IIF-SAS Final Report for the Project Titled:
"**Forecasting in Unknown Territory: Towards Physically Motivated Learning for Local Wind Fields**"

Ahmed Aziz Ezzat, Ph.D.,
Department of Industrial & Systems Engineering,
Rutgers, The State University of New Jersey,
Piscataway, NJ, USA.

December 2022

A Physically Motivated Machine Learning Approach for Very-Short-Term Wind Speed and Power Forecasting

Feng Ye, Joseph Brodie, Travis Miles, Ahmed Aziz Ezzat
Rutgers, The State University of New Jersey

Abstract

Very-short-term wind forecasts (i.e., wind speed and power predictions issued at sub-hourly forecast horizons) are indispensable to the reliable operation of wind energy systems. The dominant consensus in the wind forecasting literature and practice is that data-driven approaches are, arguably, the "right" tools for such short-term horizons. This is in contrast to numerical weather predictions (NWP), or hybrid methods thereof, for which the value is typically substantiated as the forecast horizon becomes longer ($>$ 1-3 hours). In this work, we propose a probabilistic machine learning (ML) model that leverages NWP information—which is typically available to the farm or power system operator at the time of forecast—in order to make very-short-term wind speed and power forecasting (up to an hour ahead). Instead of directly using NWPs as input regressors as in most hybrid approaches, we indirectly invoke NWP information in selecting key hyperparameters within the ML model, thereby guiding it to adhere to certain physical principles related to local wind field formation and propagation. Using real-world data from the U.S. NY/NJ Bight where several offshore wind projects are currently in-development, we show that such indirect integration of NWPs within an ML approach outperforms several prevalent forecasting methods, including persistence forecasts, which are known to be highly competitive at ultra-short-term horizons. We envision this work to serve as an exemplar for leveraging the rich, yet coarser-resolution information of NWPs in improving ML-based ultra-short-term wind forecasting models.

Keywords: Forecasting, Spatio-Temporal Learning, Wind Energy.

1. Introduction

Very-short-term wind forecasting refers to the prediction of wind speed and wind power at ultra-short-term forecast horizons (In this work, starting from 10 minutes ahead up to an hour). The value of such very-short-term forecasts stems from their high utility to several critical wind farm and power system operations, including but not limited to economic dispatch, market participation, and reserve planning (Pinson, 2013; Xie et al., 2013; Lorca & Sun, 2014; Safta et al., 2016; Modarresi et al., 2018), turbine- and farm-level asset prognostics and management (Taylor & Jeon, 2018; Golparvar et al., 2021; Papadopoulos et al., 2021, 2022), as well as production estimation, control and optimization (Howland & Dabiri, 2019; Nasery & Ezzat, 2022).

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 (Giebel & Kariniotakis, 2017). NWPs are physics-based approaches that numerically solve a system of discretized differential conservation equations of mass, momentum, and energy in the atmosphere, and forecast wind speed or other weather conditions based on the solution of the physical equations Lange & Focken (2006). Despite their value, there is an overall consensus in the wind forecasting literature and practice that invoking NWP information is particularly valuable (and in fact, necessary) as the forecast horizon becomes longer. In the literature, a precise threshold between what constitutes a "short" temporal horizon does not appear to exist, but it is typically in the range between 1-6 hours when NWPs (or hybrid methods thereof) tend to outperform purely data-driven models (Giebel et al., 2011).

On the other hand, for relatively shorter forecast horizons, data-driven methods, and in particular those that are primarily based on statistical and machine learning (ML), are typically regarded as the right tools for predicting wind speed and power. By relying on an implicit assumption of persistence, i.e., a similarity between what *has been* seen and what *will* be seen, statistical and ML methods are able to effectively learn, "mine," and extrapolate patterns, trends, and correlations from sheer volumes of historical measurements into the near future (Pinson & Madsen, 2012; Khodayar et al., 2017; Safari et al., 2018; Ju et al., 2019), thereby overcoming the computational and modeling limitations of NWPs at shorter forecast horizons. Few non-exhaustive examples of statistical and ML methods which have been pervasively used for short-term wind forecasting include autoregressive-based models (Erdem & Shi, 2011; Gneiting et al., 2006), deep learning models, in particular long short-term memory and recurrent neural networks (Feng et al., 2017; Liu et al., 2020), and geostatistical approaches (Yan et al., 2016; Ezzat et al., 2019; Ezzat, 2020; Lenzi & Genton, 2020).

Despite their demonstrated success in short-term wind forecasting, statistical and ML methods are often criticized for being largely *physics-agnostic*, that is, they are often developed and trained without careful consideration of the first principles governing the physical process being studied (in this case, the wind field). As a result, they may be vulnerable to model specifications that violate those fundamental physical principles. This has driven an active area of research collectively dubbed as physics-informed or physics-guided statistical/machine learning (Gneiting, 2002; Karniadakis et al., 2021), wherein existing physical information and/or domain knowledge is leveraged to inform the learning and/or prediction stages in an ML pipeline, thereby offering an alternative paradigm to the black-box, purely data-driven approach.

To circumvent this limitation, we propose a probabilistic ML model for very-short-term wind speed and power forecasting, for which the key distinguishing feature is the ability to actually make use of NWP information that is typically available to the farm or power system operator at the time of forecast. The proposed approach is based on a spatio-temporal Gaussian process (GP) (Rasmussen & Williams, 2006), within which NWPs are leveraged to implicitly guide the selection of key hyperparameters in the covariance function in order to encode information about the local wind flow in the region under study. This *indirect integration* of NWPs within an ML-based forecasting model breaks away from the mainstream approach of hybrid forecasting, wherein NWPs are *directly integrated* as input regressors to ML-based wind forecasting models (Chen et al., 2013; Xu et al., 2015; Hu et al., 2021). We believe that our approach is less prone than hybrid methods to carry over the deficiencies of NWPs when directly downscaled to higher resolutions and very-short-term forecast horizons. Common examples of those NWP deficiencies at fine-scale resolutions and horizons include forecast inaccuracies, as well as temporal, spatial, and scaling biases (Sweeney et al., 2020).

Thus, our approach falls under the umbrella of data-driven approaches described above, but is not a purely "data-driven" method. Instead, the construction of the model, as pointed out earlier, is guided by the physics unique to the local wind field, without the need to explicitly re-model those physics. Our driving hypothesis is that such physically motivated data-driven approach yields substantial improvements in the predictive and explanatory power relative to the purely data-driven approaches which are indifferent to the physical properties of wind fields, as well as to those that are primarily physics-based.

We train and test our proposed approach 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 (BOEM, 2017). 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 ML-based ultra-short-term wind forecasting.

The remainder of this paper is organized as follows. Section 2 describes the real-world data used in this paper. Section 3 reviews the concept of spatio-temporal asymmetry for wind fields, which will then be used in deciphering the potential role of NWPs in data-driven ultra-short-term forecasting. In Section 4, we introduce our proposed forecasting method, which is then followed by

Section 5 where we present and analyze our results. Finally, we conclude the report in Section 6.

2. Data Description

Our dataset comprises 10-min wind speed measurements, at 100-m altitude, collected using two floating LiDAR buoys that have been recently deployed by the New York State Energy Research & Development Authority (NYSERDA) called E05 and E06, respectively, in order to further our understanding of the wind resource in proximity to the wind energy areas in the NY/NJ Bight (NYSERDA, 2019). Figure 1(a) depicts the rose plot of the wind data, showing the distribution of the wind speed and direction in that region. The rose plot suggests a north-westerly prevailing wind. The two buoys are ∼77 Km apart, and their exact coordinates are 39°58'10"N and 72°43'00"W for E05 and 39°32'50"N and 73°25'45"W for E06.

We also obtain a set of hourly NWP forecasts from a meso-scale meteorological model operated by Rutgers University, called RU-WRF (short for the Rutgers University Weather Research & Forecasting model) (RUCOOL, 2019). The model has been recently validated by the National Renewable Energy Laboratory (NREL) (Optis et al., 2020), and has been continously updated and improved since then (Dicopoulos et al., 2021). For this study, both measurements and NWP outputs span the month of December 2019. Figure 1(b) shows the histograms of the actual observations versus their correspondent NWPs, while Figure 1(c) shows a 12-day time series of the actual wind speed measurements versus statistically interpolated NWPs for E05 (top) and E06 (bottom).

3. Spatio-temporal Data Analysis

Let $Z(\mathbf{s}, t)$ denote a spatio-temporal random process, 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

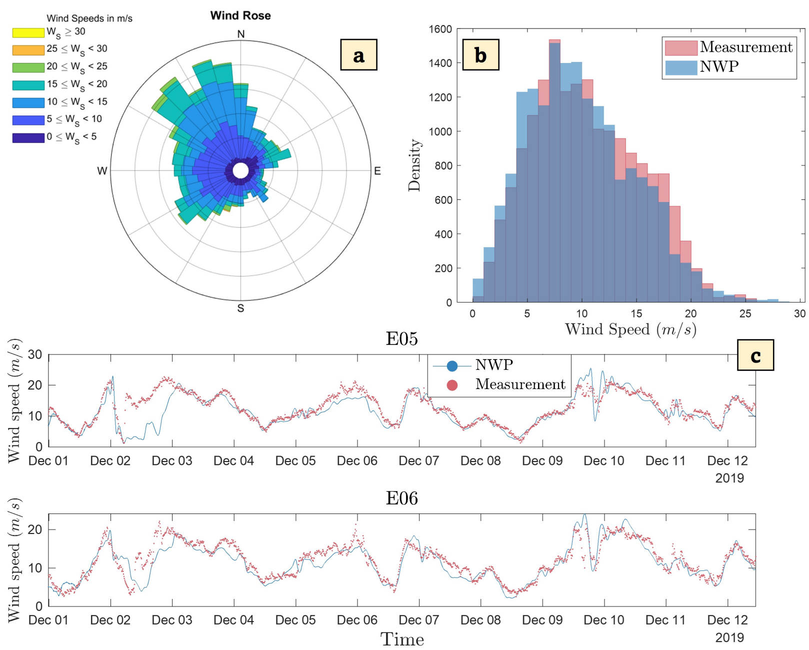

Fig. 1. (a) Wind rose plot for the actual wind speed measurements recorded in the NY/NJ Bight (Data from two buoys combined to produce this figure); (b) Histograms of actual wind speed measurements (mean = 10.52 m/s) versus NWP wind speed forecasts (mean = 9.87 m/s); (c) actual wind speed measurements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

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 $C(\mathbf{h}, u) : \mathbb{R}^2 \times \mathbb{Z}^+ \to \mathbb{R}$, 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 spatio-temporal statistical literature is through the so-called separable approach, wherein $C(\mathbf{h}, u)$ is expressed as $C(\mathbf{h}, u) = C^{\mathbf{s}}(\mathbf{h}) \times C^t(u)$, such that $C^{\mathbf{s}}(\mathbf{h})$ and $C^t(u)$ are two independent kernels for space and time, respectively (Cressie & Wikle, 2015). As such, one could independently model the covariance structure across space and time through separate covariance functions or kernels. Popular selections for $C^{\mathbf{s}}(\mathbf{h})$ and $C^t(u)$ include the Gaussian, squared exponential and Matérn kernels (Rasmussen & Williams, 2006). The final spatio-temporal covariance matrix can be computed efficiently as the Kronecker product of two smaller spatial and temporal covariance matrices, as in 1.

$$\boldsymbol{\Sigma}^{sep} = \boldsymbol{\Sigma}^{\mathbf{s}} \otimes \boldsymbol{\Sigma}^t, \tag{1}$$

such that $\boldsymbol{\Sigma}^{sep}$ is the resulting covariance matrix, whereas $\boldsymbol{\Sigma}^{\mathbf{s}}$ and $\boldsymbol{\Sigma}^t$ are the spatial and temporal covariance matrices, respectively.

Despite its simplicity, a major limitation of the abovementioned separable approach is that it assumes, by design, that space-time correlations are symmetric, that is, we have $cor\{Z(\mathbf{s}_i, t), Z(\mathbf{s}_{i'}, t+u)\} = cor\{Z(\mathbf{s}_{i'}, t), Z(\mathbf{s}_i, t+u)\}$. 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 due to the impact of the prevailing transport effect (Cressie & Wikle, 2015; Salvaña & Genton, 2020), thereby making the assumption of symmetry physically "unjustifiable" for local wind fields.

To demonstrate this using our data, we use an estimator of asymmetry (defined as lack of symmetry), expressed as in (2) (Stein, 2005; Ezzat et al., 2018).

$$a(\mathbf{s}_i, \mathbf{s}_{i'}, u) := \delta(\mathbf{s}_i, \mathbf{s}_{i'}, u) - \delta(\mathbf{s}_{i'}, \mathbf{s}_i, u), \tag{2}$$

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, which is a a measure of dissimilarity between a pair spatio-temporal observations, and is expressed as in (3).

$$\delta\left(\mathbf{s}_i, \mathbf{s}_{i'}, u\right) = \frac{1}{2(N-u-1)} \sum_{k=1}^{N-u-1} \left\{y\left(\mathbf{s}_i, k+u\right) - y\left(\mathbf{s}_{i'}, k\right)\right\}^2. \tag{3}$$

In (3), N is the number of observations, $\delta\left(\mathbf{s}_i, \mathbf{s}_{i'}, u\right)$ means that the measurements taken at site $\mathbf{s}_{i'}$ are u time lag behind that at site $\mathbf{s}_i$, while $\delta\left(\mathbf{s}_{i'}, \mathbf{s}_i, u\right)$ 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.

To test for statistical significance, we perform a t-test for each time lag, $u \in \{1, ..., 36\}$ (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. Figure 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 Figure 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 the rose plot in Figure 1(a). Another piece of information conveyed by Figure 2 is that the maximum asymmetry levels appear to take place in time lags of $\sim$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

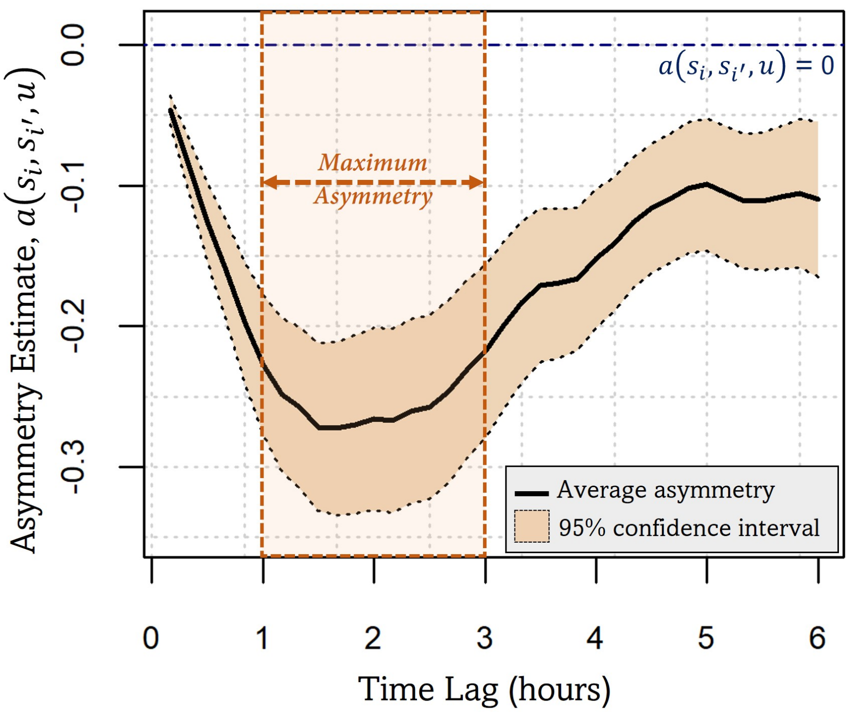

Fig. 2. Average difference in empirical semi-variograms versus the time lag, along with 95% t-test confidence intervals. Noticeable levels of asymmetry are observed, peaking at ∼1-3-hour time lag, which is line with the expected duration for wind conditions to propagate across the wind field—the average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

integration are discussed in Section 4.

4. Methodology

We first introduce spatio-temporal Gaussian processes in Section 4.1, then discuss the role of NWPs in Section 4.2.

4.1. Spatio-temporal Gaussian Processes (GPs)

Let $\mathbf{Z} = [z(\mathbf{s}_1, t_1), z(\mathbf{s}_1, t_2), \ldots, z(\mathbf{s}_1, t_T), \ldots, z(\mathbf{s}_n, t_T)]^T$ 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 (4).

$$z(\mathbf{s}_i, t_j) = m(\mathbf{s}_i, t_j) + \gamma(\mathbf{s}_i, t_j), \tag{4}$$

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 $\mathbf{\Sigma} + \delta\mathbf{I}$, where δ is the noise parameter, and $\mathbf{I}$ is the identify matrix. The entries of $\mathbf{\Sigma}$ 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 $\boldsymbol{Z}$ and a set of testing data $\boldsymbol{Z}_*$ follows a multivariate Gaussian distribution, as shown in (5).

$$\begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{Z}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{\Sigma}_* \\ \mathbf{\Sigma}_*^T & \mathbf{\Sigma}_{**} \end{bmatrix} \right), \tag{5}$$

where $\mathbf{m} = [m(\mathbf{s}_1, t_1), ..., m(\mathbf{s}_n, t_T)]^T$ is the vector of mean function evaluations at the training data, and $\mathbf{m}_*$ is similarly defined for the testing data. The matrix $\mathbf{\Sigma}_*$ holds the covariance values between $\boldsymbol{Z}$ and $\boldsymbol{Z}_*$, while $\mathbf{\Sigma}_{**}$ 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 (6).

$$\mathcal{P}\left(\boldsymbol{Z}_* \mid \boldsymbol{Z}, m(\cdot), \gamma(\cdot,\cdot)\right) \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \widehat{\mathbf{\Sigma}}), \tag{6}$$

where

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \mathbf{m}_* + \mathbf{\Sigma}_*^T (\mathbf{\Sigma} + \delta\mathbf{I})^{-1} (\boldsymbol{Z} - \mathbf{m}), \\ \widehat{\mathbf{\Sigma}} &= \mathbf{\Sigma}_{**} - \mathbf{\Sigma}_*^T (\mathbf{\Sigma} + \delta\mathbf{I})^{-1} \mathbf{\Sigma}_*. \end{aligned} \tag{7}$$

We note that using GPs for wind forecasting may entail some data transformations to ensure the validity of GP assumptions. For our case study, non-positivity was not an issue since offshore wind speeds are typically high (average = 10.52 m/s)—See the histogram of Fig 1(b). Despite not enforcing a non-positivity constraint, the resulting GP-based mean forecasts were all non-negative, whereas less than 0.1% of the 95% prediction intervals included values that are slightly lower than zero. We acknowledge, however, that for other datasets, one may wish to invoke data pre-processing transformations before using GPs.

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

Defining the GP kernel $C(\mathbf{h}, u)$ is essential for spatio-temporal GPs. The analysis in Section 3 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 $\boldsymbol{\Theta} \in \mathbb{R}^2$, creating a spatial-temporal random process with an asymmetric covariance, $C_a(\mathbf{h}, u)$, expressed as in (8). This general covariance structure (despite still lacking a closed-form representation) is referred to as a Lagragian reference framework in the geostatistical literature (Cox & Isham, 1988; Gneiting et al., 2007).

$$C_a(\mathbf{h}, u) = \mathbb{E}_{\Theta}\{C_s(\mathbf{h} - \boldsymbol{\Theta} u)\}. \tag{8}$$

Assuming $C_s(x) = \exp(-x^2)$ and letting $\boldsymbol{\Theta} \sim \mathcal{N}(\boldsymbol{\tau}, \boldsymbol{\Psi})$ gives us the closed-form expression in (9), which is referred to as Schlather's covariance model (Schlather, 2010).

$$\begin{aligned} C_a(\mathbf{h},u) = {} & \frac{1}{\sqrt{|\mathbf{I}_{2\times 2} + 2\boldsymbol{\Psi} u^2|}} \\ & \times \exp\left\{ -\left(\mathbf{h} - \boldsymbol{\tau} u\right)^T \left(\mathbf{I}_{2\times 2} + 2\boldsymbol{\Psi} u^2\right)^{-1} \left(\mathbf{h} - \boldsymbol{\tau} u\right) \right\}, \end{aligned} \tag{9}$$

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

The choice of $\boldsymbol{\tau}$ and $\boldsymbol{\Psi}$ (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 $\boldsymbol{\Psi}$ can severely limit (or even reverse) the benefits of an asymmetric approach. In our prior works (Ezzat et al., 2018, 2019), $\boldsymbol{\tau}$ and $\boldsymbol{\Psi}$ have been either preset, or estimated using historical measurements. We believe, however, that local measurements are not necessarily the best descriptors of 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 $\boldsymbol{\Psi}$ as in (10) and (11), respectively.

$$\boldsymbol{\tau} = [\tau_1, \tau_2]^T = [\bar{v}, \bar{w}]^T, \tag{10}$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} \\ \Psi_{2,1} & \Psi_{2,2} \end{bmatrix} = \begin{bmatrix} cov(\mathbf{v}, \mathbf{v}) & cov(\mathbf{v}, \mathbf{w}) \\ cov(\mathbf{w}, \mathbf{v}) & cov(\mathbf{w}, \mathbf{w}) \end{bmatrix}, \tag{11}$$

where $\mathbf{v} = [v_1, ..., v_{T+h}]^T$ and $\mathbf{w} = [w_1, ..., w_{T+h}]^T$ 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 $cov(\cdot, \cdot)$ denotes the sample covariance.

5. Real-world Case Study

We test our method using a rolling forecasting scheme, for forecast horizons, $h \in \{10, ..., 60\}$ 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 , which is equivalent to $1,200$ testing instances (6 forecasts/hour $\times$100 rolls $\times$2 locations). 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.

5.1. Wind Speed Forecasting Results

We compare the wind speed forecasts obtaiend 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. The purpose of this benchmark is to demonstrate the merits of a physically meaningful covariance function proposed in our model relative to a black-box, physics-agnostic kernel choice.

Table 1: Forecasting results (in MAE), for wind speed (up) and wind power (bottom), averaged over both sites (E05 and E06). Bold-faced values indicate best performance (i.e. lower errors). Avg. and %IMP denote average performance, and percentage improvement, across different forecast horizons, $h = 10, ..., 60$ mins.

	Wind Speed (m/s)					
	Data-driven				**Physics-based**	
h (minutes)	Proposed	GP	ARMA	PER	NWP	HYB
10	**.373**	.376	.382	.376	1.62	1.51
20	**.505**	.510	.513	.509	1.57	1.28
30	**.685**	.692	.692	.695	1.48	1.28
40	**.825**	.832	.829	.833	1.46	1.30
50	**.909**	.921	.942	.925	1.50	1.36
60	**1.05**	1.07	1.09	1.07	1.56	1.44
Avg.	**.725**	.734	.742	.734	1.53	1.36
% IMP	-	1.20%	2.23%	1.22%	52.6%	46.7%

	Wind Power (dimensionless)					
	Data-driven				**Physics-based**	
h (minutes)	Proposed	GP	ARMA	PER	NWP	HYB
10	**.030**	.031	.034	.031	.187	.153
20	**.043**	.044	.046	.044	.191	.154
30	**.085**	.085	.085	.085	.170	.164
40	**.096**	.100	.101	.100	.297	.148
50	**.105**	.111	.109	.111	.291	.143
60	**.129**	.133	.134	.134	.292	.139
Avg.	**.081**	.084	.085	.084	.238	.150
% IMP	-	3.17%	4.02%	3.39%	65.8%	45.8%

2. **ARMA**: The autoregressive moving average model is a statistical approach, trained separately for each location (no spatial correlations considered). Bayesian Information Criterion (BIC) and Maximum Likelihood Estimation (MLE) are used to dynamically update the model order parameters and the model coefficients, respectively, at each forecasting roll.

3. **PER**: Persistence forecasting assumes wind conditions persist in the forecast horizon. This method is known to be highly competitive for ultra-short-term horizons, with a decaying predictive skill as the forecast horizon becomes longer.

4. **NWP**: Those are the hourly (physics-based) NWP model outputs, which we statistically interpolate (using cubic splines) to the target 10-min resolution.

5. **HYB**: This is a hybrid model that calibrates NWPs using local observations via a simple model output statistics (MOS) regression approach (Glahn & Lowry, 1972).

Table 1 (top) shows the mean absolute error (MAE) values for the wind speed forecasts for all models at various forecast horizons, $h \in \{10, ..., 60\}$ minutes ahead. First, we clearly note how data-driven methods (including ours) are performing significantly better than the benchmarks that directly invoke NWP information in the forecast (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, and PER), with average percentage improvements, ranging between 1.20 - 2.23%. Finally, we would like to stress how our method in particular outperforms the persistence forecast (PER), which is 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. Figure 3 depicts the probabilistic forecasts for five consecutive days, with 95% forecast intervals, suggesting a faithful agreement with the actual observations.

5.2. Wind Power Forecasting Results

We transform the wind speed forecasts into wind power predictions using actual power curves. Currently, there are no existing wind farms in the NY/NJ Bight (where the wind measurements are obtained), so we use actual power curves that have been constructed using the method of bins (IEC, 2017; Golparvar et al., 2021) applied on SCADA data obtained from an operational wind farm in the US (Ding, 2019). The method of bins is the industrial standard for power curve modeling and entail the discretization of the wind speed domain into a number of bins (typically with a bin width of 0.5 m/s) and then averaging the power output within each bin, thus producing bin-specific power estimates.

We scale the power output to the $[0, 1]$ interval, such that a value of 1 constitutes 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 1 (bottom) 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 fairly higher than those in the wind speed domain, which aligns with the theoretical cubic speed-to-power functional relationship.

6. Conclusions

In this work, we proposed a data-driven, spatio-temporal model for ultra-short-term wind speed and power forecasting. Unlike purely data-driven methods, or on the other hand, those that are primarily physics-based, we indirectly leverage numerical weather predictions in loosely guiding the selection of key physically meaningful parameters within the ML-based model (in particular, in

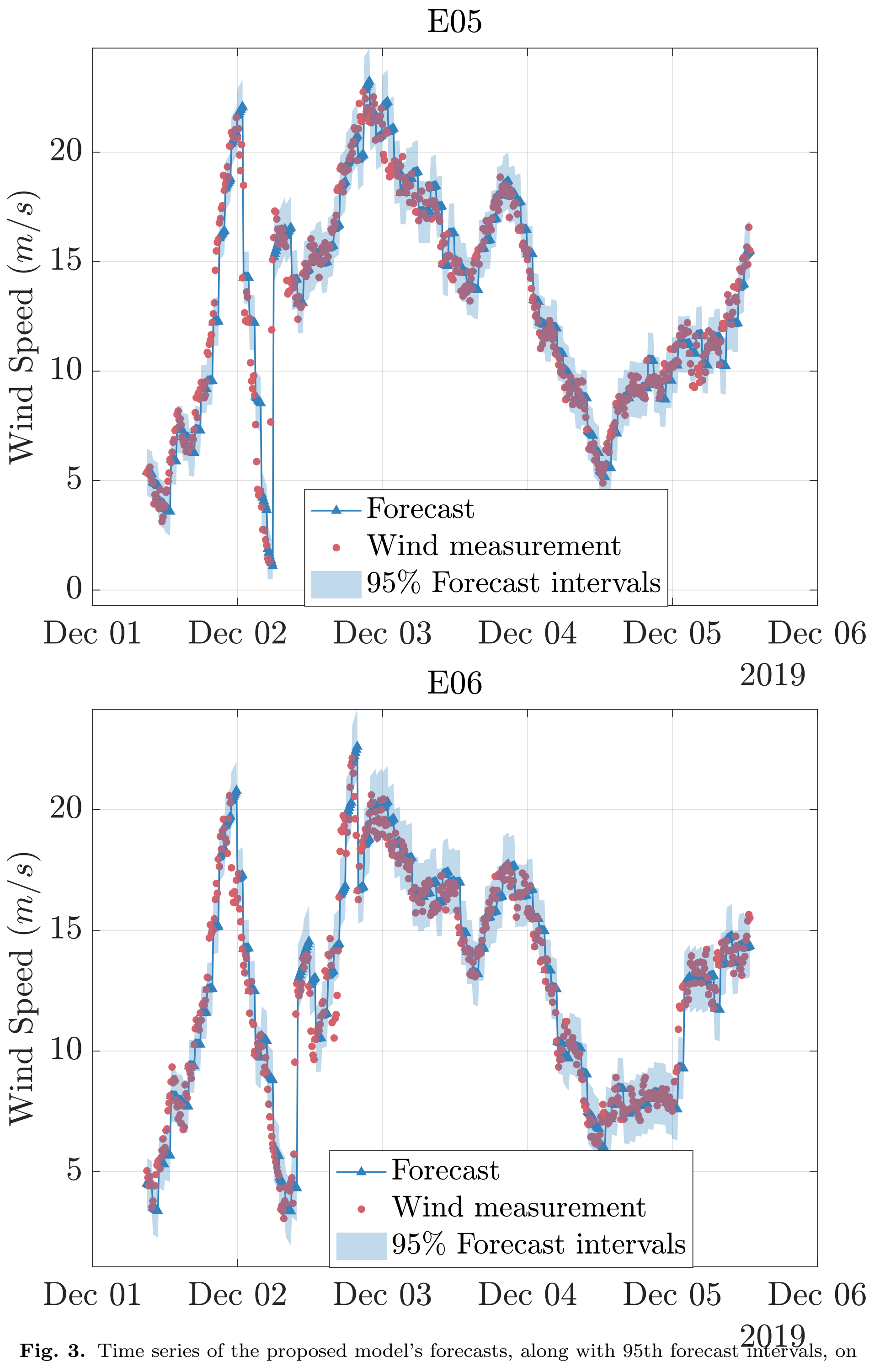

Fig. 3. Time series of the proposed model's forecasts, along with 95th forecast intervals, on top of actual observations for E05 (top) and E06 (bottom).

informing the kernel choice of the GP model). We show that, for ultra-short-term horizons, such indirect integration leads to noticeable forecast accuracy improvements, in terms of both wind speed and power prediction, relative to purely data-driven models (that do not invoke NWPs), or those that directly use NWPs as inputs. Further research will investigate the merit of our approach for longer forecast horizons and larger spatial networks.

References

BOEM (2017). Lease and grant information. `https://www.boem.gov/renewable-energy/lease-and-grant-information`.

Chen, N., Qian, Z., Nabney, I. T., & Meng, X. (2013). Wind power forecasts using gaussian processes and numerical weather prediction. *IEEE Transactions on Power Systems*, *29*, 656–665.

Cox, D. R., & Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, *415*, 317–328.

Cressie, N., & Wikle, C. K. (2015). *Statistics for Spatio-Temporal Data*. John Wiley & Sons.

Dicopoulos, J., Brodie, J. F., Glenn, S., Kohut, J., Miles, T., Seroka, G., Dunk, R., & Fredj, E. (2021). 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* (pp. 1–7). IEEE.

Ding, Y. (2019). *Data science for wind energy*. CRC Press.

Erdem, E., & Shi, J. (2011). Arma based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, *88*, 1405–1414.

Ezzat, A. A. (2020). Turbine-specific short-term wind speed forecasting considering within-farm wind field dependencies and fluctuations. *Applied Energy*, *269*, 115034.

Ezzat, A. A., Jun, M., & Ding, Y. (2018). Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. *IEEE Transactions on Sustainable Energy*, *9*, 1437–1447.

Ezzat, A. A., Jun, M., & Ding, Y. (2019). Spatio-temporal short-term wind forecast: A calibrated regime-switching method. *The Annals of Applied Statistics*, *13*, 1484 – 1510.

Feng, C., Cui, M., Hodge, B.-M., & Zhang, J. (2017). A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Applied Energy*, *190*, 1245–1257.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). The state-of-the-art in short-term prediction of wind power: A literature overview. *ANEMOS. plus*, .

Giebel, G., & Kariniotakis, G. (2017). Wind power forecasting—a review of the state of the art. *Renewable energy forecasting*, (pp. 59–109).

Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (mos) in objective weather forecasting. *Journal of Applied Meteorology and Climatology*, *11*, 1203–1211.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, *97*, 590–600.

Gneiting, T., Genton, M., & Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. *Statistical Methods for Spatio-Temporal Systems*, (pp. 151–175).

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., & Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space–time method. *Journal of the American Statistical Association*, *101*, 968–979.

Golparvar, B., Papadopoulos, P., Ezzat, A. A., & Wang, R.-Q. (2021). A surrogate-model-based approach for estimating the first and second-order moments of offshore wind power. *Applied Energy*, *299*, 117286.

Howland, M. F., & Dabiri, J. O. (2019). Wind farm modeling with interpretable physics-informed machine learning. *Energies*, *12*.

Hu, S., Xiang, Y., Zhang, H., Xie, S., Li, J., Gu, C., Sun, W., & Liu, J. (2021). Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction. *Applied Energy*, *293*, 116951.

IEC (2017). Wind Energy Generation Systems - Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines. *IEC 61400-12-1*, . International Electrotechnical Commission.

Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., & Rehman, M. U. (2019). A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting. *Ieee Access*, *7*, 28309–28318.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*, 422–440.

Khodayar, M., Kaynak, O., & Khodayar, M. E. (2017). Rough deep neural architecture for short-term wind speed forecasting. *IEEE Transactions on Industrial Informatics*, *13*, 2770–2779.

Lange, M., & Focken, U. (2006). *Physical Approach to Short-Term Wind Power Prediction*. Springer.

Lenzi, A., & Genton, M. G. (2020). Spatiotemporal probabilistic wind vector forecasting over Saudi Arabia. *The Annals of Applied Statistics*, *14*, 1359 – 1378.

Liu, Y., Qin, H., Zhang, Z., Pei, S., Jiang, Z., Feng, Z., & Zhou, J. (2020). Probabilistic spatiotemporal wind speed forecasting based on a variational bayesian deep learning model. *Applied Energy*, *260*, 114259.

Lorca, A., & Sun, X. A. (2014). Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems*, *30*, 1702–1713.

Modarresi, M. S., Xie, L., Campi, M. C., Garatti, S., Care, A., Thatte, A. A., & Kumar, P. (2018). Scenario-based economic dispatch with tunable risk levels in high-renewable power systems. *IEEE Transactions on Power Systems*, *34*, 5103–5114.

Nasery, P., & Ezzat, A. A. (2022). Yaw-adjusted wind power curve modeling: A local regression approach. *Renewable Energy*, .

NYSERDA (2019). Nyserda floating lidar buoy data. `https://oswbuoysny.resourcepanorama.dnvgl.com/`.

Optis, M., Kumler, A., Scott, G. N., Debnath, M. C., & Moriarty, P. J. (2020). *Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership*. Technical Report National Renewable Energy Lab.(NREL), Golden, CO (United States).

Papadopoulos, P., Coit, D., & Ezzat, A. A. (2021). Seizing opportunity: Maintenance optimization in offshore wind farms considering accessibility, production, and crew dispatch. *IEEE Transactions on Sustainable Energy*, (pp. 1–1).

Papadopoulos, P., Coit, D. W., & Ezzat, A. A. (2022). Stochos: Stochastic opportunistic maintenance scheduling for offshore wind farms. *arXiv preprint arXiv:2207.02274*, .

Pinson, P. (2013). Wind Energy: Forecasting Challenges for Its Operational Management. *Statistical Science*, *28*, 564 – 585.

Pinson, P., & Madsen, H. (2012). Adaptive modelling and forecasting of offshore wind power fluctuations with markov-switching autoregressive models. *Journal of Forecasting*, *31*, 281–313.

Rasmussen, C., & Williams, C. (2006). *Gaussian Processes for Machine Learning*. Cambridge: MIT Press.

RUCOOL (2019). Rutgers weather research and forecasting model. `https://tds.marine.rutgers.edu/thredds/dodsC/cool/ruwrf/wrf_4_1_3km_processed/WRF_4.1_3km_Processed_Dataset_Best.html`.

Safari, N., Mazhari, S., & Chung, C. (2018). Very short-term wind power prediction interval framework via bi-level optimization and novel convex cost function. *IEEE Transactions on Power Systems*, *34*, 1289–1300.

Safta, C., Chen, R. L.-Y., Najm, H. N., Pinar, A., & Watson, J.-P. (2016). Efficient uncertainty quantification in stochastic economic dispatch. *IEEE Transactions on Power Systems*, *32*, 2535–2546.

Salvaña, M. L. O., & Genton, M. G. (2020). Nonstationary cross-covariance functions for multivariate spatio-temporal random fields. *Spatial Statistics*, *37*, 100411.

Schlather, M. (2010). Some covariance models based on normal scale mixtures. *Bernoulli*, *16*, 780 – 797.

Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association*, *100*, 310–321.

Sweeney, C., Bessa, R. J., Browell, J., & Pinson, P. (2020). The future of forecasting for renewable energy. *Wiley Interdisciplinary Reviews: Energy and Environment*, *9*, e365.

Taylor, J. W., & Jeon, J. (2018). Probabilistic forecasting of wave height for offshore wind turbine maintenance. *European Journal of Operational Research*, *267*, 877–890.

Xie, L., Gu, Y., Zhu, X., & Genton, M. G. (2013). Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Transactions on Smart Grid*, *5*, 511–520.

Xu, Q., He, D., Zhang, N., Kang, C., Xia, Q., Bai, J., & Huang, J. (2015). A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on sustainable energy*, *6*, 1283–1291.

Yan, J., Li, K., Bai, E.-W., Deng, J., & Foley, A. M. (2016). Hybrid probabilistic wind power forecasting using temporally local gaussian process. *IEEE Transactions on Sustainable Energy*, *7*, 87–95.