottom), showing faithful alignment between the model and the ground truth.

TABLE I

FORECAST ERRORS (IN MAE), FOR WIND SPEED (LEFT) AND POWER (RIGHT), AVERAGED OVER BOTH SITES (E05 & E06). BOLD-FACED VALUES DENOTE BEST PERFORMANCE. AVG. AND %IMP ARE AVERAGE PERFORMANCE AND % IMPROVEMENT, OVER FORECAST HORIZONS, $h = 10, \dots, 60$ MIN.

	Wind Speed (m/s)							Wind Power (dimensionless)						
	Data-driven					Physics-based		Data-driven					Physics-based	
h (minutes)	Proposed	GP	ARMA	PER	LSTM	NWP	HYB	Proposed	GP	ARMA	PER	LSTM	NWP	HYB
10	.373	.376	.382	.376	.548	1.62	1.45	.030	.031	.034	.031	.048	.187	.153
20	.505	.510	.513	.509	.627	1.57	1.39	.043	.044	.046	.044	.056	.191	.154
30	.685	.692	.692	.695	.770	1.48	1.39	.085	.085	.085	.085	.087	.170	.164
40	.825	.832	.829	.833	.926	1.46	1.35	.096	.100	.101	.100	.105	.297	.148
50	.909	.921	.942	.925	1.05	1.50	1.39	.105	.111	.109	.111	.112	.291	.143
60	1.05	1.07	1.09	1.07	1.15	1.56	1.41	.129	.133	.134	.134	.130	.292	.139
Avg.	.725	.734	.742	.734	.845	1.53	1.39	.081	.084	.085	.084	.090	.238	.150
% IMP	-	1.23%	2.29%	1.23%	14.2%	52.6%	47.8%	-	3.57%	4.71%	3.57%	10.0%	66.0%	46.0%

VI. CONCLUSIONS

In this work, we proposed a probabilistic model for ultra-short-term wind forecasting. Unlike purely data-driven methods, we indirectly integrate NWPs to guide the selection of key physically meaningful parameters within the data-science-based model. We show that such indirect integration leads to large forecast accuracy gains, in terms of both wind speed and power, relative to purely data-driven models (that do not invoke NWPs), or those that directly use NWPs as inputs. An accurate estimation of prevailing flow parameters is key to our approach. Thus, we plan to explore ways to invoke multi-resolution, spatio-temporal NWPs over a grid to more effectively estimate the prevailing flow, and to leverage exogenous information such as temperature, pressure, or turbulence. We also intend to extensively evaluate our approach over longer horizons, larger datasets, and spatial networks.

REFERENCES

- [1] A. Lorca and X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1702–1713, 2014.
- [2] M. S. Modarresi, L. Xie, M. C. Campi, *et al.*, "Scenario-based economic dispatch with tunable risk levels in high-renewable power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5103–5114, 2018.
- [3] P. Papadopoulos, D. Coit, and A. A. Ezzat, "Seizing opportunity: Maintenance optimization in offshore wind farms considering accessibility, production, and crew dispatch," *IEEE Transactions on Sustainable Energy*, pp. 1–1, 2021.
- [4] P. Papadopoulos, D. W. Coit, and A. A. Ezzat, "STOCHOS: Stochastic opportunistic maintenance scheduling for offshore wind farms," *IIEE Transactions*, pp. 1–15, 2022.
- [5] M. F. Howland and J. O. Dabiri, "Wind farm modeling with interpretable physics-informed machine learning," *Energies*, vol. 12, no. 14, 2019.
- [6] G. Giebel and G. Kariniotakis, "Wind power forecasting—a review of the state of the art," *Renewable energy forecasting*, pp. 59–109, 2017.
- [7] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2770–2779, 2017.
- [8] N. Safari, S. Mazhari, and C. Chung, "Very short-term wind power prediction interval framework via bi-level optimization and novel convex cost function," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1289–1300, 2018.
- [9] A. A. Ezzat, "Turbine-specific short-term wind speed forecasting considering within-farm wind field dependencies and fluctuations," *Applied Energy*, vol. 269, p. 115 034, 2020, ISSN: 0306-2619.
- [10] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting," *IEEE Access*, vol. 7, pp. 28 309–28 318, 2019.
- [11] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge: MIT Press, Jan. 2006.
- [12] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using gaussian processes and numerical weather prediction," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 656–665, 2013.
- [13] Q. Xu, D. He, N. Zhang, *et al.*, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Transactions on sustainable energy*, vol. 6, no. 4, pp. 1283–1291, 2015.
- [14] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 9, no. 2, e365, 2020.
- [15] BOEM, *Lease and grant information*, <https://www.boem.gov/renewable-energy/lease-and-grant-information>, 2017.
- [16] J. Dicapoulos, J. F. Brodie, S. Glenn, *et al.*, "Weather research and forecasting model validation with NREL specifications over the new york/new jersey bight for offshore wind development," in *OCEANS 2021: San Diego-Porto*, IEEE, 2021, pp. 1–7.
- [17] M. Optis, A. Kumler, G. N. Scott, M. C. Debnath, and P. J. Moriarty, "Validation of RU-WRF, the custom atmospheric mesoscale model of the rutgers center for ocean observing leadership," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2020.
- [18] N. Cressie and C. K. Wikle, *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 2015.
- [19] M. L. O. Salvaña and M. G. Genton, "Nonstationary cross-covariance functions for multivariate spatio-temporal random fields," *Spatial Statistics*, vol. 37, p. 100411, 2020, ISSN: 2211-6753.
- [20] M. L. Stein, "Space-time covariance functions," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 310–321, 2005.
- [21] A. A. Ezzat, M. Jun, and Y. Ding, "Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1437–1447, 2018.
- [22] M. Schlather, "Some covariance models based on Normal scale mixtures," *Bernoulli*, vol. 16, no. 3, pp. 780–797, 2010.
- [23] A. A. Ezzat, M. Jun, and Y. Ding, "Spatio-temporal short-term wind forecast: A calibrated regime-switching method," *The Annals of Applied Statistics*, vol. 13, no. 3, pp. 1484–1510, 2019.
- [24] M.-S. Ko, K. Lee, J.-K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 1321–1335, 2020.
- [25] H. R. Glahn and D. A. Lowry, "The use of model output statistics (MOS) in objective weather forecasting," *Journal of Applied Meteorology and Climatology*, vol. 11, no. 8, pp. 1203–1211, 1972.
- [26] "Wind Energy Generation Systems - Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines," *IEC 61400-12-1*, 2017, International Electrotechnical Commission.
- [27] B. Golparvar, P. Papadopoulos, A. Ezzat, and R.-Q. Wang, "A surrogate-model-based approach for estimating the first and second-order moments of offshore wind power," *Applied Energy*, vol. 299, p. 117 286, 2021.
- [28] P. Nasery and A. Ezzat, "Yaw-adjusted wind power curve modeling: A local regression approach," *Renewable Energy*, vol. 202, pp. 1368–1376, 2023.
- [29] Y. Ding, *Data Science for Wind Energy*. CRC Press, 2019.