

AIRU-WRF: A Physics-Guided Spatio-Temporal Wind Forecasting Model and its Application to the U.S. Mid Atlantic Offshore Wind Energy Areas

Feng Ye^a, Joseph Brodie^b, Travis Miles^c, Ahmed Aziz Ezzat^a

^a*Industrial & Systems Engineering, Rutgers University, Piscataway, 08854, NJ, USA*

^b*AKRF Inc., New York, 10016, NY, USA*

^c*Marine & Coastal Sciences, Rutgers University, Piscataway, 08901, NJ, USA*

Abstract

The reliable integration of wind energy into modern-day electricity systems heavily relies on accurate short-term wind forecasts. We propose a spatio-temporal model called AIRU-WRF (short for the AI-powered Rutgers University Weather Research & Forecasting), which combines numerical weather predictions (NWP) with local observations in order to make wind speed forecasts that are short-term (minutes to hours ahead), and of high resolution, both spatially (site-specific) and temporally (minute-level). In contrast to purely data-driven methods, we undertake a “physics-guided” machine learning approach which captures salient physical features of the local wind field *without* the need to explicitly solve for those physics, including: (i) modeling wind field advection and diffusion via physically meaningful kernel functions, (ii) integrating exogenous predictors that are both meteorologically relevant and statistically significant; and (iii) linking the multi-type NWP biases to their driving mesoscale weather conditions. Tested on real-world data from the U.S. Mid Atlantic where several offshore wind projects are in-development, AIRU-WRF achieves notable improvements, in terms of both wind speed and power, relative to various forecasting benchmarks including physics-based, hybrid, statistical, and deep learning methods.

Keywords: Offshore Wind Energy, Physics-Informed Learning, Probabilistic Forecasting, Spatio-Temporal Modeling

1. Introduction

Offshore wind (OSW) is one of the fastest growing sources of renewable energy worldwide. Several countries have set ambitious targets to increase the penetration of OSW into their electricity systems. For instance, the United States (U.S.) plans to install 30 Gigawatts (GW) of OSW capacity by 2030, distributed across five major geographical regions. Among those, the U.S. Mid Atlantic and Northeastern U.S. (the focus of this work) is set to contribute more than a third of the total planned capacity. To meet this target, several GW-scale OSW projects are currently under development in this region [1, 2].

The reliable integration of those soon-to-be-operational OSW farms into the power grid will be contingent on accurate, high-resolution, short-term wind forecasts. To that end, we propose the AI-powered Rutgers University Weather Research & Forecasting (AIRU-WRF) model, in order to make accurate OSW forecasts that are short-term (minutes to hours ahead), and of high resolution, both spatially (site-specific) and temporally (minute-level). Such hyper-detailed forecasts—practically unattainable by stand-alone physics-based models—are pivotal to several OSW operations, including power estimation [3, 4], economic dispatch and reserve planning [5, 6], operations and maintenance [7, 8, 9], among others.

Obtaining high-resolution, short-term wind forecasts, however, is challenging, mainly due to the limitations of mesoscale numerical weather predictions (NWP) at finer spatial and temporal scales. The inaccurate parameterization of sub-grid physical processes, spatial averaging assumptions, and local effects largely compromise the value of NWP for (near) real-time operations. Figure 1 shows two days of hourly wind speed forecasts from a state-of-the-art mesoscale NWP model, along with 10-min co-located wind speed observations in the U.S. Mid Atlantic. When downscaled to the 10-min, site-specific resolution at which the forecasts are needed, NWP suffer from notable errors and biases, of multiple types, including temporal biases (early/late predictions) and shift biases (under- or over-prediction) [10].

To remedy those limitations, statistical and machine learning (ML) methods have emerged as a powerful approach for high-resolution, short-term wind forecasting [10, 12]. One criticism often pointed at statistical and ML methods, however, is that they are, by and large, *physics-agnostic*, i.e., they are formulated with little consideration of the physics of wind field formation and propagation, and hence, may be susceptible to model specifications that

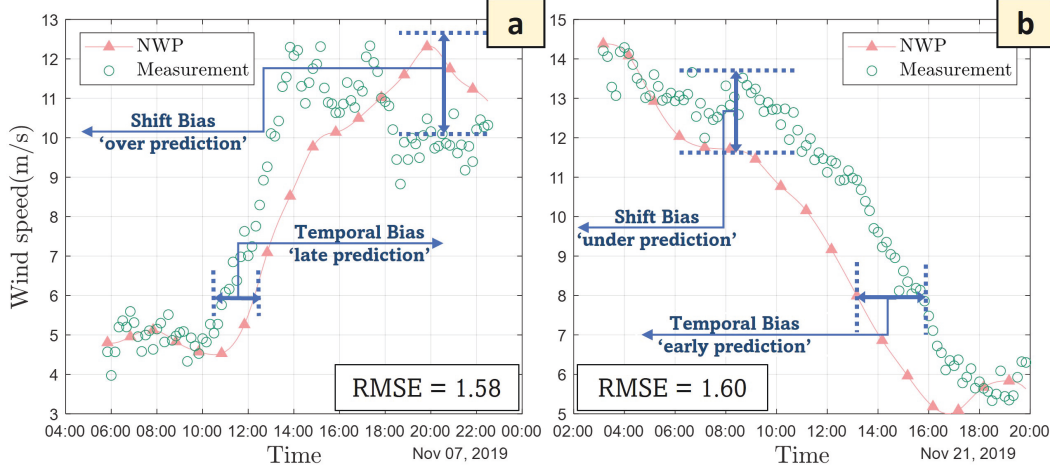


Figure 1: Two days of meso-scale NWP from the RU-WRF meso-scale model [11], along with co-located wind speed observations. Both data and forecasts are obtained in proximity to the planned OSW energy areas in the NY/NJ Bight. Despite their value, NWP often exhibit notable forecast biases, of multiple types and magnitudes, when directly downscaled to higher spatio-temporal resolutions.

violate those first principles. This has driven an active area of research in ML referred to as *physics-informed*, or *physics-guided* learning. Physics-guided learning is broadly defined as the integration of physical principles, laws, or physics-based information within ML models in order to guide them to adhere to certain physical aspects underpinning the driving physical process.

One direct approach to loosely inject physics within ML-based wind forecasting is by using NWP outputs as regressors to a statistical- or ML-based formulation. This approach is often referred to as a “hybrid” forecasting model as it attempts to calibrate the physics-based NWP at a set of target locations and time resolutions by learning a functional mapping that adjusts future NWP closer to incoming observations [13, 14, 15]. Our work broadly belongs to this family of hybrid approaches, but departs from the vast majority of the methods therein by bearing on physical knowledge to guide the selection of certain parameters, features, and statistical constructs in the ML-based forecasting model, rather than relying on a purely data-driven correction of NWP outputs. As such, our method does not seek to “replace” NWP models, but rather borrow strength across both the physical and statistical learning paradigms. In light of that, the unique aspects of this work are summarized as follows:

(A₁) The vast majority of wind forecasting models either overlook the spatio-temporal dependence in wind fields, or at best, model it using physically irrelevant kernel functions which unrestrictedly learn the spatial and temporal correlations in the historical data. In contrast, AIRU-WRF adopts a special class of physically meaningful kernels, collectively dubbed as the “Lagrangian reference framework,” for which we select the parameters in part using NWP. We show how this physically aligned covariance modeling approach encodes the first principles of wind field formation and propagation, and how it translates into significant forecast accuracy gains.

(A₂) We connect the multi-type NWP biases (such as those shown in Figure 1) with their driving mesoscale weather conditions via a physically justifiable calibration model, wherein we construct exogenous predictors that are both physically meaningful (in terms of meteorological relevance) and statistically significant (in terms of explanatory power). We dynamically update this feature input set, yielding a parsimonious statistical representation that is shown to enhance the predictive quality of the final forecasts.

(A₃) Unlike several methods that solely focus on single-valued (or point) forecasting, AIRU-WRF makes probabilistic predictions, even at locations where no or sparse data is available. We leverage this capability to generate wind field forecast “maps” (in the form of evolving two-dimensional images) for effective visualization and communication of the forecast outputs to OSW energy stakeholders and end-users.

(A₄) In terms of practical relevance, we focus on the OSW regions in the U.S. Mid Atlantic where several GW-scale projects are under development. Our models, results, and analyses can potentially provide timely insights to the developers and operators of those soon-to-be-operational wind farms.

The remainder of this paper is organized as follows. Section 2 describes the data used in this study and its relevance to the U.S. OSW wind industry. Section 3 introduces the building blocks of AIRU-WRF, followed by Section 4 where forecast evaluations, results, and discussions are presented. Section 5 concludes the paper and highlights future research directions.

2. Data Description

This work has been motivated by the ongoing large-scale OSW developments in the U.S. Mid Atlantic, and in particular, the New York (NY)/New Jersey (NJ) Bight (shown in Figure 2) [2]. We make use of two sources of data from this geographical region, with varying spatial and temporal resolutions:

(i) A set of hub-height wind speed observations collected by two floating Lidar buoys (E05 and E06); and (ii) a set of co-located NWP outputs, obtained via a state-of-the-art mesoscale meteorological model (RU-WRF). Details of both sets of data are described next.

2.1. Local observations from the NY/NJ Bight

A set of wind speed observations is obtained from two floating Lidar buoys: E05 Hudson North (E05) and E06 Hudson South (E06) [16]. The two buoys are ~ 77 km apart. The observations are recorded in 10-min intervals. We focus on the 100-m altitude, which is a common hub-height of typical wind turbines. The available data spans a total duration of 6 months, distributed into two distinct periods, representing winter and summer intervals, respectively. Specifically, the winter period spans from November 1st, 2019 to February 29th, 2020, while the summer period spans from May 5th to June 30th, 2020. Throughout both those periods, the prevailing wind direction was mostly westerly, with average wind speeds of 10.42 m/s and 8.64 m/s for the winter and summer periods, respectively, and average wind directions of 304.44° and 234.98° , for the winter and summer periods, respectively. Our choice of the study periods is attributed to the quality and completeness of the hub-height observational data during those times of the year, relative to other time periods where long streaks of missing observations prevented us from conducting extensive and reliable forecasting experiments. Our analysis of the correspondent year-round NWP data suggests that the wind conditions observed during the study period cover a wide spectrum of the meteorological regimes that typically manifest in this region.

2.2. Numerical Weather Predictions from RU-WRF

The Rutgers University Center for Ocean Observing Leadership (RU-COOL) at Rutgers University runs a daily real-time version of WRF, called RU-WRF, which is tailored to the U.S. Mid Atlantic OSW energy areas [17]. RU-WRF runs a parent nest at 9 km resolution out to 120 hours and a child nest centered on the NJ shelf at 3 km resolution out to 48 hours, generating hourly forecasts of multiple meteorological variables, which are listed in Table 1. The model initial and boundary conditions are obtained from the National Weather Service (NWS) Global Forecasting System (GFS), with the model re-initialized from GFS daily at 00Z. The continuous model data archive from December 2019 to March 2023 [18] includes hourly output compiled from the 3-km domain model.

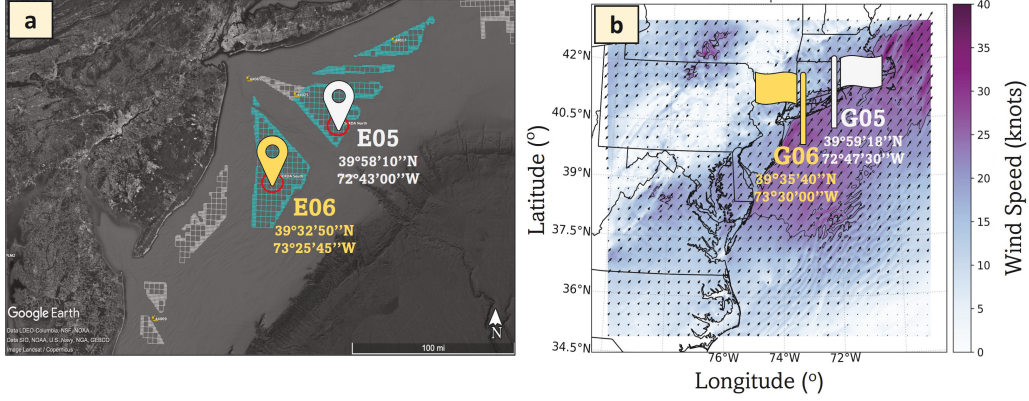


Figure 2: (a) Locations of the two buoys, E05 and E06, in the NY/NJ Bight. Green and grey regions represent OSW energy planning and lease areas, respectively [2]; (b) Locations of the two closest RU-WRF grid points, G05 and G06, on top of the NWP forecasts on November 7th, 2019 at 18:00 GMT.

RU-WRF has been recently validated by the National Renewable Energy Laboratory (NREL) [11]. The physics parameterizations and model setup in the version of RU-WRF used here are consistent with [19], however, the vertical levels have been increased from 40 to 48 to enhance resolution in the boundary layer near the sea surface, the surface layer scheme is now Mellor–Yamada–Nakanishi–Niino (MYNN) [20]. The model is a non-data assimilative system, however, it does utilize a novel ocean sea surface temperature satellite product designed to capture the nearshore coastal upwelling [21] at each daily initialization to provide an accurate bottom boundary condition within the Mid Atlantic Bight (MAB). Generally, the modeling philosophy is to leverage the highly accurate initial conditions of the NWS GFS model, but enhance local simulations through the use of higher horizontal and vertical resolutions and accurately represent the essential ocean feature of the MAB, the coastal upwelling. The nearest NWP grid points to E05 and E06 are denoted as G05 and G06, respectively, and are shown in Figure 2(b), on top of the NWPs for a select day and time in November, 2019.

3. Methodology

Let $Y(\mathbf{s}, t)$ be the random variable denoting the hub-height, spatio-temporal wind speed, where $\mathbf{s} \in \mathbb{R}^2$ is a pair of spatial coordinates, and $t \in \mathbb{Z}^+$ denotes time such that $\mathbb{Z}^+$ is the set of non-negative integers. Similarly, let $\mathbf{X}(\mathbf{s}, t) = [X_1(\mathbf{s}, t), \dots, X_p(\mathbf{s}, t)]^T$ be a set of p meteorological variables, other

Table 1: NWP variables, their descriptions, and units, as extracted from RU-WRF.

NWP variable	Description	Unit
WIND SPEED	Wind speed forecast at 100-m altitude	m/s
SWDOWN	Surface downwelling shortwave Flux	W/m ²
LWUPB	Surface upwelling longwave flux	W/m ²
GLW	Surface downwelling longwave flux	W/m ²
SNOWNC	Accumulated total grid scale snow and ice	mm
TEMP	Sea surface temperature	K
DIF_FRAC	Diffuse fraction of surface shortwave irradiance	-
LANDMASK	Land mask (1 For Land, 0 For Water)	-
LAKEMASK	Lake mask (1 For Lake, 0 For Non-Lake)	-
PBLH	Height of the top of the planetary boundary layer (PBL)	m
HUMIDITY	Relative humidity (surface level)	%
PRESSURE	Sea level pressure	hPa
MDBZ	Maximum radar reflectivity	dBZ
U	Eastward wind component at 100-m altitude	m/s
V	Northward wind component at 100-m altitude	m/s
WINDGUST	Wind gust, computed by mixing down momentum from the level at the top of the planetary boundary layer	m/s

than wind speed, for which hourly NWP forecasts are available. Examples of such variables are the NWP outputs that are listed in Table 1.

At our disposal are two sets of data with distinct spatial and temporal resolutions: (i) a set of hub-height, spatio-temporal wind speed observations denoted by $\mathbf{y}(\mathbf{s}, t) = [y(\mathbf{s}_1, t_1), \dots, y(\mathbf{s}_n, t_c)]^T$, where n is the number of measurement locations and t_c denotes the current time, and (ii) a set of correspondent NWP forecasts for $Y(\mathbf{s}, t)$, denoted by $\hat{\mathbf{y}}(\mathbf{s}, t) = [\hat{y}(\mathbf{s}_1, t_1), \dots, \hat{y}(\mathbf{s}_n, t_k)]^T$, as well as for $\mathbf{X}(\mathbf{s}, t)$, which are denoted by $\hat{\mathbf{x}}(\mathbf{s}, t)$. Note that (i) $t_k > t_c$ because NWP forecasts are available for both the historical observations (up to t_c), as well as for future forecast horizons (beyond t_c); and (ii) $\mathbf{y}$ and $\hat{\mathbf{y}}$ may not have the same temporal resolution (In our context, 10-min for the actual data $\mathbf{y}$, while NWPs, $\hat{\mathbf{y}}$, are of hourly resolution). Hereinafter, notation for *variables* will be in uppercase, while that for *data* will be in lowercase.

The overarching formulation of AIRU-WRF is shown in (1), where $\mu(\mathbf{s}, t)$ and $\eta(\mathbf{s}, t)$ are two independent spatio-temporal functions, serving two distinct purposes, while $\epsilon(\mathbf{s}, t)$ is the Gaussian white noise process. In specific, $\mu(\mathbf{s}, t)$ is intended to capture large-scale, low-frequency variations in the wind field that manifest themselves over relatively coarse time scales and spatial resolutions. We refer to this as the “mesoscale.” In contrast, $\eta(\mathbf{s}, t)$ charac-

terizes the higher-resolution, site-specific variations that $\mu(\mathbf{s}, t)$ typically fails to capture. We refer to this as the “local” or “sub-mesoscale” variation.

$$Y(\mathbf{s}, t) = \underbrace{\mu(\mathbf{s}, t|\mathbf{y}, \hat{\mathbf{y}}, \hat{\mathbf{x}})}_{\text{Mesoscale variation}} + \underbrace{\eta(\mathbf{s}, t|\mathbf{y}, \hat{\mathbf{x}})}_{\text{Local variation}} + \underbrace{\epsilon(\mathbf{s}, t)}_{\text{White noise}}. \quad (1)$$

Next, we discuss the role and formulation of $\mu(\mathbf{s}, t)$ and $\eta(\mathbf{s}, t)$ in Sections 3.1 and 3.3, respectively.

3.1. Physics-guided modeling of $\mu(\mathbf{s}, t)$

The role of $\mu(\mathbf{s}, t)$ is to characterize the larger-scale variation in the wind field, which are mostly driven by physical phenomena that manifest themselves over relatively longer time scales (hours to days) and coarser spatial resolutions (mesoscale). Examples of those large-scale variations include trends, diurnal and semi-diurnal cycles, regime alternations, etc. Our assumption herein is that NWP can play a key role in capturing such larger-scale fluctuations by virtue of their embedded physics. As such, we can think of $\mu(\mathbf{s}, t)$ as an NWP calibration, except that it is a physically motivated one.

The first step of this calibration is to interpolate the hourly NWP variables $\hat{Y}(\mathbf{s}, t)$ and $\hat{\mathbf{X}}(\mathbf{s}, t)$ into the 10-min resolution at which the forecasts are to be made, yielding the interpolated NWP variables $\tilde{Y}(\mathbf{s}, t)$ and $\tilde{\mathbf{X}}(\mathbf{s}, t)$, respectively. For interpolation of both sets, we use cubic splines. Let us then denote by $\tilde{\mathbf{G}}(\mathbf{s}, t) = [\tilde{G}_1(\mathbf{s}, t), \dots, \tilde{G}_m(\mathbf{s}, t)]^T$ the set of m spatio-temporal explanatory variables, which are to be included as regressors in modeling $\mu(\mathbf{s}, t)$. For example, $\tilde{\mathbf{G}}(\mathbf{s}, t)$ may include the “most informative” subset of $\tilde{\mathbf{X}}(\mathbf{s}, t)$, in addition to other exogenous variables that possess some degree of explanatory power in calibrating NWP. In light of that, we propose the following form for $\mu(\mathbf{s}, t)$:

$$\mu(\mathbf{s}, t) = \underbrace{\mathbf{a}^T \tilde{\mathbf{Y}}^\ell(\mathbf{s}, t) + \mathbf{b}^T \tilde{\mathbf{G}}(\mathbf{s}, t)}_{\text{additive bias correction}} + \underbrace{\mathbf{c}^T \tilde{\mathbf{G}}(\mathbf{s}, t) \tilde{Y}(\mathbf{s}, t)}_{\text{multiplicative bias correction}}, \quad (2)$$

where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are sets of unknown parameters (to be estimated), and $\tilde{\mathbf{Y}}^\ell(\mathbf{s}, t) = [\tilde{Y}(\mathbf{s}, t), \dots, \tilde{Y}(\mathbf{s}, t - \ell)]^T$ denotes the set of (interpolated) lagged NWP forecasts of wind speed, up to lag ℓ . The motivation behind (2) is to correct the multi-type biases of NWP by linking them to their driving mesoscale weather conditions encoded in $\tilde{\mathbf{G}}(\mathbf{s}, t)$ and $\tilde{\mathbf{Y}}^\ell(\mathbf{s}, t)$.

The formulation in (2) targets two distinct types of NWP biases: Additive bias refers to the systematic NWP inaccuracies that can be corrected

by scaling. Examples are shift biases (over- or under-prediction), temporal biases (early/late prediction), and spatial biases (location-dependent errors)—Recall Figure 1. The inclusion of lagged values in $\tilde{\mathbf{Y}}^\ell(\mathbf{s}, t)$ is motivated by their potential role in especially correcting the temporal biases. The multiplicative bias term, on the other hand, addresses nonlinear biases that cannot be simply adjusted by scaling, such as regime-specific NWP errors (e.g., errors that are more pronounced at higher or lower winds).

The next key question is to identify the variables constituting the elements of the set $\tilde{\mathbf{G}}(\mathbf{s}, t)$. Our approach to construct $\tilde{\mathbf{G}}(\mathbf{s}, t)$ is to include variables which are both physically meaningful (in terms of meteorological relevance) *and* statistically significant (in terms of explanatory power), thus forming the basis of our physics-guided calibration. As a starter, $\tilde{\mathbf{G}}(\mathbf{s}, t)$ will include a subset of the NWP variables in $\tilde{\mathbf{X}}(\mathbf{s}, t)$ which are known to possess a meteorological association with wind field physics. Out of the NWP outputs of Table 1, prior physical knowledge suggests the potential inclusion of six features: air pressure, surface temperature, wind gust, relative humidity, eastward and northward wind components, since those variables are physically known to contribute, either directly or indirectly, to local wind field formation and propagation. From a purely statistical perspective, those features also exhibit intermediate to strong association with the wind speed observations, with Pearson’s correlations ranging on average between 0.87 for wind gust and -0.20 for surface temperature.

In addition to those six variables, we construct additional features that are not readily forecast by mesoscale models, but are derived from NWP outputs. We postulate the construction of two additional features: the spatio-temporal pressure differential and the geostrophic wind, which are described below.

Spatio-temporal pressure differentials:

Coarsely simplifying the physical processes involved, winds are generated by the movement of air from high- to low-pressure locations; the larger the pressure difference, the stronger the winds. If $P(\mathbf{s}_i, t)$ denotes the pressure at location $\mathbf{s}_i$ at time t , then, the spatio-temporal pressure differential between two locations $\mathbf{s}_i$ and $\mathbf{s}_j$ at two time instances t and $t + d$, is defined as:

$$\Delta_p(\mathbf{s}_i, \mathbf{s}_j, d) = P(\mathbf{s}_i, t) - P(\mathbf{s}_j, t + d), \quad (3)$$

where $d \in \mathbb{Z}$ is a time lag. Note that d can be both positive or negative, that is, we consider both past and future lags.

Geostrophic wind:

The geostrophic wind is the flow that results from a balance between the Coriolis acceleration from the Earth's rotation and the horizontal pressure gradient force at relatively high altitudes ($\sim 1000\text{m}$). Geostrophic wind has been shown to improve the accuracy of short-term wind speed forecasts [22]. Specifically, the geostrophic wind can be expressed as in (4), with u_g and v_g denoting its eastward and northward components, respectively.

$$\begin{aligned} u_g &= -\frac{g}{2\Omega \sin \rho} \frac{\partial H}{\partial y_c}, \\ v_g &= \frac{g}{2\Omega \sin \rho} \frac{\partial H}{\partial x_c}, \end{aligned} \tag{4}$$

where H is the geopotential height, g is the gravitational acceleration, Ω is the Earth rotation rate, ρ is latitude, and x_c and y_c are local eastward and northward Cartesian coordinates. To compute u_g and v_g , the geopotential height H is first estimated through the hydrostatic equation, which is expressed for isothermal atmosphere as in (5).

$$H = H_i + \frac{R\bar{T}}{g} \ln \left(\frac{p_i}{p_0} \right), \tag{5}$$

where H_i , p_i are the geopotential height and pressure at barometer i , respectively, p_0 is the reference pressure (850 hPa), R is the gas constant ($287 \text{ JK}^{-1} \text{ kg}^{-1}$) and $\bar{T}$ is the layer-averaged temperature between p_i and p_0 , which we approximate by the surface temperature. We acknowledge that relaxing the assumption of isothermal atmosphere by leveraging temperature information from RU-WRF may potentially provide more physically meaningful estimates of the geostrophic winds. This is an area of ongoing research.

From there, we follow a similar procedure to that proposed in [22] where a spatial response surface of the geopotential height, as in (6), is estimated. A video showing the geopotential height response surface over time is included in the supplemental materials appended to this article (SM-1).

$$H(x_c, y_c) = c_0 + c_1 x_c + c_2 y_c. \tag{6}$$

Setting $\frac{\partial H}{\partial x_c} = c_1$ and $\frac{\partial H}{\partial y_c} = c_2$, we can compute u_g and v_g , both of which are then used to estimate the final geostrophic wind as $Y_g(\mathbf{s}, t) = \sqrt{u_g^2 + v_g^2}$.

3.2. Selecting the “right” features in $\mu(\mathbf{s}, t)$

Putting the above pieces together, we have the following eight features as the potential constituents of $\tilde{\mathbf{G}}(\mathbf{s}, t)$: wind gust, air pressure, surface temperature, relative humidity, eastward and northward wind components, spatio-temporal pressure differential, and geostrophic wind. We also consider including lagged versions for each of those eight variables, as motivated by their potential role in correcting temporal biases.

Suppose we only include four hourly lags for each of the eight variables listed above. This corresponds to 24 lags in 10-min resolution (4 hours $\times$ 6 ten-minute intervals per hour). Hence, we end up with 8 variables $\times$ 24 lags = 192 regressors for inclusion in $\tilde{\mathbf{G}}(\mathbf{s}, t)$. However, not all features are expected to be relevant at all times. In fact, more often than not, using an excessively large set of predictors does not coincide with the best predictive performance [23] (the law of parsimony in ML). This also aligns with the prior physical knowledge: The drivers of NWP bias change over space-time, resulting in distinct NWP bias types and magnitudes. Thus, including a feature in the set $\tilde{\mathbf{G}}(\mathbf{s}, t)$ at a certain time does not justify its inclusion at other time instances. A dynamic feature selection mechanism is therefore needed to continuously identify and update a minimally sized subset of information-rich exogenous variables, constituting the elements of the set $\tilde{\mathbf{G}}(\mathbf{s}, t)$.

Given the time resolution at which the first set of forecasts are to be made (10-min ahead), advanced feature selection techniques that rely on iterative model estimation are practically prohibitive. Thus, we revert to simple (but effective) measures of explanatory power, namely partial autocorrelation functions (PACFs) to determine the time lag ℓ of $\tilde{\mathbf{Y}}^\ell$, and Pearson’s correlation to select the features in $\tilde{\mathbf{G}}(\mathbf{s}, t)$. To ensure parsimony and avoid multicollinearity, we impose a simple rule: For the features in $\tilde{\mathbf{G}}(\mathbf{s}, t)$, we only select the most correlated lagged version of the same variable, i.e. the one that has the maximal correlation with the target response.

Figure 3(a-h) shows the Pearson’s correlations (in absolute value) between all eight variables of Section 3.1 and the actual wind speed observations across all of the forecasting rolls. Looking at Figure 3, two insights are immediately drawn. First, it is clear how the explanatory power of a feature can significantly vary over time. As an example, relative humidity (RH) can record correlations that reach up to ~ 0.8 at some time instances (Rolls 320-330), while having extremely low correlations at others (around Rolls 95-100). This suggests that this feature should only be included when it matters, i.e. at times when it can positively contribute towards explaining the variability

in the predictand. Second, we note how selecting a lagged version of an exogenous feature can noticeably enhance its explanatory power. As a case in point, a lagged version of the northward wind component V (red solid line) appears to be a much more informative predictor (correlation around 0.6, Roll ~ 85) relative to the value of V itself (correlation around 0.45; blue solid line). This is also valid with most features, in particular RH, V , P , GW, and STPD. Figure 3(k) shows the change in m (the number of features in $\tilde{\mathbf{G}}(\mathbf{s}, t)$) over time, showing how our feature selection dynamically adds or drops certain features over time, depending on their statistical relevance at the time of forecast, while keeping the input feature space reasonably sparse.

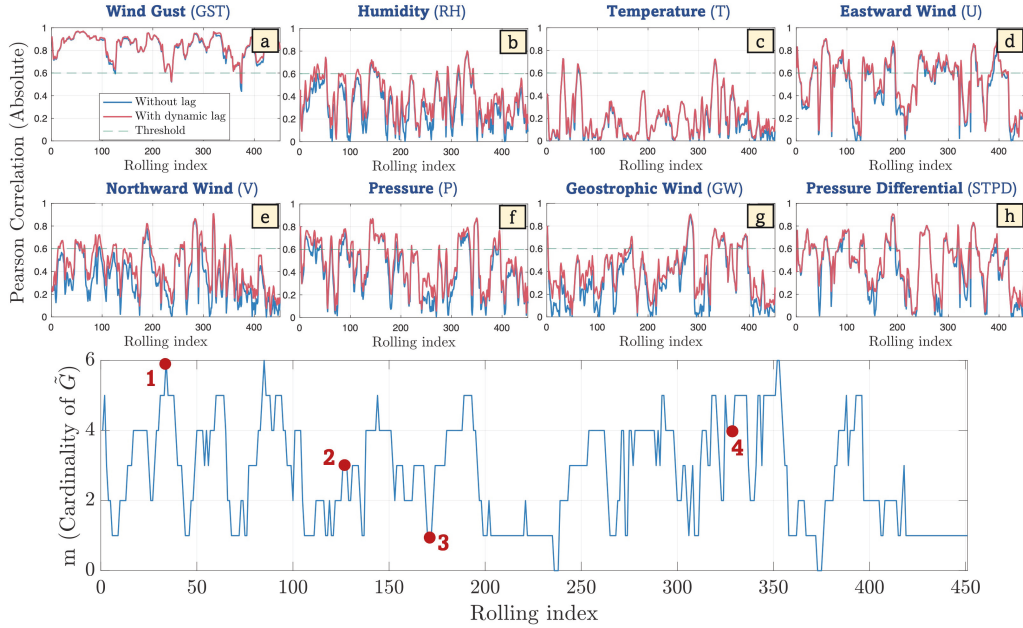


Figure 3: (a)-(h) Pearson's correlations between eight variables and the wind speeds over 451 forecast rolls (The green horizontal line depicts a threshold of 0.6). Blue and red lines correspond to the correlation of the variable (with the predictand) versus that of a lagged version thereof. Panel (k) shows the change in m (cardinality of $\tilde{\mathbf{G}}$) over time. Red points 1-4 correspond to the following selected subsets (subscript denotes the chosen lag in 10-min intervals): (1) $\tilde{\mathbf{G}} = \{\text{GST}_{-6}, \text{RH}_{18}, \text{T}_7, \text{U}_{10}, \text{P}_{-18}, \text{STPD}_1\}$; (2) $\tilde{\mathbf{G}} = \{\text{GST}_{-2}, \text{U}_{-4}, \text{STPD}_{-1}\}$; (3): $\tilde{\mathbf{G}} = \{\text{GST}_{-3}\}$; (4): $\tilde{\mathbf{G}} = \{\text{GW}_0, \text{GST}_{10}, \text{RH}_{14}, \text{P}_{18}\}$

3.3. Physics-guided modeling of $\eta(\mathbf{s}, t)$

While $\mu(\mathbf{s}, t)$ is intended to capture large-scale fluctuations in the wind field, the role of $\eta(\mathbf{s}, t)$, on the other hand, is to characterize the higher-frequency variations (site-specific, minutes to hours) which are typically driven by sub-mesoscale local effects that NWP models may fail to capture. We decide to model $\eta(\mathbf{s}, t)$ as a spatio-temporal Gaussian Process (GP) [24, 25]. 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 the vector of spatio-temporal residuals, such that $Z(\mathbf{s}, t) = Y(\mathbf{s}, t) - \mu(\mathbf{s}, t)$. We regard the vector $\mathbf{z}$ as a realization of a spatio-temporal GP, $\mathcal{Z}(\cdot) \sim \mathcal{GP}(\mathcal{M}(\mathbf{s}, t), K(\boldsymbol{\gamma}, w))$, such that $\mathcal{M}(\mathbf{s}, t)$, and $K(\boldsymbol{\gamma}, w)$ are the GP mean and covariance (or kernel) functions, respectively, wherein $\boldsymbol{\gamma} \in \mathbb{R}^2$ and $w \in \mathbb{Z}^+$ are the spatial and temporal lags, respectively, while $\mathbb{Z}^+$ is the set of non-negative integers.

The key challenge in GPs is to propose a suitable, mathematically permissible form for $K(\boldsymbol{\gamma}, w) : \mathbb{R}^2 \times \mathbb{Z}^+ \rightarrow \mathbb{R}$, which adequately captures the spatio-temporal dependence and enables GP-based forecasting. The most prevalent approach to specify $K(\boldsymbol{\gamma}, w)$ in the literature is through the so-called separable approach, which decomposes the dependence structure over space and time such that $K(\boldsymbol{\gamma}, w) = K^s(\boldsymbol{\gamma}) \times K^t(w)$, wherein $K^s(\boldsymbol{\gamma})$ and $K^t(w)$ are two covariance structures for space and time, respectively [24]. Popular selections for $K^s(\boldsymbol{\gamma})$ and $K^t(w)$ include the squared exponential and Matérn covariance functions [26].

Despite its simplicity, the disconnect between space and time in the separable approach yields model specifications that violate the physical property of wind advection, i.e., the propagation of wind along a certain prevailing direction. This is because the separable approach assumes that space-time correlations are symmetric, i.e. $\text{cor}(Y(\mathbf{s}_i, t), Y(\mathbf{s}_{i'}, t+\omega)) = \text{cor}(Y(\mathbf{s}_{i'}, t), Y(\mathbf{s}_i, t+\omega))$ [27, 28]. In reality, wind advection induces an asymmetry in the value of information, that is, along-wind dependence is expected to be stronger than opposite-wind dependence. In other words, $\text{cor}(Y(\mathbf{s}_i, t), Y(\mathbf{s}_{i'}, t+\omega)) - \text{cor}(Y(\mathbf{s}_{i'}, t), Y(\mathbf{s}_i, t+\omega)) > 0$ when i' is downstream of i .

To capture this physical property, we propose to adopt a class of covariance models known in the geostatistical literature as the Lagrangian reference framework [29, 30], which is capable of mimicking the “advection” of spatio-temporal information by having the following form:

$$K(\boldsymbol{\gamma}, w) = \mathbb{E}_{\Theta} \{ \psi(\boldsymbol{\gamma} - \Theta w) \}, \quad (7)$$

where $\psi(\cdot)$ is a positive-definite kernel. The advection vector $\Theta \in \mathbb{R}^2$ represents the prevailing flow and hence, its specification is instrumental

to Lagrangian models. Here, we model Θ as a bivariate normal random vector, since wind is not only governed by advection, but also diffusion, i.e. the random propagation of wind in other directions [31, 32]. Letting $\Theta \sim \mathcal{N}_d(\mu_\Theta, \Sigma_\Theta)$, $\psi(x) = \exp(-x^2)$, and $\mathbf{F} = \mathbf{I}_{d \times d} + 2\Sigma_\Theta w^2$, yields the following expression for the Lagrangian correlation model:

$$K_{LG}(\gamma, w) = \frac{1}{\sqrt{|\mathbf{F}|}} \exp \left\{ -(\gamma - \mu_\Theta w)^T \mathbf{F}^{-1} (\gamma - \mu_\Theta w) \right\}, \quad (8)$$

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

We propose to estimate μ_Θ and Σ_Θ in light of the NWP forecasts of the eastward and northward wind components as reasonable representations of the prevailing flow during the forecast horizon. As shown in (9) and (10), our estimates for μ_Θ and Σ_Θ will be the spatio-temporal empirical averages and covariances of the NWP forecasts in the window $[t_c - T, t_c + H_T]$, where t_c is the current time, while T and H_T are the lengths of the training data and of the forecast horizon, both in 10-min intervals.

$$\mu_\Theta = [\mu_1, \mu_2]^T = [\bar{u}, \bar{v}]^T, \quad (9)$$

$$\Sigma_\Theta = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix} = \begin{bmatrix} \text{cov}(\mathbf{u}, \mathbf{u}) & \text{cov}(\mathbf{u}, \mathbf{v}) \\ \text{cov}(\mathbf{v}, \mathbf{u}) & \text{cov}(\mathbf{v}, \mathbf{v}) \end{bmatrix}, \quad (10)$$

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

Our final covariance function $K_\eta(\mathbf{s}, t)$ is a convex combination of a separable correlation function and the Lagrangian correlation function of (8):

$$K_\eta(\gamma, w) = \alpha \left[\lambda \overbrace{(K_{SE}^s(\gamma) \times K_{SE}^t(w))}^{\text{Separable Kernel}} + (1 - \lambda) \overbrace{K_{LG}(\gamma, w)}^{\text{Lagrangian Kernel}} \right] + \mathbb{I}_{\{||\gamma||=|w|=0\}} \delta, \quad (11)$$

where $\alpha > 0$ is the marginal variance parameter, $\lambda \in [0, 1]^T$ is the asymmetry coefficient denoting the strength of symmetry (or lack thereof), and $\delta > 0$ denotes the noise variance, whereas $\mathbb{I}(\cdot)$ is the indicator function and $||\cdot||$ is the Euclidean norm. The terms $K_{SE}^s(\cdot)$ and $K_{SE}^t(\cdot)$ are squared exponential (SE) correlation functions for space and time, respectively.

3.4. Estimation and probabilistic forecasting

The parameters to be estimated are: (1) $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ denote the sets of parameters in the calibration term $\mu(\mathbf{s}, t)$ in (2); (2) $\boldsymbol{\mu}_\Theta$ and $\boldsymbol{\Sigma}_\Theta$ are the advection vector parameters; (3) the marginal variance α and asymmetry parameter λ in (11), and the GP mean parameter $\mathcal{M}(\mathbf{s}, t) = \beta_0$; (4) the range parameters of $K_{SE}^s(\cdot)$ and $K_{SE}^t(\cdot)$, denoted by r_s and r_t , respectively; and (5) the noise variance, δ . We estimate the parameters sequentially: First, we estimate $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ via least squares, then, use the residuals to estimate the remainder of the parameters using maximum likelihood estimation, except for $\boldsymbol{\mu}_\Theta$ and $\boldsymbol{\Sigma}_\Theta$ which are estimated as in (9)-(10).

The joint predictive distribution of the final set of spatio-temporal forecasts, $\hat{\mathbf{f}} = [\hat{f}(\mathbf{s}_1, t_c + 1), \dots, \hat{f}(\mathbf{s}_n, t_c + H_T)]^T$, is fully characterized by virtue of the GP framework. In specific, the predictive mean and variance at each look-ahead time and spatial location $\hat{f}(\mathbf{s}_i, t_c + h)$ are obtained as:

$$\hat{f}(\mathbf{s}_i, t_c + h) = \hat{\mu}(\mathbf{s}_i, t_c + h) + \hat{\beta}_0 + \hat{\mathbf{k}}^T \bar{\mathbf{K}}_\eta^{-1} (\mathbf{y} - \bar{\boldsymbol{\mu}} - \hat{\beta}_0 \mathbf{I}), \quad (12)$$

$$\begin{aligned} \hat{\sigma}^2(\mathbf{s}_i, t_c + h) = & \hat{K}_\eta(\mathbf{0}, 0) - \hat{\mathbf{k}}^T \bar{\mathbf{K}}_\eta^{-1} \hat{\mathbf{k}} \\ & + (1 - (\hat{\mathbf{k}}^T \bar{\mathbf{K}}_\eta^{-1} \mathbf{I}))^T (\mathbf{I}^T \bar{\mathbf{K}}_\eta^{-1} \mathbf{I})^{-1} (1 - [\hat{\mathbf{k}}^T \bar{\mathbf{K}}_\eta^{-1} \mathbf{I}]), \end{aligned} \quad (13)$$

where $h \in \{1, \dots, H_T\}$ is the forecast horizon (in 10-min intervals), $\mathbf{I}$ is an $n \cdot H_T \times 1$ column of 1's, and $\bar{\boldsymbol{\mu}} = [\mu(\mathbf{s}_1, t_1), \dots, \mu(\mathbf{s}_n, t_c)]^T$ is the vector of evaluations of $\mu(\mathbf{s}, t)$ for the training data. Similarly, $\bar{\mathbf{K}}_\eta$ is the training covariance matrix evaluated using the estimated kernel $\hat{K}_\eta(\cdot, \cdot)$. The vector $\hat{\mathbf{k}}$ contains the pairwise covariances between $\mathbf{z} = \mathbf{y} - \bar{\boldsymbol{\mu}}$ and $z(\mathbf{s}_i, t_c + h)$.

Two unique advantages of AIRU-WRF are: (1) its ability to make full probabilistic inference about the joint distribution of the spatio-temporal forecasts [33], which can be used to generate trajectories that naturally embed the spatial and temporal dependence for operational decisions (e.g., for use within a stochastic program) [34, 8]; (2) The ability to produce “spatial wind field forecast maps,” in the form of two-dimensional images, including at locations where no measurements are available. Both of those capabilities are demonstrated in Section 4.

4. Real-world Experiments and Discussions

We focus on short-term forecasts (up to six hours) in 10-min resolution, i.e. $h \in \{1, \dots, 36\}$. We evaluate AIRU-WRF at E05 and E06 (where data are available), and visualize its forecasts at nearby sites where no data exist.

4.1. Training and testing

We test our approach using a rolling forecasting scheme, wherein for each forecast roll, we perform our feature selection procedure, re-train the model, obtain the forecasts, then roll by 6 hours, and repeat the whole process. For the four-month winter period, this is correspondent to 451 rolls, yielding a total of $6 \text{ forecasts/hour} \times 6\text{-hour horizon} \times 451 \text{ rolls} \times 2 \text{ spatial sites} = 32,472$ testing instances. For the two-month summer period, this is correspondent to 216 rolls, yielding a total of $6 \text{ forecasts/hour} \times 6\text{-hour horizon} \times 216 \text{ rolls} \times 2 \text{ spatial sites} = 15,552$ testing instances. We find that five days of historical data is a sufficient training data size to balance model fitting and computational efficiency. The maximum time to complete one forecasting roll (including feature selection, training, and forecasting) was 2.52 minutes on a standard laptop with an i7-6700HQ Intel processor, 2.60GHz base frequency, and 16 GB RAM. We also find that using NWP for sub-hourly forecasts (i.e., $h < 6$, or ultra-short-term forecasts) is, on average, not helpful. So we exclude the term $\mu(\mathbf{s}, t)$ when making sub-hourly forecasts, and include it for all other forecast horizons (i.e. for $h \geq 6$).

4.2. Benchmarks and evaluation metrics

We compare AIRU-WRF against five benchmarks, $\mathcal{B}1$ - $\mathcal{B}5$:

- ($\mathcal{B}1$) GOP: The Geostatistical Output Perturbation (GOP), proposed in [35], is a hybrid model to calibrate mesoscale NWP using local observations. GOP is regarded as a special case of AIRU-WRF by assuming $\mu^{\text{GOP}}(\mathbf{s}, t) = \mathbf{a}^T \tilde{\mathbf{H}}(\mathbf{s}, t) + (\mathbf{b}^T \tilde{\mathbf{H}}(\mathbf{s}, t)) \tilde{Y}(\mathbf{s}, t)$, where $\tilde{\mathbf{H}}(\mathbf{s}, t)$ is the set of NWP variables directly obtained from RU-WRF (namely, air pressure, surface temperature, wind gust, relative humidity, eastward and northward wind components), while $\eta^{\text{GOP}}(\mathbf{s}, t)$ is assumed to be a GP with a spatial SE kernel. AIRU-WRF generalizes on GOP on two main fronts: (1) Extension to the spatio-temporal setting via a physically meaningful covariance function; (2) A physics-guided calibration of the NWP through the construction and dynamic updating and selection of physically meaningful predictors encoded in the set $\mathbf{G}(\mathbf{s}, t)$.
- ($\mathcal{B}2$) NWP: Those are the raw forecasts from RU-WRF (statistically interpolated to the 10-min resolution).
- ($\mathcal{B}3$) ARIMAX(p, q, d): We train a separate Autoregressive Integrated Moving Average with Exogenous inputs (ARIMAX) model for each forecast

location, with $\tilde{\mathbf{H}}$ as the set of exogenous inputs (the same set used in GOP). All model parameters (p , q , and d) as well as the autoregressive, moving average, and exogenous coefficients are optimized using the `pmdarima` package in `Python`.

- (B4) LSTM: Long short-term memory networks is a class of deep learning models based on recurrent neural networks that is well-suited for time series data [36]. We fit a separate LSTM for each location and use stochastic gradient descent to optimize the LSTM network parameters through the `Deep Learning Toolbox` in `Matlab`. A grid search was used to tune the number of hidden units for the LSTM (set at 100), number of epochs (set at 100), and the learning rate (start = 0.005, decreased by a factor of 0.1 every 30 epochs).
- (B5) PER: Persistence (PER) forecasting is a standard benchmark that assumes present weather conditions will persist into the future. It is known to perform well at ultra-short-term horizons.

To evaluate the forecasts, we use the Mean Absolute Error (MAE) and the Continuous Ranked Probability Score (CRPS) for point and probabilistic wind speed forecasting, respectively. We only compute CRPS for probabilistic approaches, namely AIRU-WRF, GOP, and ARIMAX. We convert the wind speed forecasts from all competing models into wind power predictions using statistically constructed wind power curves (more on that in Section 4.4). We use the power curve error (PCE) loss to evaluate the wind power predictions [37].

4.3. Wind speed forecasting results

Tables 2 and 3 show the MAE and CRPS values for the six models (including AIRU-WRF), at both sites, aggregated in hourly intervals for winter and summer periods, respectively. Overall, AIRU-WRF outperforms statistical methods (ARIMAX, PER) by 16.8-27.1%, physics-based models (NWP) by 16.6-22.5%, hybrid methods (GOP) by 18.5-21.4%, and deep learning-based methods (LSTM) by 19.1-29.1%. A deeper analysis of the results in Tables 2 and 3 yields useful insights, as discussed next.

First, statistical methods (ARIMAX, PER) appear to perform well in ultra-short-term horizons (first three hours, $h < 18$) relative to NWPs. This aligns with the general consensus in the wind forecasting literature [10, 38]. The trend is reversed for longer forecast horizons (last three hours, $h > 18$)

Table 2: Wind speed forecasting results for the four-month winter period (November, 2019 to February, 2020): MAE (left) and CRPS (right), aggregated in hourly intervals, for sites E05 (Top) and E06 (Bottom). Bold-faced values denote the best performance.

E05 (39°58'10"N and 72°43'00"W)									
MAE							CRPS		
Horizon (hrs)	AIRU-WRF	GOP	NWP	ARIMAX	LSTM	PER	AIRU-WRF	GOP	ARIMAX
1	0.753	0.922	1.657	0.794	0.791	0.743	0.575	0.742	0.643
2	1.267	1.628	1.601	1.333	1.300	1.287	0.957	1.205	1.039
3	1.451	1.750	1.592	1.587	1.716	1.708	1.094	1.282	1.212
4	1.478	1.798	1.586	1.860	2.150	2.164	1.133	1.319	1.407
5	1.561	1.859	1.591	2.055	2.499	2.495	1.186	1.358	1.536
6	1.651	2.052	1.741	2.287	2.782	2.801	1.274	1.490	1.656
Average	1.360	1.668	1.631	1.653	1.873	1.866	1.037	1.233	1.249
% Improvement	-	18.5%	16.6%	17.7%	27.4%	27.1%	-	15.9%	17.0%
E06 (39°32'50"N and 73°25'45"W)									
MAE							CRPS		
Horizon (hrs)	AIRU-WRF	GOP	NWP	ARIMAX	LSTM	PER	AIRU-WRF	GOP	ARIMAX
1	0.727	0.975	1.621	0.767	0.805	0.729	0.556	0.753	0.614
2	1.271	1.702	1.691	1.347	1.372	1.277	0.969	1.291	1.018
3	1.530	1.902	1.730	1.663	1.855	1.753	1.193	1.447	1.256
4	1.558	1.978	1.801	1.942	2.235	2.156	1.216	1.470	1.463
5	1.592	1.973	1.706	2.077	2.566	2.504	1.235	1.440	1.515
6	1.584	1.988	1.659	2.137	2.827	2.779	1.252	1.441	1.557
Average	1.377	1.753	1.701	1.656	1.943	1.866	1.070	1.307	1.237
% Improvement	-	21.4%	19.1%	16.8%	29.1%	26.2%	-	18.1%	13.5%

Table 3: Wind speed forecasting results for the two-month summer period (May, 2020 to June, 2020): MAE (left) and CRPS (right) values, aggregated in hourly intervals, for sites E05 (Top) and E06 (Bottom). Bold-faced values denote the best performance.

E05 (39°58'10"N and 72°43'00"W)									
MAE							CRPS		
Horizon (hrs)	AIRU-WRF	GOP	NWP	ARIMAX	LSTM	PER	AIRU-WRF	GOP	ARIMAX
1	0.763	1.176	2.086	0.904	0.895	0.817	0.614	0.931	0.729
2	1.413	2.044	2.120	1.649	1.558	1.471	1.052	1.548	1.319
3	1.702	2.054	2.048	1.999	1.988	1.849	1.293	1.484	1.545
4	1.872	2.193	2.095	2.295	2.338	2.206	1.501	1.518	1.688
5	1.987	2.301	2.158	2.461	2.532	2.490	1.565	1.731	1.703
6	1.997	2.441	2.023	2.621	2.722	2.760	1.566	1.851	1.794
Average	1.622	2.035	2.088	1.988	2.006	1.932	1.265	1.511	1.463
% Improvement	-	20.3%	22.3%	18.4%	19.1%	16.0%	-	16.2%	13.5%
E06 (39°32'50"N and 73°25'45"W)									
MAE							CRPS		
Horizon (hrs)	AIRU-WRF	GOP	NWP	ARIMAX	LSTM	PER	AIRU-WRF	GOP	ARIMAX
1	0.787	1.103	2.098	0.856	0.882	0.805	0.636	0.913	0.685
2	1.501	1.856	2.059	1.671	1.593	1.514	1.072	1.408	1.267
3	1.807	2.085	2.115	2.030	2.056	1.958	1.382	1.530	1.570
4	1.870	2.251	2.115	2.228	2.361	2.266	1.478	1.583	1.629
5	1.880	2.342	2.085	2.380	2.612	2.505	1.495	1.702	1.706
6	1.951	2.583	2.172	2.618	2.901	2.809	1.580	1.956	1.815
Average	1.633	2.037	2.107	1.964	2.068	1.976	1.274	1.515	1.445
% Improvement	-	19.8%	22.5%	16.9%	21.0%	17.4%	-	15.9%	11.9%

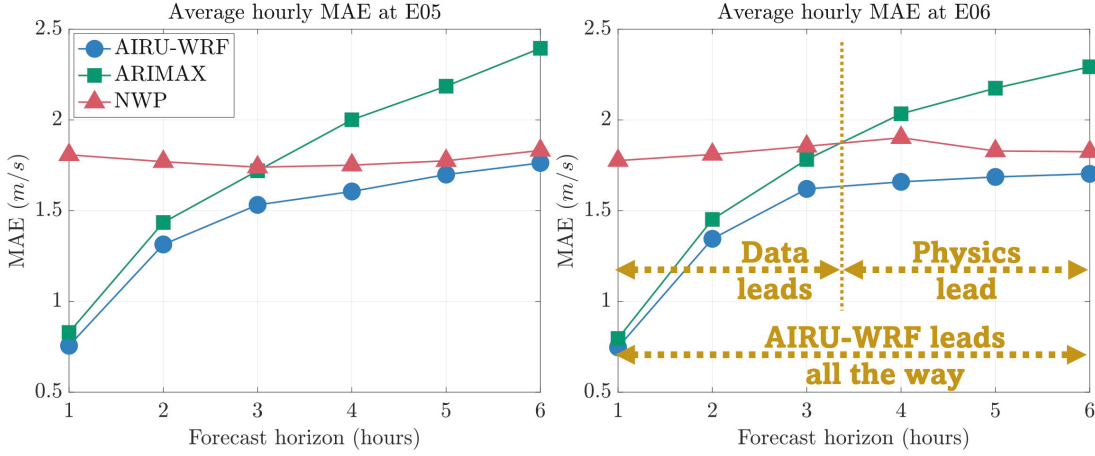


Figure 4: Average MAE (across $451 + 216 = 667$ forecast rolls), aggregated in hourly intervals, for the forecasts made via AIRU-WRF (blue circles), NWP (red triangles), and ARIMAX (green squares) for E05 (left) and E06 (right).

by virtue of the embedded physics within NWP’s enabling them to reliably extrapolate at longer horizons. AIRU-WRF, on the other hand, outperforms both sets of methods in almost all forecast horizons (with the exception of the first hour in E05, where AIRU-WRF comes in as a close second to PER). This is clearly demonstrated in Figure 4 showing how AIRU-WRF borrows strength across the physical and statistical learning paradigms, yielding forecasts that are superior to both across all forecast horizons.

Second, Figure 4, as well as Tables 2 and 3 show that the improvement from AIRU-WRF over NWP’s is maximal in short-term horizons where the “learning from data” aspect of AIRU-WRF provides a significant advantage relative to a purely physics-based model. This margin of improvement decays as the forecast horizon becomes longer, with forecasts from AIRU-WRF gradually converging towards their NWP counterparts, yet still maintaining a non-negligible lead, especially for E06. On the contrary, the advantage of AIRU-WRF over purely statistical methods peaks at longer forecast horizons (up to 43% at the sixth hour), where the embedded physics in AIRU-WRF enable it to safely extrapolate at longer horizons where solely depending on data-driven learning is typically unreliable.

Third, comparing a deep learning (DL) method like LSTM with statistical methods (ARIMAX, PER) suggests that their performance is comparable in short-term forecast horizons ($\sim$ first four hours), with the latter gradually overtaking the former as the forecast horizon extends. We speculate that DL may need a much larger dataset (in terms of the number of locations and

measurements) to unleash its full potential. Having a dense network of observations in the U.S. Mid Atlantic (and other similarly under-explored OSW areas) is, however, not practical, at least in the foreseeable future. AIRU-WRF, on the other hand, performs significantly better than all physics-free methods (be it statistical- or DL-based), in almost all forecast horizons.

Finally, comparing AIRU-WRF to a seminal hybrid method (GOP) demonstrates the merit of the physics-guided modeling of $\mu(\mathbf{s}, t)$ and $\eta(\mathbf{s}, t)$. The largest improvement from AIRU-WRF relative to GOP is realized at shorter forecast horizons, where the benefit of invoking a physically meaningful spatio-temporal kernel materializes. Noticeable improvements at longer horizons are still maintained, and are likely attributed to the construction and integration of physically relevant features within AIRU-WRF.

Figure 5(c) shows AIRU-WRF’s probabilistic forecasts suggesting a faithful alignment with actual observations. The ability of AIRU-WRF to adjust forecast biases is demonstrated in Figure 5(a)-(b), where the true versus forecast values for AIRU-WRF show a more symmetric clustering around the 45° line, relative to those from RU-WRF. Moreover, unique to AIRU-WRF is its ability to make forecasts at locations where no observations are available. Those are used to produce spatial wind field forecast “maps”, in the form of evolving two-dimensional images for a region of interest. Figure 6 shows examples of those wind field forecast maps at two separate time instances on a select day, on top of the OSW energy lease areas in the NY/NJ Bight. A video showing the evolution of those wind field forecast maps is included in the supplemental materials (SM-2). Those forecast maps can be highly effective in communicating AIRU-WRF’s outputs to key stakeholders in the OSW energy industry.

4.4. Wind power forecasting results

In order to showcase the value of our approach to the OSW energy industry, we convert the wind speed forecasts obtained via AIRU-WRF into wind power predictions. As there are no wind farms currently located in the NY/NJ Bight region, we make use of actual power curves constructed via the method of bins [39, 40] using SCADA data from an operational wind farm in the United States [41]. The power output is then scaled to the $[0, 1]$ interval, where the maximum rated capacity is represented by a value of 1. Using the constructed power curve, we transform the corresponding wind speed forecasts from the six competing models into wind power predictions.

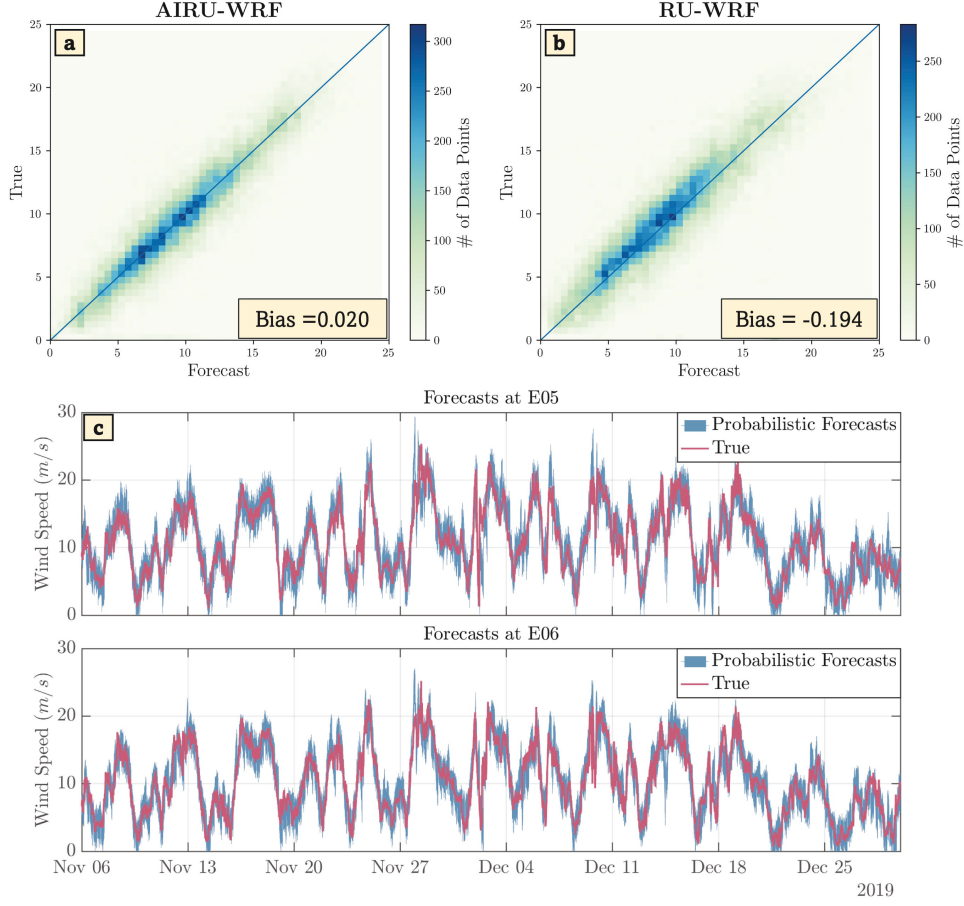


Figure 5: (a)-(b) Comparing AIRU-WRF and RU-WRF: On average, AIRU-WRF’s forecasts are noticeably closer to the true values, yielding a $\sim 90\%$ bias reduction (defined as the average signed difference between data and forecasts). (c) AIRU-WRF’s probabilistic forecasts (10% and 90% percentiles shown), on top of the true observations.

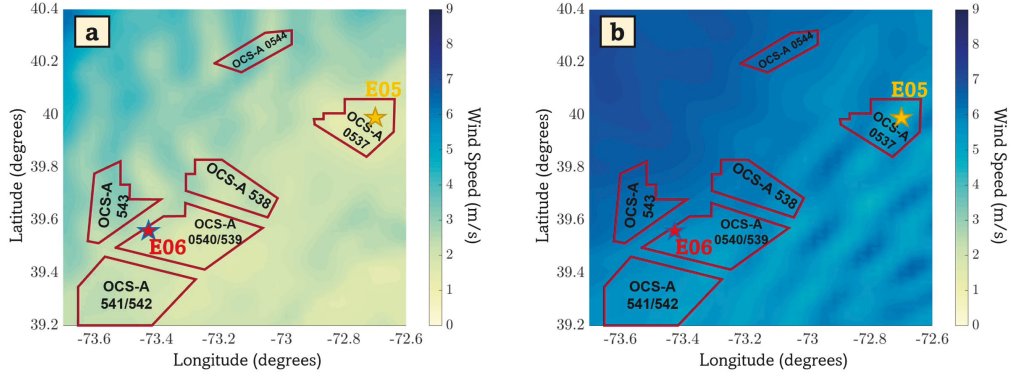


Figure 6: AIRU-WRF’s spatial wind field forecast maps as evolving 2-dimensional images at two separate time instances on the same day. Stars denote E05 and E06 locations, while red polygons roughly depict the OSW energy lease areas, as of August, 2021 [2].

Table 4: Wind power forecasting results for the combined winter and summer periods (total of six months): Average PCE values for sites E05 (Top) and E06 (Bottom).

Bold-faced values denote the best performance.

E05 (39°58'10"N and 72°43'00"W)						
g	AIRU-WRF	GOP	NWP	ARIMAX	LSTM	PER
0.5	0.051	0.061	0.062	0.063	0.071	0.068
0.6	0.052	0.062	0.065	0.064	0.070	0.068
0.7	0.053	0.064	0.069	0.064	0.070	0.069
0.73	0.054	0.064	0.070	0.065	0.070	0.069
0.8	0.054	0.065	0.071	0.065	0.070	0.069
E06 (39°32'50"N and 73°25'45"W)						
g	AIRU-WRF	GOP	NWP	ARIMAX	LSTM	PER
0.5	0.052	0.064	0.065	0.066	0.076	0.071
0.6	0.053	0.065	0.067	0.066	0.076	0.072
0.7	0.054	0.066	0.069	0.066	0.075	0.072
0.73	0.054	0.066	0.070	0.067	0.075	0.072
0.8	0.054	0.067	0.071	0.067	0.075	0.073

The resulting power predictions are evaluated using the PCE loss [37], which assigns unequal weights for under- and over-prediction, as in (14).

$$PCE(P, \hat{P}) = \begin{cases} g \left[P(\mathbf{s}, t_c + h) - \hat{P}(\mathbf{s}, t_c + h) \right] & \text{if } \hat{f}(\mathbf{s}, t_c + h) \leq Y(\mathbf{s}, t_c + h), \\ (1 - g) \left[\hat{P}(\mathbf{s}, t_c + h) - P(\mathbf{s}, t_c + h) \right] & \text{if } \hat{f}(\mathbf{s}, t_c + h) > Y(\mathbf{s}, t_c + h), \end{cases} \quad (14)$$

where $P(\mathbf{s}, t_c + h)$ and $\hat{P}(\mathbf{s}, t_c + h)$ are the normalized power observations and forecasts at $t_c + h$ and the $\mathbf{s}$ th location, and g is the under-estimation weight, which is typically set at values higher than 0.5. Table 4, shows the average PCE values across all horizons for values of g ranging between 0.5 and 0.8 with 0.1 increment, as well as $g = 0.73$ (the value reported in [42, 37]), for sites E05 and E06. AIRU-WRF is shown to significantly outperform all of its competitors, suggesting that the improvements attained in wind speed forecasting are translated into predictive gains in the wind power domain.

5. Conclusions

Accurate short-term wind forecasts are indispensable for the optimal operation of (offshore) wind farms. We have proposed a physics-guided machine-learning-based forecasting model, called AIRU-WRF, which yields significant improvements, in terms of both point and probabilistic evaluations, over a wide array of forecasting benchmarks, thus testifying to the merit of AIRU-WRF to the OSW sector in the US and elsewhere.

This work opens the door for several interesting future research avenues. For instance, meteorological models are often simultaneously run at multiple spatial resolutions (e.g., 3 and 9 km). AIRU-WRF can be extended to integrate the multi-scale physics represented by the nested NWP resolutions through a multi-resolution statistical modeling framework. In addition, extending the forecast horizon (e.g., to a day-ahead), varying the spatio-temporal resolution (e.g. turbine vs. farm-level), and extending AIRU-WRF to directly forecast wind power output, are all topics of ongoing research.

Supplemental Material

SM-1 is a video of the geopotential height over time. SM-2 is a video of AIRU-WRF’s wind field forecast maps. Both videos are appended to this article and are also accessible at [43].

Acknowledgment

This work is supported in part by the U.S. National Science Foundation (ECCS-2114422) and in part by the National Offshore Wind Research & Development Consortium (Project # 192900-133). Funding for the RU-WRF model has been provided by the New Jersey Board of Public Utilities.

References

- [1] NYSERDA, Research and development roadmap version 2.0, Tech. rep., New York State Energy Research and Development Authority (2019).
- [2] BOEM, Lease Information, Bureau of Ocean Energy Management, <https://www.boem.gov/renewable-energy/lease-and-grant-information> (2021).

- [3] G. Lee, Y. Ding, M. G. Genton, L. Xie, Power curve estimation with multivariate environmental factors for inland and offshore wind farms, *Journal of the American Statistical Association* 110 (509) (2015) 56–67.
- [4] P. Nasery, A. A. Ezzat, Yaw-adjusted wind power curve modeling: A local regression approach, *Renewable Energy* 202 (2023) 1368–1376.
- [5] Y. Jiang, T. H. Ortmeyer, Propagation-based network partitioning strategies for parallel power system restoration with variable renewable generation resources, *IEEE Access* 9 (2021) 144965–144975.
- [6] N. Barry, M. Chatzos, W. Chen, D. Han, C. Huang, R. Joseph, M. Klamkin, S. Park, M. Tanneau, P. Van Hentenryck, et al., Risk-aware control and optimization for high-renewable power grids, *arXiv preprint arXiv:2204.00950* (2022).
- [7] P. Papadopoulos, D. W. Coit, A. A. Ezzat, Seizing opportunity: Maintenance optimization in offshore wind farms considering accessibility, production, and crew dispatch, *IEEE Transactions on Sustainable Energy* 13 (1) (2021) 111–121.
- [8] P. Papadopoulos, D. Coit, A. Ezzat, STOCHOS: Stochastic opportunistic maintenance scheduling for offshore wind farms, *IIEE Transactions* (2022) 1–15.
- [9] P. Papadopoulos, F. Fallahi, M. Yildirim, A. A. Ezzat, Joint optimization of maintenance and production in offshore wind farms: Balancing the short-and long-term needs of wind energy operation, *arXiv preprint arXiv:2303.06174* (2023).
- [10] C. Sweeney, R. J. Bessa, J. Browell, P. Pinson, The future of forecasting for renewable energy, *Wiley Interdisciplinary Reviews: Energy and Environment* 9 (2) (2020) e365.
- [11] M. Optis, A. Kumler, G. N. Scott, M. C. Debnath, P. J. Moriarty, Validation of RU-WRF, the custom atmospheric mesoscale model of the Rutgers Center for Ocean Observing Leadership, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2020).

- [12] A. A. Ezzat, M. Jun, Y. Ding, Spatio-temporal short-term wind forecast: A calibrated regime-switching method, *The Annals of Applied Statistics* 13 (3) (2019) 1484 – 1510.
- [13] N. Chen, Z. Qian, I. T. Nabney, X. Meng, Wind power forecasts using gaussian processes and numerical weather prediction, *IEEE Transactions on Power Systems* 29 (2) (2013) 656–665.
- [14] L. Dong, L. Wang, S. F. Khahro, S. Gao, X. Liao, Wind power day-ahead prediction with cluster analysis of NWP, *Renewable and Sustainable Energy Reviews* 60 (2016) 1206–1212.
- [15] S. Hu, Y. Xiang, H. Zhang, S. Xie, J. Li, C. Gu, W. Sun, J. Liu, Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction, *Applied Energy* 293 (2021) 116951.
- [16] NYSERDA, Nyserda floating lidar buoy data, <https://oswbuoysny.resourcepanorama.dnvgl.com/> (2019).
- [17] J. Dicopoulos, J. F. Brodie, S. Glenn, J. Kohut, T. Miles, G. Seroka, R. Dunk, E. Fredj, 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.
- [18] RUCOOL, 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 (2019).
- [19] M. Optis, A. Kumler, J. Brodie, T. Miles, Quantifying sensitivity in numerical weather prediction-modeled offshore wind speeds through an ensemble modeling approach, *Wind Energy* 24 (9) (2021) 957–973.
- [20] J. B. Olson, T. Smirnova, J. S. Kenyon, D. D. Turner, J. M. Brown, W. Zheng, B. W. Green, A description of the mynn surface-layer scheme (2021).
- [21] S. C. Murphy, L. J. Nazzaro, J. Simkins, M. J. Oliver, J. Kohut, M. Crowley, T. N. Miles, Persistent upwelling in the mid-atlantic bight

detected using gap-filled, high-resolution satellite sst, *Remote Sensing of Environment* 262 (2021) 112487.

- [22] X. Zhu, K. P. Bowman, M. G. Genton, Incorporating geostrophic wind information for improved space–time short-term wind speed forecasting, *The Annals of Applied Statistics* 8 (3) (2014) 1782 – 1799.
- [23] C. Feng, M. Cui, B.-M. Hodge, J. Zhang, A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, *Applied Energy* 190 (2017) 1245–1257.
- [24] N. Cressie, C. K. Wikle, *Statistics for Spatio-Temporal Data*, John Wiley & Sons, 2015.
- [25] A. A. Ezzat, Turbine-specific short-term wind speed forecasting considering within-farm wind field dependencies and fluctuations, *Applied Energy* 269 (2020) 115034.
- [26] C. Rasmussen, C. Williams, *Gaussian Processes for Machine Learning*, Cambridge: MIT Press, 2006.
- [27] M. L. Stein, Space–time covariance functions, *Journal of the American Statistical Association* 100 (469) (2005) 310–321.
- [28] A. A. Ezzat, M. Jun, Y. Ding, Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting, *IEEE Transactions on Sustainable Energy* 9 (3) (2018) 1437–1447.
- [29] D. R. Cox, V. Isham, A simple spatial-temporal model of rainfall, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 415 (1849) (1988) 317–328.
- [30] M. L. O. Salvaña, A. Lenzi, M. G. Genton, Spatio-temporal cross-covariance functions under the Lagrangian framework with multiple advections, *Journal of the American Statistical Association* (2022) 1–16.
- [31] M. Schlather, Some covariance models based on normal scale mixtures, *Bernoulli* 16 (3) (2010) 780 – 797.
- [32] M. L. O. Salvana, M. G. Genton, Nonstationary cross-covariance functions for multivariate spatio-temporal random fields, *Spatial Statistics* 37 (2020) 100411.

- [33] M. Arrieta-Prieto, K. R. Schell, Spatio-temporal probabilistic forecasting of wind power for multiple farms: A copula-based hybrid model, *International Journal of Forecasting* 38 (1) (2022) 300–320.
- [34] P. Pinson, Wind energy: Forecasting challenges for its operational management, *Statistical Science* 28 (4) (2013) 564 – 585.
- [35] Y. Gel, A. E. Raftery, T. Gneiting, Calibrated probabilistic mesoscale weather field forecasting, *Journal of the American Statistical Association* 99 (467) (2004) 575–583.
- [36] M.-S. Ko, K. Lee, J.-K. Kim, C. W. Hong, Z. Y. Dong, K. Hur, Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting, *IEEE Transactions on Sustainable Energy* 12 (2) (2020) 1321–1335.
- [37] A. S. Hering, M. G. Genton, Powering up with space-time wind forecasting, *Journal of the American Statistical Association* 105 (489) (2010) 92–104.
- [38] F. Ye, J. Brodie, T. Miles, A. A. Ezzat, Ultra-short-term probabilistic wind forecasting: Can numerical weather predictions help?, Accepted, 2023 IEEE PES General Meeting (2023).
- [39] Wind Energy Generation Systems - Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, IEC 61400-12-1 International Electrotechnical Commission (2017).
- [40] B. Golparvar, P. Papadopoulos, A. A. Ezzat, R.-Q. Wang, A surrogate-model-based approach for estimating the first and second-order moments of offshore wind power, *Applied Energy* 299 (2021) 117286.
- [41] Y. Ding, *Data Science for Wind Energy*, CRC Press, 2019.
- [42] P. Pinson, C. Chevallier, G. N. Kariniotakis, Trading wind generation from short-term probabilistic forecasts of wind power, *IEEE Transactions on Power Systems* 22 (3) (2007) 1148–1156.
- [43] F. Ye, A. Ezzat, Research Multimedia: [M2] AIRU-WRF: A physics-guided spatio-temporal wind forecasting model and its application to the U.S. Mid-Atlantic offshore wind energy areas, *Last Accessed: August*,

2023.

URL <https://sites.rutgers.edu/azizezzat/research-videos-multimedia/>