Methods for Pacific OCS Wind Characterization for Offshore Wind Development

by

Macy Frost Chang

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science

in

Engineering - Civil & Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Fotini Katopodes Chow, Chair
Professor Scott Moura
Professor Bruce Buffett

Spring 2022

Methods for Pacific OCS Wind Characterization for Offshore Wind Development

Copyright 2022
by
Macy Frost Chang

Abstract

Methods for Pacific OCS Wind Characterization for Offshore Wind Development

by

Macy Frost Chang

Master of Science in Engineering - Civil & Environmental Engineering

University of California, Berkeley

Professor Fotini Katopodes Chow, Chair

The rapid development of the U.S. offshore wind industry has necessitated accurate assessment and prediction of offshore wind profiles to forecast generated power. Two areas on the Pacific Outer Continental Shelf (OCS), the Humboldt and Morro Bay Wind Energy Areas (WEAs), are currently under review for potential floating wind turbine farms. Scientific buoys sponsored by the Department of Energy and deployed by the Pacific Northwest National Laboratory in these two WEAs have made observational hub-height wind time series available for the first time, opening novel opportunities for model validation and intercomparison. Predictive models are a necessary tool for Pacific offshore wind development, as collection of sufficient observational data is rendered prohibitive by the high capital costs and maintenance requirements of the scientific buoys. Predictions of wind speeds 100 meters above mean sea level are compared between a conventional physical law approach, known as the stability-corrected logarithmic law (S-C Log Law), and three machine learning (ML) approaches, specifically the methods of random forest (RF), Gaussian process regression (GPR), and Long Short-Term Memory neural network (LSTM). To simulate realistic industry applications, predictor variables for the ML approaches are constrained to surface-level measurements, and the ML algorithms are trained and tested at different locations. The S-C Log Law and LSTM produce the most accurate predictions, with RMSE of 1.33 m/s and 1.38 m/s respectively, even with LSTM's large train-test extrapolation distance between the two sites (631 km). Error metrics for all three ML methods generally improve when a longer training time is implemented and when the algorithms are trained and tested in the same location, with LSTM surpassing S-C Log Law and achieving RMSE as low as 0.68 m/s. It is concluded that ML techniques, notably LSTM and GPR, can present equivalent or improved offshore wind speed prediction capabilities over the S-C Log Law, a widely-accepted physical law extrapolation method. These ML techniques are advantageous for wind energy purposes, as they can be used for predicting other wind parameters and generating short-term forecasts. Increased future availability of Pacific OCS turbine-height wind data would be vital for determining ML performance degradation over varying extrapolation distances.

To my mom, Angie Frost

Contents

List of Figures

List of Tables

List of Acronyms

A2e	Atmosphere to Electrons
ABL	atmospheric boundary layer
AMSL	above mean sea level
ANN	artificial neural network
ARIMA	autoregressive integrated moving average
BOEM	Bureau of Ocean Energy Management
BPTT	back-propagation through time
CI	confidence interval
COARE	Coupled Ocean–Atmosphere Response Experiment
cRMSE	centered root-mean-square error
DOE	Department of Energy
DOI	Department of the Interior
EMD	Earth-mover's distance
ERA5	European Centre for Medium-Range Weather Forecasts Reanalysis v5
GPR	Gaussian process regression
HUM	Humboldt [Wind Energy Area]
LiDAR	Light Detection and Ranging
LSTM	Long Short-Term Memory [neural network]
MAE	mean absolute error
MSE	mean squared error
ML	machine learning
MOR	Morro Bay [Wind Energy Area]
MOST	Monin-Obukhov similarity theory
MSE	mean squared error
MVN	multivariate normal distribution
NDBC	National Data Buoy Center
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
NWP	numerical weather prediction
OCS	Outer Continental Shelf
OOBE	Out-of-bag error
OWA	Offshore Wind Accelerator

PCA	principal component analysis
PDF	probability density function
PNNL	Pacific Northwest National Laboratory
R^2	coefficient of determination
RAP	Rapid Refresh
RBF	radial basis function
RF	random forest
RI	Richardson number
RI_B	bulk Richardson number
RMSE	root-mean-square error
RNN	recurrent neural network
S-C Log Law	stability-corrected logarithmic law
SST	sea surface temperature
SVM	Support Vector Machine
SVR	Support Vector Regression
TKE	turbulent kinetic energy
VI	variable importance
WEA	Wind Energy Area
WIND	Wind Integration National Dataset
WRF	Weather Research and Forecasting

Acknowledgments

I would like to express deep gratitude to my research supervisor Professor Tina K. Chow for providing invaluable guidance and time toward this project, especially amidst the chaos presented by the COVID-19 pandemic. Tina's motivation and expertise is a huge inspiration for my academic goals, and her sincerity and thoughtfulness is greatly influential to me as a person. I would also like to thank Professors Scott Moura and Dan Arnold for the encouraging and insightful conversations that made the journey into the land of machine learning much less daunting, and Professor Bruce Buffett for lending me his generous time and patience while digging into the theory with me.

I would additionally like to thank Raghavendra Krishnamurthy and the other Pacific Northwest National Laboratory scientists for allowing me to use their buoy data and providing clarification on some vital details. I also extend gratitude toward my fellow graduate student cohort, especially Adam Wise for providing great advice toward this project. I would lastly like to thank my wonderful cooperative community in Berkeley for the countless conversations and unrelenting support through the many challenges of this year.

Chapter 1

Introduction

1.1 U.S. Wind Energy

Wind energy is a clean energy technology that is continuing to rapidly expand and evolve across the globe. It is considered one of the most sustainable sources of energy due to its domestic availability, low operational carbon emissions, and near-zero usage of water resources [1]. These advantages, along with large annual industry investments [2] and its cost-competitiveness with conventional fossil fuel energy generation [3], have allowed wind energy capacity to grow 15% annually over the past 10 years, making it the current largest source of renewable energy in the United States [4]. With record installations of wind turbines worldwide in 2020, global wind power capacity now stands at 743 GW, supplying over 6% of global electricity demand [5].

Technological advances and stakeholder interest are allowing new areas to become feasible for wind farm construction. Bottom-mounted and floating wind turbine structures allow for construction and operation of wind farms in offshore locations, where wind speeds tend to be higher and more consistent than over land [6]. The United States Department of Energy (DOE) is currently in the process of expanding potential wind energy areas into the offshore arena. Two offshore wind sites on the Atlantic coast are already in operation and demonstrate the economic and energetic feasibility of offshore wind energy siting [7], [8].

1.2 Pacific OCS Offshore Wind: Opportunities and Challenges

The Pacific Outer Continental Shelf (OCS), the area of ocean extending to about 200 nautical miles from the U.S. West Coast, is an area that has recently come into consideration for offshore wind development. In May of 2021, the Biden administration and California

governor Gavin Newsom announced plans to develop floating offshore wind farms in two Pacific OCS locations off the coast of Morro Bay and Humboldt County, which together have the potential to provide up to 4.6 GW of electricity, equivalent to powering 1.6 million homes [9]. The Bureau of Ocean Energy Management (BOEM) has recently titled these two locations as "Wind Energy Areas" (WEAs) after collaborating with federal agencies, the state of California, local tribes and communities, and the public to consider potential impacts on the environment as well as ocean resources and commerce.

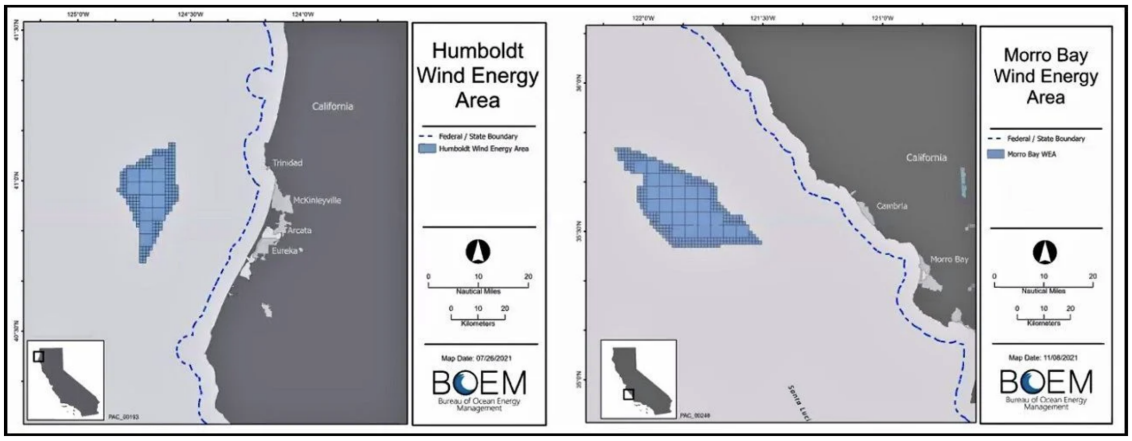

Figure 1.1: The geographic areas of the Humboldt and Morro Bay Wind Energy Areas (WEAs). These areas are currently undergoing Environmental Assessment by the U.S. Bureau of Ocean Energy Management. (Images from BOEM.gov [10])

The development of offshore wind farms requires extensive research of environmental factors that affect wind farm layout, predicted power output, and fatigue load estimations. Until recently, only a few surface-monitoring buoys scattered along the Pacific OCS have provided direct observational data of offshore meteorological and oceanic characteristics. These buoys measure only surface variables and do not provide vertical profiles of winds and turbulence in the surface atmospheric boundary layer (ABL), which is critical information for wind energy developers.

As of September 2020, a research campaign funded by the DOE has made direct, consistent observations of above-surface wind characteristics in the Pacific OCS available for the first time. The Pacific campaign consists of two specialized research buoys stationed in the Humboldt and Morro Bay WEAs. These buoys measure wind speed and direction up to 250 meters above mean sea level (AMSL) as well as a large set of surface variables.

The specifics of the DOE campaign and buoy instrumentation are detailed further in the Background section.

1.3 Motivation

Data from the DOE buoy campaign provide many novel opportunities for offshore wind research on the U.S. West Coast. This data can be used to assess the wind resource at potential turbine heights at these Pacific OCS sites. The data also introduce new possibilities for validating wind prediction models in the Pacific OCS region. More specifically, the observational dataset is vital for identifying model biases and optimizing the performance of statistical, physical, and intelligent learning wind prediction methods [11].

New strategies for wind speed extrapolation from surface measurements could prove to be more adaptable and site-specific than complex and computationally-expensive numerical weather prediction (NWP) models. Additionally, improved wind modeling accuracy and new intelligent modeling approaches may help to enhance confidence in turbine and wind plant power production estimates, which would have significant impacts on downstream analyses, including grid integration and expansion [12]–[14] as well as life-cycle cost analyses of floating offshore wind energy production [15]. Considering the recency of the Pacific OCS wind data and the current Environmental Assessment status of the WEAs in the region, an imminent research need is to test out different wind prediction models and assess which techniques provide the best estimates for energy developers.

A large suite of wind prediction techniques are in widespread use by the wind industry and could potentially be tested in these offshore areas. One category of techniques is physical modeling, wherein meteorological variables are input into physical law equations to calculate other weather characteristics. Another contemporary group of techniques are intelligent learning methods, which most often use machine learning (ML) architecture to draw complex statistical relationships between input and output data and subsequently make predictions. In this study, prediction methods from both of these categories are tested by comparing their accuracy and adaptability across the Humboldt and Morro Bay WEAs. Model performance is analyzed to provide novel insights into which technique(s) produce the most accurate wind predictions and offer the most adaptivity for other Pacific OCS wind prediction applications.

1.4 Research Objective

The first aim of the research is to analyze relevant hub-height (100 meters above mean sea level) wind characteristics in the Humboldt and Morro Bay WEAs, then to elucidate relationships between these hub-height winds and surface variables. These relationships will

be drawn from analyzing correlations of 100-meter winds to surface wind speed, air and sea temperature, atmospheric pressure, and temporal characteristics.

The second aim of the research is to compare vertical wind extrapolation performance of a widely-used physical law approach, the stability-corrected logarithmic law (denoted as S-C Log Law), versus the machine learning (ML) methods of random forest, Gaussian process regression, and Long Short-Term Memory neural network (denoted respectively as RF, GPR, and LSTM). Input variable importance will be assessed by the weighted effect of each input feature on the prediction. ML model testing will occur in a "round-robin" fashion (training on one buoy's data and testing on the other) to minimize training location bias and assess the adaptability of each ML algorithm to other Pacific OCS offshore locations. This style of testing will help gauge the greater applicability of these ML methods to other Pacific offshore areas where only surface-level data is available.

Chapter 2

Background

2.1 Environmental Logistics

Offshore wind demonstrates many advantages over onshore wind energy, particularly with respect to wind power potential and reliability. One of the most notable differences is spatial availability, as offshore locations offer vast, open areas to develop large-scale wind farms. Additionally, wind speeds in offshore locations are generally higher than those in neighboring coastal locations due to lower surface roughness and the absence of topographic obstructions [16], [17]. The topographic uniformity of offshore locations also contributes to typically lower turbulence and wind shear, which are important considerations for wind turbine energy efficiency and fatigue load [18]. In some areas, the greater wind power availability over the ocean allows for more than triple the energy generation potential of typical land-based wind power [19]. Offshore winds in general are known to be especially strong in evening hours, when solar power production sharply decreases, which helps to balance renewable energy contributions to the grid [20].

The offshore arena also presents novel challenges that are not present in land-based energy operations. Offshore wind turbines are more expensive to build, install, and maintain than their onshore counterparts due to the remote location siting and the deteriorative effects of waves and seaspray [21], [22]. Transmission lines are also more difficult to install and must traverse long distances across the oceanic benthic zone [23]. These factors cause offshore wind projects to be much more capital-intensive in both construction and maintenance, which has proven prohibitive to U.S. offshore wind development prior to the last few years. As of 2021, less than 5% of global wind power capacity is supplied from offshore wind turbines [5].

2.2 U.S. Offshore Wind Development Status

Interest driven by private industry and efforts of the U.S. Bureau of Ocean Energy Management (BOEM) are helping to meet the high capital costs associated with offshore wind development. These offshore wind planning operations are thus gaining traction; the current Biden-Harris administration released plans to install 30 GW of offshore wind energy by 2030 [24] and BOEM is reviewing ten Construction and Operations Plans for offshore wind projects in U.S. waters, with at least five more reviews currently being considered [25]. The Department of the Interior (DOI) is currently holding lease sales for more sites along the East Coast and planning to hold future lease sales for Pacific and Gulf Coast sites [26].

Offshore wind development is primarily taking place off of the Atlantic coast because of the relatively shallow and accessible seabed of the Atlantic OCS. In the areas that are less than 25 meters in depth, bottom-mounted turbine substructures can be installed, which comprise most of the offshore wind structures installed to date worldwide [27]. Large portions of the Atlantic OCS maintain depths above 60 meters, where it is still possible to install monopiles and four-legged jackets to house bottom-mounted turbines [28]. In contrast, the Pacific continental shelf declines steeply, exceeding depths of 600 meters in the potential wind energy sites [29]. The depths of the Pacific OCS render conventional bottom-mounted turbine installation impossible in most places and have hindered any offshore wind development on the U.S. West Coast to date.

New developments in offshore wind turbine technology are opening the doors to Pacific OCS wind farms. The massive wind resource potential in this area can be accessed by installing floating offshore wind turbines [30], in which a turbine is mounted on a floating substructure connected to the seabed with mooring cables. Many design iterations of these floating structures exist and have been deployed by private industries in other countries, adventing in 2017 with the Hywind Scotland commercial floating wind farm [31]. Fueling this industry advancement farther out to sea is the enormous space and energy potential presented by deeper maritime waters, which have the potential to more than double the gross U.S. offshore wind resource [32]. These floating turbine designs can access oceanic areas of up to 1000 meters deep [33], where winds are typically stronger and more consistent than adjacent near-shore locations [34].

2.3 Humboldt & Morro Bay Wind Energy Areas

The Humboldt Wind Energy Area (WEA) is located 20-30 miles offshore of Humboldt County and covers an area of 207 square miles, with depths ranging from 600-1000 meters. Farther south, the Morro Bay WEA, 22 miles offshore between Monterey Bay and Morro Bay, spans nearly 400 square miles and sports depths of 800-1000 meters [20], [35].

Both locations have average wind speeds of 8-10 meters per second at turbine hub height elevations, which are projected to be around 90-120 meters above mean sea level, and relatively steady seasonal and diurnal wind patterns. These are highly favorable conditions for maximizing wind energy efficiency and turbine longevity [36]–[38]. These areas were selected not only based on their wind energy potential, but also a multitude of logistical considerations, including compatibility with Department of Defense training and testing operations, compliance with federal, state, and local environmental permitting agencies, and local electrical load and transmission capabilities [12], [37], [39].

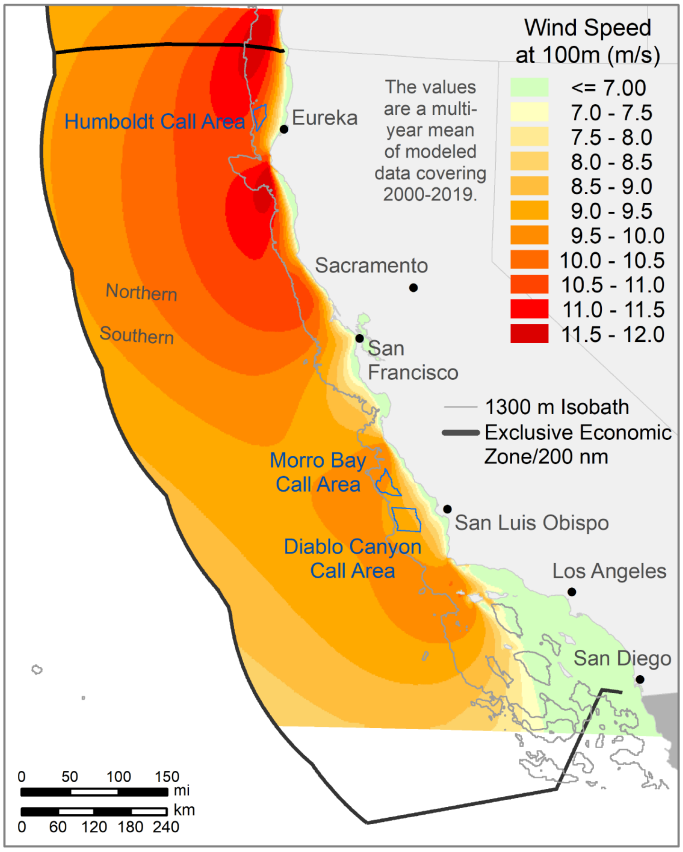

Figure 2.1: Estimates of average 100-meter wind speeds across the Pacific OCS. Data is sourced from the CA20 dataset, generated from the Weather Research & Forecasting (WRF) numerical weather prediction model. (Figure from Optis et al., 2020 [40])

2.4 DOE LiDAR Buoy Program

Vertical profiles of offshore wind are able to be measured through the use of Light Detection and Ranging (LiDAR) instruments, which can interpret wind speed and direction at multiple heights by emitting and receiving laser light pulses. In 2014, the U.S. Department of Energy's Wind Energy Technologies Office launched their LiDAR Buoy Program, consisting of two AXYS WindSentinel buoys (Buoys #120 and #130) capable of measuring large, comprehensive sets of meteorological and oceanographic data at their anchored location [41]. The data is made publicly available on a daily basis to support offshore wind energy development in the United States. The two buoys were deployed in U.S. East Coast WEAs from 2014 to 2017 and have undergone various instrument upgrades and LiDAR validation campaigns [42].

In September 2020, the two DOE LiDAR buoys were deployed in the approximate centroids of the Humboldt and Morro Bay WEAs under management of the Pacific Northwest National Laboratory (PNNL). The LiDAR buoys provide 10-minute average and select 1-second meteorological and oceanographic data, including longitudinal wind speed and direction from 40 to 250 meters AMSL at 20 meter vertical spatial resolution [43]. The Morro Bay deployment ended in October 2021, while the Humboldt deployment is currently continuing until May 2022 [42].

The LiDAR-scanning instrument on the DOE buoys is the Leosphere WindCube 866 v2, a state-of-the-art pulsed-wave Doppler-LiDAR technology capable of withstanding harsh oceanic conditions and retrieving robust, high-resolution wind field data from its buoy-mounted location. It operates by emitting an infrared laser beam at an inclination angle of 28° in the four principal directions and one beam in the vertical direction, then deriving a three-dimensional wind vector from these measurements [44]. Each five-beam sample is completed in approximately 5 seconds [1]. Previous to deployment off of California, both buoys were independently validated against a stationary, industry-standard LiDAR at Martha's Vineyard Coastal Observatory, achieving commercial acceptance through Carbon Trust's Offshore Wind Accelerator (OWA) program [45]. The data gathered by these LiDAR buoys provide the first high-temporal resolution observational time series data of wind across rotor-swept heights in the Pacific OCS.

Chapter 3

Literature Review

3.1 Fidelity of LiDAR Wind Measurements

Light Detection and Ranging (LiDAR) works based on the "coherent detection principle" of the Doppler-shifted frequency of backscattered radiation from a laser light beam. In other words, a LiDAR instrument sends out a laser light pulse, from which the amount of backscattered light returns information about atmospheric components such as cloud density and aerosol concentrations [46]. Detection of wind speed requires a precise measurement of the small difference in transmitted and backscattered light frequency, measured along the laser line-of-sight beam direction [47]. Doppler LiDAR systems are becoming increasingly popular due to improved compactness, reliability, and accessibility of LiDAR instruments [48]. This remote sensing technique is now widely deployed in commercial and wind research applications to acquire meteorological data from the near-surface atmospheric boundary layer (ABL), such as wind speed and direction, shear and low-level jets, and turbulence characteristics [38], [49]–[51].

LiDAR for meteorological applications began in the early 1960s and has since been in continuous use for aviation weather forecasting [52], [53], aerosol detection [54], and global atmospheric dynamics characterization for numerical weather prediction (NWP) model initialization [55]. For wind energy applications, coherent Doppler LiDAR is particularly useful because of its powerful, high-resolution scanning capabilities over large areas of the surface ABL. The favorable spatial resolution and depth is complemented by the short intervals between scans, meaning that velocity fields can be obtained at a temporal resolution as fine as 1 Hz [56]. The size and ease-of-use of commercial LiDAR instruments also make this technology a more adaptable, cost-effective, and sometimes necessary substitute for meteorological masts [47]. This is especially true in the offshore environment, considering the unique challenges of installing meteorological masts in these locations.

The fidelity of wind LiDAR measurements has been examined by comparison with anemome-

ters on nearby meteorological masts in land-based sites [57], [58] and in offshore areas. Kindler et al. (2007) compared 8 months of Doppler-LiDAR measurements up to 100 meters against an adjacent meteorological mast at the FINO-1 offshore platform in the North Sea [59]. They reported a correlation coefficient of 0.99 and average regression slope of 0.97 for the 10-minute wind speeds. Similarly, Shu et al. (2016) compared measurements from a platform-mounted Doppler-LiDAR in an offshore site in Hong Kong against an offshore-mounted meteorological mast with measurements taken up to 40.7 meters [18]. The researchers found, between the LiDAR and mast measurements, that wind direction measurements were in good agreement, 10-minute average wind speed measurements were near-identical, and correlation coefficients were greater than 0.99. Peña et al. (2009) performed a similar analysis with the continuous-wave ZephIR wind LiDAR at the Horns Rev offshore wind farm in Denmark, finding excellent agreement between 10-minute wind speeds up to 161 meters [60]. These studies lend confidence to the offshore wind-profiling capabilities of fixed-platform LiDAR.

3.2 Offshore Wind Observation Methods

3.2.1 Fixed-Platform Measurements

The lack of available observational data is an ongoing research dilemma faced in the Pacific OCS offshore environment. Traditionally, wind measurements in a wind energy research context are measured by anemometers attached to meteorological masts, which are free-standing towers equipped with meteorological instruments to measure wind speed, air temperature, and other environmental parameters. In recent years, LiDAR technology has increasingly come into use often as a substitute or cross-reference for a meteorological mast.

Most of the studies that have conducted analyses of above-surface offshore wind characteristics have been limited to areas with a preexisting oceanic platform for mounting a meteorological mast or LiDAR. Ortiz-Suslow et al. (2019) used ultrasonic anemometer data from the Floating Instrument Platform off of Southern California to take measurements of atmospheric turbulence [61]. The researchers evaluated turbulence dynamics against Kolmogorov's inertial subrange power law and found variance from the theoretical values based on mean environmental state and air-sea dynamical regime. Bodini et al. (2019) used fixed-platform LiDAR data from the Woods Hole Oceanographic Institution's Air-Sea Interaction Tower to derive offshore atmospheric turbulence dissipation rates up to 200 meters [1]. Smaller turbulence dissipation rates than typical onshore values were found, which are conducive to more persistent turbine wakes. Shu et al. (2016) used meteorological mast and LiDAR data from an offshore platform in Hong Kong to determine offshore wind speed, shear, and turbulence intensity profiles [18]. They found that due to thermal instability, turbulence intensity decreased with increasing wind speed below a certain value, which are

helpful metrics for evaluating offshore turbine efficiency and fatigue load. Additional studies with similar methods to the above can be found in [62]–[64].

The data sourcing methods from the above studies are unlikely to be applicable to Pacific OCS wind energy domains. The rough oceanic conditions and depth of the Pacific OCS render the cost of oceanic platform or bottom-mounted meteorological mast installment prohibitively high [65]. It is also important to consider the specific needs of wind energy research. Observational data needs to be accurate and have high spatial resolution up to the uppermost turbine heights to characterize the wind across the entire rotor-swept area. Offshore turbines are expected to be larger than onshore ones; rotor diameters are expected to reach heights up to 175 meters for 8 MW designs and 240 meters for 15 MW designs [66] compared to a typical height of 90 meters for 3 MW onshore turbines [2]. Meteorological masts are typically shorter than these rotor-swept heights and only measure wind at a few discrete points of the vertical profile, which is not useful for gathering detailed measurements of the atmospheric wind field surrounding a large turbine [47]. Floating LiDAR buoys provide a solution to the practical shortcomings of meteorological masts and the infrastructural challenges presented by the extensive depths of the Pacific OCS seabed.

3.2.2 Floating LiDAR Instruments

Studies on floating LiDAR technology have primarily focused on its correctional capabilities for the translational and rotational motions of the buoy. Bischoff et al. (2015) analyzed the effects of buoy motion between a buoy-mounted ZephIR 300 LiDAR instrument and a stationary reference [67]. They found that continuous wave LiDAR systems can accurately correct pitch and roll motions to obtain precise 10-minute average horizontal wind speed measurements. Wolken-Möhlmann et al. (2010) studied the influence of motion on both pulsed and continuous-wave LiDAR [68]. The influence of tilting was found to introduce the most error into wind speed readings, but all errors were able to be successfully corrected, especially by accounting for the change of beam direction. Désert et al. (2021) investigated the mechanisms responsible for turbulence intensity measurement error of a buoy-mounted WindCube v2 and successfully applied a correction using a model-based method [69]. Tiana-Alsina et al. (2015) employed the use of a cardanic (motion-compensating) frame for a continuous-wave LiDAR buoy [70]. They found that 10-minute average wind speed measurements had excellent correlation to fixed-platform LiDAR data both with and without the frame, except in cases with large angular amplitudes and short periods. Similarly to Bischoff et al. (2015), they noted the degradation of 1-second data without motion compensation, but found significant improvement in this data with use of the cardanic frame.

The literature reviewed supports the fidelity of floating LiDAR data for near-surface ABL offshore wind characterization. Both continuous-wave and pulsed-wave LiDAR instruments, including the Leosphere WindCube v2, are in current compliance with standardized require-

ments given for floating LiDAR systems, as delineated in the Carbon Trust OWA Roadmap [45], [67]. An important finding from the literature reviewed is the reduced quality of unprocessed 1-second buoy-mounted LiDAR data, primarily due to buoy motions of tilting and heaving. The literature indicates that these errors induced by buoy motion are correctable and do not impact the accuracy of 10-minute average data.

Buoy-mounted LiDAR data has only recently come into use for offshore wind characterization and comparison with NWP offshore wind model estimates. Shaw et al. (2020) provided a climatological analysis of surface and hub-height winds as well as an analysis of oceanographic observations based on the data from the DOE buoy deployments in New Jersey and Virginia [71]. Debnath et al. (2021) used the DOE floating LiDAR data from the BOEM lease areas off of New Jersey to characterize occurrences of high wind shear and low-level jet events [38]. They found over 100 of these events throughout the year, exhibiting strong seasonal variability and correlating to stable atmospheric stratification and a positive air-sea temperature difference. Sheridan et al. (2021) used data from near-shore New Jersey and Virginia deployments of the DOE LiDAR buoys to compare the performances of four commonly-used reanalysis models [72]. All four datasets underestimated the wind speed at 90 meters above mean sea level and demonstrated increased error in conditions of stable stratification, high wind shear, and tropical and winter storms. Optis et al. (2020) conducted a large validation study of 16 Weather Research and Forecasting (WRF) model setups to the New Jersey LiDAR buoy data in order to choose the best model inputs, parameterizations, grid specifications, and physics schemes for the CA20 offshore wind dataset [40]. After choosing the appropriate model scheme, they found that WRF-modeled winds at 100-meter heights in the Pacific offshore WEAs were significantly higher than predicted by the Wind Integration National Dataset (WIND) Toolkit, implicating new model-based economic and energy production scenarios for the Pacific OCS wind resource.

As offshore wind analyses which utilize floating LiDAR data are limited to only the past few years, it is clear that there is much more research potential to be achieved using buoy-mounted LiDAR technology. The majority of offshore wind studies reviewed derive observational data of hub-height offshore wind characteristics from an oceanic platform-mounted LiDAR or mast, or extrapolate this data from surface buoys or scatterometers. In the former case, spatial availability is extremely limited, and new platform structures are extremely challenging to construct in U.S. West Coast offshore locations. In the latter case, conflicts in the research arise as to whether extrapolated surface data is meaningful enough to serve as reliable data without other sources of validation. Many vertical extrapolation methods have nontrivial underlying assumptions that are not well-tested in offshore locations, which can subsequently produce inaccurate wind estimates [66]. Based on these considerations, floating LiDARs appear to be the most promising and robust data source for validating hub-height wind speeds in remote, offshore locations.

Figure 3.1: One of the two AXYS WindSentinel LiDAR buoys operated by the Pacific Northwest National Laboratory (PNNL) for DOE. The LiDAR instrument is the small white cube atop the platform on the large mast. (Image from BOEM.gov)

3.3 Wind Prediction Methods: Physical Law Approaches

Direct in-situ measurements by anemometers or LiDAR instruments are best for precise wind measurements, but these limited measurements are often insufficient for wind energy purposes. The reasoning behind this is simple: for energy developers, wind data must often be extrapolated in space, such as to a neighboring wind farm or development site, or in time, such as a day-ahead forecast. Scalable wind power assessment therefore relies on the accurate estimation of wind speed at rotor-swept heights from predictive models. This and the following section will discuss the key considerations involving two categories of wind prediction approaches: physical law calculations and intelligent learning models.

3.3.1 Numerical Weather Prediction

Currently, most of the information used to characterize the wind resource on the Pacific OCS is derived from NWP model estimates. NWP uses mathematical models of the atmosphere and ocean to simulate weather conditions based on past and present physical weather in-

puts [73]. In the offshore Pacific region, data is most commonly sourced from WRF NWP datasets such as the WIND Toolkit and more recent CA20 dataset [40]. Information on offshore wind is also often obtained from reanalysis products such as European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) [74] and Rapid Refresh (RAP) from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction [75].

A prerequisite for accurate, high-resolution modeling of the Pacific OCS wind resource is validation of the model with direct observations. One major issue with using NWP datasets for the Humboldt and Morro Bay WEAs is the lack of validation with observed vertical wind profiles from these Pacific OCS areas of interest. The WIND Toolkit, despite serving as the basis for many past wind resource assessments in this area [76], has been minimally validated in offshore locations [66]. These sparse data used for WIND Toolkit model validation in U.S. offshore locations are almost entirely sourced from surface-level buoy-based measurements from the NOAA National Data Buoy Center (NDBC) [77]. The more recently released CA20 dataset, despite increased data comparison with buoys and coastal radars, still lacks validation against high-quality Pacific OCS wind observations at the expected rotor-swept heights of offshore wind turbines [40].

Even with the available observational data, validation of NWP models against sea surface-level and land-based observations is not guaranteed to be representative of offshore hub-height wind modeling accuracy. In other words, model validation at buoy heights (generally 10 meters or below) is not necessarily appropriate for offshore wind energy applications of the model, nor is model validation at coastal or other onshore wind observation sites. Researchers at the National Renewable Energy Laboratory (NREL) compared WRF simulation data against coastal, sea surface-elevation, and offshore hub-height level observations in two Atlantic offshore WEAs [66]. The study found that diurnal wind speed cycles at offshore hub-heights are substantially different from sea surface-level measurements below, but the WRF simulations do not capture the magnitude of these differences. Likewise, despite how land-based weather stations generally have measurement heights above those of buoys, the combination of large land-sea wind speed gradients and interpolation of WRF wind speeds to observation stations presents too much uncertainty to provide meaningful estimates of rotor-height offshore wind speeds.

NWP models, such as WRF, are the most complex of the physical approaches, as they calculate numerous weather factors across many spatiotemporal intervals based on historical data and intricate physical relationships. They are extremely useful for assessing weather and climate data at mesoscale resolutions and larger. However, in the interest of fine-scale wind estimates, running these models is relatively computationally expensive and prone to inaccuracies based on certain meteorological and oceanic conditions as well as interpolation errors [72], [78]. Simplified physical law approaches that extrapolate wind data in one vertical column are therefore more practical for validation against stationary LiDAR observations.

3.3.2 Logarithmic Profile

The most common, simplified forms of a physical model to perform vertical wind speed extrapolation are the power law [79] and the neutral logarithmic law, based on the Monin-Obukhov similarity theory (MOST) [80]. The most basic forms of both methods calculate a vertical wind profile based on single-point surface wind measurements, providing a quick and simple estimate of turbine-height wind speeds and justifying popularity in the wind energy sector.

The power law and neutral logarithmic profile are commonly used to assess hub-height offshore wind conditions from surface buoys or scatterometer measurements. Onea & Rusu (2014) used the neutral logarithmic law to transform in-situ measurements and reanalysis data from the northwest Black Sea to 80-meter wind speeds [81]. They analyzed the seasonal and spatial distribution of wind energy density and found that the area presents energy potential similar to nearby offshore wind farms and development sites. Jang et al. (2009) used the power law to extrapolate 10-meter winds from the QuikSCAT satellite to hub-heights, concluding from wind speed and wind energy density estimations that the western and southern Korean Peninsula are suitable for offshore wind power production [82]. Bethel (2021) extrapolated wind speeds to 100-meter height from four NDBC buoys using the simplified logarithmic law to estimate wind energy generation in the Caribbean Sea [83]. Bethel found that electricity production would be greatest in the central basin compared to the eastern or western borders due to Caribbean Low Level Jet activity.

Despite their historical usage in wind research, more recent research suggests that these conventional physical law extrapolations of wind speed demonstrate limited applicability to the complicated maritime wind climate [66], [84]. Both the simplified logarithmic profile and power law were developed under assumptions of neutral atmospheric stratification in the surface ABL as well as homogenous terrain. These conditions are not trivial in the surface ABL, where atmospheric stability conditions are known to modulate wind conditions across all potential rotor-swept heights [85], [86]. As on land, non-neutral conditions are common in offshore areas; unstable conditions occur when thermal convection generates increased turbulence, and stable conditions occur when turbulence is suppressed by a positive vertical temperature profile. Thus, more realistic profiles of offshore wind speed can be obtained using modified wind profiling models that include atmospheric stability corrections.

One of these modified profiling techniques is the stability-corrected logarithmic law, discussed in detail in the Methods section. Sathe et al. (2011) included surface layer stability-correction terms in the logarithmic law and found agreement with meteorological mast measurements at two sites in the North Sea, though only in unstable and near-neutral conditions [87]. The authors emphasized the importance of making diabatic (heat transfer) corrections to conventional methods for obtaining a valid extended wind profile. Van Wijk et al. (1990) had previously conducted a study in the North Sea evaluating the effect of stability correc-

tions on the wind speed profile, finding that the stability correction improved agreement with observations relative to the conventional logarithmic profile [88]. Domagalski et al. (2019) compared physical wind extrapolation methods at a coastal site in Norway [89]. The researchers demonstrated that under unstable atmospheric conditions, the stability-corrected logarithmic law significantly improved estimates of 100-meter wind speeds compared to the conventional logarithmic law, though little improvement was attained under stable conditions. Overall, it was found that the stability-corrected logarithmic law is a more sophisticated approach to offshore wind speed extrapolation and demonstrates significant improvements over the neutral logarithmic law in all but stable atmospheric conditions.

3.4 Wind Prediction Methods: Intelligent Learning Approaches

The physical laws used for vertical speed wind extrapolation exhibit two immediate shortcomings for wind energy purposes: there are known decreases in prediction accuracy under certain atmospheric conditions, and there is no built-in architecture for temporal forecasting. In response to these limitations, extensive research has focused on intelligent learning models, namely machine learning (ML) methods, for projecting hub-height wind speeds from surface data and providing short-term forecasts. These models can be trained on meteorological data to predict wind characteristics at other spatial or temporal instances. Additionally, while some methods are relatively simple and assume the output to be a linear combination of predictor variables, such as autoregressive integrated moving average (ARIMA), many other ML methods are designed to capture nonlinear or chaotic patterns in the data [11]. A handful of ML methods used in wind prediction research includes Support Vector Regression, multivariate linear regression, artificial neural network, random forest, decision tree regression, radial basis function, and Gaussian process regression [11], [90].

Though all of the aforementioned ML methods have shown significant promise in the field of wind prediction, a select few are of interest in this study based on past research or particular algorithm characteristics. Random forest (RF) is a relatively simple machine learning method that uses regression trees to train on subsets of the training data, which are then averaged to generate a prediction [91]. RF serves as an appropriate baseline reference ML method, given its minimal manual selection of model parameters and its performance against standard extrapolation approaches [92]. Gaussian process regression (GPR) is a probabilistic, non-parametric process that infers a probability distribution over all possible values by assuming that all variables follow one joint Gaussian distribution [93]. GPR is extremely applicable to wind prediction as it works well on small datasets, handles missing data better than parametric models, and provides uncertainty estimates with its predictions. Finally, the Long Short-Term Memory (LSTM) neural network, an improved type of recurrent neural network (RNN), is a state-of-the-art deep learning method that works off

the concept of cellular memory states, which allow for long-term patterns to be preserved for dynamic prediction modeling [94]. LSTM has only recently gained popularity in wind prediction research, but offers great promise in terms of handling diurnal and seasonal patterns and nonlinear changes in wind behavior. The details of these three models are further discussed in the Methods section.

3.4.1 Random Forest (RF)

Random forest (RF) is a powerful ML method that commonly appears in many wind speed and power assessment studies. Bodini & Optis (2020) compared the performance of random forest against the standard power and logarithmic laws at four sites in the central United States and found mean absolute error improvements of 28% and 23% over these respective physical extrapolation approaches [92]. They noted substantial improvement of ML performance when multiple input features, including near-surface winds and the time of day, were included as training parameters. Vassallo et al. (2020) found RF to produce high-accuracy forecasts of one to six-hours ahead wind speed at both the FINO1 offshore platform in the North Sea and the SGP C1 observatory in Oklahoma [95]. Optis et al. (2021) tested RF against a single-column WRF model and both conventional and long-term stability-corrected logarithmic laws [65]. They found that RF produced the most accurate near-surface Atlantic offshore wind speed predictions, and its performance did not degrade when tested on a location 83 kilometers away from the training site, suggesting strong spatial extrapolation capabilities of this technique for the U.S. Atlantic coast. The researchers noted air-sea temperature difference to be an important ML training variable, attributed to its correspondence to atmospheric stability.

3.4.2 Gaussian Process Regression (GPR)

Gaussian process regression (GPR) is another well-studied ML method that has shown great promise for wind speed and power prediction. Lee & Baldick (2014) found GPR to excel over a persistence model at time-series learning for short-term wind power forecasting [96]. Chen et al. (2014) found similar performance metrics as Lee & Baldick for up to one day-ahead wind power forecasting using a combined GPR-NWP model [97]. Wind speed, air temperature, and humidity were found to have a significant influence on power predictions and were thus used as input training variables. Hu & Wang (2015) developed a hybrid GPR model with an empirical-wavelet transform with the ability to produce robust wind speed forecasts at an onshore wind farm [98]. The researchers noted that the predictive distributions and prediction intervals given by GPR, which are not typical of other common ML approaches, are extremely useful for wind energy applications. Yu et al. (2013) proposed combinations of GPR, Support Vector Regression (SVR), and ARIMA with a Gaussian mixture copula model for wind speed prediction at three U.S. wind energy sites [99]. The combined-GPR

model was most effective at characterizing multi-seasonal trends and stochastic variability in wind speed, demonstrating the strong performance of GPR for making long-term wind predictions.

3.4.3 Long Short-Term Memory Neural Network (LSTM)

Long Short-Term Memory neural network (LSTM) has more recently gained popularity in the wind research field and also demonstrates robust wind prediction capabilities. Bethel (2021) achieved high accuracy hour-ahead forecasts of the wind resource in the Caribbean Basin using LSTM [83]. The author noted that inclusion of air temperature and pressure surface variables would likely extend the range of accuracy for the forecasts. Shao et al. (2021) utilized LSTM with hyperparameter tuning optimized by the firework algorithm and found greater accuracy of hub-height wind speed prediction when compared to ARIMA, SVR, and other simpler RNN architectures [11]. Chen et al. (2021) used LSTM in combination with the back-propagation neural network algorithm to perform short-term wind speed forecasting at a wind farm in China and found the Pearson correlation between predicted and true values to be greater than 99% for both test datasets [94]. Banik et al. (2020) utilized LSTM to generate prediction intervals for one-hour ahead wind power uncertainty forecast, finding superior performance to benchmark models such as the Elman RNN and ARIMA [100]. More literature of recent LSTM applications to wind resource characterization can be found in [27], [101]–[103].

3.4.4 Inter-model Comparison

Due to the ubiquity of different ML algorithms in popular use for time series prediction, as well as the recency of LSTM applications to wind energy research, no wind-related literature was found to draw direct comparisons between RF, GPR, and LSTM. The reviewed research most often compared one of these algorithms to simpler methods, as discussed above. In some other cases, RF and/or GPR were compared against a generic artificial neural network (ANN) or an ANN class such as a multilayer perceptron (MLP) or recurrent neural network, which can be thought of as simpler versions of LSTM. Richmond et al. (2020) found RF to produce the most robust hub-height wind speed predictions for a computer-modeled offshore wind farm, though the ANN and radial basis function (RBF), which typically serves as the basis of covariance calculations in GPR, produced the most numerically accurate results on average [90]. Larger amounts of training data were shown to improve the performance of GPR and RBF more than the other models. Shi et al. (2018) modified the traditional RF model with improved feature selection and decision tree reorganization and found that it outperformed the Support Vector Machine (SVM) and simple neural network architectures for 15-minute wind power forecasting at a land-based wind farm [104]. Liu et al. (2018)

found that GPR outperformed both an MLP neural network and SVM model for wind power forecasting in cases of both missing and complete data [93].

The reviewed literature illustrates a complicated picture of relative advantages and shortcomings of these ML algorithms for wind profiling purposes. RF effectively handles missing values, non-linear parameters, and outlier data, though its stable nature may limit its effectiveness in capturing stochastic variability in time series data. GPR also handles missing and nonlinear data well and provides confidence intervals for predictions, but the computational intensity required by its nonparametric design may limit efficiency for large datasets. LSTM has demonstrated high accuracy in wind predictions as it was specifically designed for retaining long-term patterns in time series data, but it requires more hyperparameter tuning to each specific problem and cannot handle missing values.

To the author's knowledge, these three ML algorithms have not been compared within one wind prediction study. Additionally, applications of these algorithms to the Pacific OCS wind resource are nearly absent, likely due to the historical sparsity of data availability and previous lack of research motivation. Comparisons of these ML algorithms against a physical law extrapolation technique will therefore offer new insights into the most suitable method(s) for Pacific OCS wind prediction.

Chapter 4

Methods

4.1 Source Data

4.1.1 Loading & Preprocessing

Data from the DOE/PNNL LiDAR buoys is retrieved from the Atmosphere to Electrons (A2e) Data Archive and Portal, which is supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy's Wind Energy Technologies Office. The data is publicly available at a2e.energy.gov [105], [106].

The data is available at 1-second and 10-minute average intervals. All analysis performed in this study uses the 10-minute average data. For the Humboldt WEA, the data included in this analysis spans from October 2020 to January 2022. Humboldt data is missing from the end of December 2020 through May 2021 due to instrument failure; thus, to retrieve a full annual cycle of wind data, the Humboldt buoy is scheduled to remain deployed until May 2022. The Morro Bay buoy provides mostly continuous data coverage from September 2020 to October 2021.

The motion-corrected Doppler LiDAR data were preliminarily filtered for appropriate signal-to-noise ratios by PNNL. The remaining values are not expected to fall outside of instrument technical specification of ± 0.1 m/s wind speed accuracy [107]. Standard filtering thresholds were also applied by PNNL to the surface-monitoring and in-situ sea surface measurement instruments.

4.1.2 Selection Criteria

Selection of predictors is based on analysis of surface variable correlations to wind speed at 100 meters above mean sea level (AMSL). These input features are denoted in Table 4.1.

Predictor Feature	Unit	Measurement Height (m AMSL)
Surface Wind Speed	m/s	4
Air Temperature	°C	3.7
Sea Surface Temperature	°C	0
Air-Sea Temp. Difference	°C	-
Atmospheric Pressure	mbar	~ 1
Time of Day (cosine)	-	-
Time of Day (sine)	-	-

Table 4.1: Input feature variables used for ML training and prediction of 100-meter above mean sea level (AMSL) wind speed.

Time of day is cyclically encoded into cosine and sine components for use by the machine learning algorithms. This transformation creates a polar representation of time in order to reduce the numerical difference in magnitude between different parts of the day, while also maintaining continuity between consecutive hours of the day. An important drawback of this transformation is that the sine and cosine components must be considered together to identify a unique time of day. Thus, algorithms which handle only one input feature at a time may deflate the overall predictive power of the encoded time features.

Due to the many instances of missing data for periods of one day or longer, train-test dataset selection is carefully conducted. Selection of each dataset is based on the criteria of:

- Parallel day-of-year data availability at both Humboldt and Morro Bay sites
- Approximately one month-long continuous data, with no longer than 1 continuous day of missing data for any variable
- No data overlap with other train-test datasets

The data selection based on the above criteria resulted in six datasets, detailed in Table 4.2. Because of the erratic data availability, assessment of machine learning algorithm capabilities of handling long-term seasonal variability is not included in this analysis.

Within the selected datasets, 2.1% of Humboldt and 1.5% of Morro Bay data remain sporadically missing. Missing values were filled using the mean hourly value within the month containing the missing value. Though random forest and Gaussian process regression do not technically require fully-complete datasets, the same filled datasets are used for all ML methods because of the complete-data requirement of LSTM.

The first 80% of data in each dataset is used for model training, with the remaining 20% used for testing. The 80:20 train-test data ratio is a standard practice in the field of machine learning. The 10-minute data is averaged to hourly data for faster overall computation time.

Dataset Month Range	Dates (Month/Day)	No. of Days
October	10/01 - 10/31	31
November	11/01 - 11/30	30
December	12/01 - 12/28	28
January	01/01 - 01/23	23
June	06/01 - 06/30	30
July-August	07/12 - 08/04	23

Table 4.2: Train-test dataset ranges, selected for data completion and continuity.

4.1.3 Data Scaling & Reduction

Prior to ML algorithm training, input features are scaled through transformation to a standardized z-score in order to prevent features with wider ranges from exerting more influence on the target prediction. This method simply implies subtraction of the data mean and division by the standard deviation of the training samples. Standardization is performed using the StandardScaler() function from the scikit-learn *preprocessing* module in python.

After data is scaled, principal component analysis (PCA) is used to reduce the dimensionality of the input data. PCA is a common statistical practice for reducing the amount of input data while maintaining the variability within data. It generates a new set of uncorrelated features that maximize the variance of the original dataset, and is thus useful for optimizing the efficiency of intelligent learning methods on large datasets. PCA is performed using the python scikit-learn *PCA* module with a lower variance threshold of 0.95, meaning at least 95% of original dataset variance is captured by the modified features.

4.2 Physical Law Calculations

4.2.1 Stability-Corrected Logarithmic Law

The physical law extrapolations of wind speed are conducted using the stability-corrected logarithmic law, denoted as S-C Log Law. The logarithmic profile describes the vertical

distribution of mean horizontal wind speeds within the surface layer of the ABL, which typically ranges up to 100 meters above the surface during the day or approximately 10% of the boundary layer depth [108].

The logarithmic profile method, along with the power law profile, are two of the most commonly-used vertical wind speed extrapolation techniques for wind energy applications [72]. However, the version of the power law which considers atmospheric stability conditions requires two known wind speed measurements at lower and upper heights to calculate the wind shear coefficient, α [109]. As this study seeks to constrain predictor variables to surface-level measurements only, the power law is excluded from the analysis.

The stability-corrected logarithmic wind profile can be written as:

$$U(z) = \frac{u_*}{\kappa}\left[\ln\left(\frac{z}{z_0}\right) - \psi_m\left(\frac{z}{L}, \frac{z_0}{L}\right)\right] \tag{4.1}$$

where U is the horizontal wind speed at height z above the surface, u_* is the friction velocity, κ is the von Kármán constant (0.4), z_0 is the surface roughness, ψ_m is the universal stability function for momentum, and L is the Obukhov length, described in detail in the following subsection.

The friction velocity u_* is a dynamic fundamental scaling velocity derived from surface stress and air density measurements. A sonic anemometer instrument is typically used to capture fine-scale wind speed fluctuations in order to calculate the friction velocity. However, the longitudinal and vertical wind speed fluctuation data measured by the 2D ultrasonic anemometers on the LiDAR buoys are not publicly available. Therefore, the above equation is reformulated into a modified logarithmic law, in which the wind speeds at the buoy surface-monitoring height are used as a reference measurement:

$$U(z) = U_{4m}\left[\frac{\ln\left(\frac{z}{z_0}\right) - \psi_m\left(\frac{z}{L}, \frac{z_0}{L}\right)}{\ln\left(\frac{z_{4m}}{z_0}\right) - \psi_m\left(\frac{z_{4m}}{L}, \frac{z_0}{L}\right)}\right] \tag{4.2}$$

where U_{4m} is the horizontal wind speed at 4 meters above mean sea level and z_{4m} is equal to 4 meters. The height of 4 meters AMSL is used here as it is the height above mean sea level of the buoy's cup anemometer and wind vane.

The surface roughness z_0 is typically determined empirically using the friction velocity. Other methods to estimate z_0 without empirically-derived friction velocity exist, such as analytical methods of the COARE model [110], [111] or statistical methods using data

from multiple vertical wind measurement points [112]. However, the former case involves computationally-intensive iterative calculations and a large suite of input parameters, some of which are not publicly available or reliably measured from the Pacific buoy deployments. In the latter case, there is a need for multiple vertical measurement points throughout the surface layer, which does not meet the constraint of this study to only use surface-level measurements as inputs.

With these considerations, calculations for atmospheric stability and logarithmic wind profiles are simplified and subsequently cross-referenced with COARE 3.6 outputs calculated by PNNL. Instead of the iterative solving used in the COARE algorithm, surface roughness is characterized according to the Davenport roughness classification of $z_0 = 0.0002$ meters for sea surface with fetch greater than 3 km. Roughness lengths of this classification scheme have been well-studied and reassessed [113], and this sea surface roughness parameter is widely used in related offshore atmospheric stability studies [114]–[116].

4.2.2 Monin-Obukhov Similarity Theory

The logarithmic law is conceptually derived from Monin-Obukhov similarity theory (MOST), which describes the vertical gradients of wind speed and temperature as a function of a dimensionless height parameter, z/L [80], [117]. The Obukhov length L represents the lowest height from the surface where turbulent kinetic energy (TKE) produced from buoyancy equates to TKE production by mechanical effects such as shear and friction.

$$L = -\frac{\overline{\theta_v} u_*^3}{\kappa g (\overline{w'\theta'})_s} = \frac{\overline{\theta_v} u_*^2}{\kappa g \theta_*} \tag{4.3}$$

Here $\overline{\theta_v}$ represents the mean virtual potential temperature, g is gravitational acceleration, $(\overline{w'\theta'})_s$ is a measure of the sensible heat flux, and θ_* is the scaled surface layer temperature [118]. In a neutral atmosphere, buoyancy is theoretically absent and L approaches infinity, while in buoyancy-forced stable and unstable conditions L is driven toward small positive and negative values, respectively. The values of L therefore provide a measure of atmospheric stability, which can be generalized into the categories in Table 4.3. [119]–[121].

4.2.3 Bulk Richardson Number

Though turbulent fluxes of momentum and heat are not publicly available through the DOE buoy data, other various empirical methods exist to obtain L [87]. An important value used in these alternative methods is Richardson number (RI) or bulk Richardson number (RI_B),

Stability Classification	Range of L
Very Unstable	$-200 \leq L < 0$
Unstable	$-500 \leq L < -200$
Neutral	$\|L\| > 500$
Stable	$200 < L \leq 500$
Very Stable	$0 < L \leq 200$

Table 4.3: Surface ABL stability classification scheme based on calculated Obukhov length (L).

which are metrics that, similarly to the Obukhov length L, represent the ratio of buoyant to shear production of TKE. The method used in this study to calculate the Obukhov length relates RI_B and L through wind speed and temperature profiles.

The Richardson number (RI) is defined as:

$$RI = \frac{g\Delta\overline{\theta_v}\Delta z}{\overline{\theta_v}(\Delta\overline{u})^2} \tag{4.4}$$

When the Richardson number in Equation 4.4 is considered at the sea surface under assumption of a no-slip bottom boundary, the bulk Richardson number (RI_B) can be calculated using only virtual potential temperatures and wind speed at a reference height (4 meters AMSL in this case), given as:

$$RI_B = \frac{gz_{4m}(\theta_{4m} - \theta_{surf})}{\theta_{avg}U_{4m}^2} \tag{4.5}$$

The surface eddy flux parameters can be used to relate the bulk Richardson number to the Obukhov length, as defined in Equation 4.3. The friction velocity u_* and the surface layer temperature scale θ_* in this context are given as [119]:

$$u_* = \frac{\Delta u\kappa}{\ln(\frac{z}{z_0}) - \psi_m(\frac{z}{L}) + \psi_m(\frac{z_0}{L})} \tag{4.6}$$

$$\theta_* = \frac{\Delta\overline{\theta_v}\kappa}{\ln(\frac{z}{z_0}) - \psi_h(\frac{z}{L}) + \psi_h(\frac{z_0}{L})} \tag{4.7}$$

which, when Equations 4.6 and 4.7 are substituted into Equation 4.3, results in:

$$RI_B = \frac{z_{4m}}{L} \frac{\ln\left(\frac{z_{4m}}{z_0}\right) - \psi_h\left(\frac{z_{4m}}{L}, \frac{z_0}{L}\right)}{\left[\ln\left(\frac{z_{4m}}{z_0}\right) - \psi_m\left(\frac{z_{4m}}{L}, \frac{z_0}{L}\right)\right]^2} \tag{4.8}$$

Stability correction functions must be chosen in order to solve for L. These functions represent universal relationships between the non-dimensionalized variables in MOST, which are related based on the Buckingham Pi Theorem [122]. A simple RI_B-based method for estimating z/L, derived from Businger et al. (1971) [123], is shown below in Equation 4.9 to illustrate the generalized relationship between RI_B and atmospheric stability. It demonstrates that a positive or negative RI_B can be used to binarize atmospheric conditions into stable and unstable, respectively. This relationship is used in this study solely to determine which stability functions to use at each timestep in terms of RI_B. It is not included in any numerical calculations because of the singularity at $RI_B = 0.2$, above which this conversion is not valid and the z/L parameter cannot be accurately calculated [111].

$$\frac{z}{L} = \begin{cases} \frac{10RI_B}{1-5RI_B} & \text{if } RI_B \geq 0 \\ 10RI_B & \text{if } RI_B \leq 0 \end{cases} \tag{4.9}$$

4.2.4 Stability-Correction Functions

Many definitions of the stability correction functions for momentum ψ_m exist in the literature. The standard correction functions most commonly used are the Businger and Dyer functions [123], [124]. Multiple studies have tested the validity of these functions and found that they do not hold up well under strongly unstable conditions, especially beyond the free convection limit [125]. Therefore, updated stability correction functions for momentum were proposed for unstable and stable conditions and are written respectively as [126], [127]:

$$\psi_{m,unstable}\left(\frac{z}{L}\right) = 1.5\ln\left(\frac{1+y+y^2}{3}\right) - \sqrt{3}\arctan\left(\frac{2y+1}{\sqrt{3}}\right) + \frac{\pi}{\sqrt{3}} \tag{4.10}$$

$$\psi_{m,stable}\left(\frac{z}{L}\right) = -a\left(\frac{z}{L}\right) - b\left(\frac{z}{L} - \frac{c}{d}\right)\exp(-d\left(\frac{z}{L}\right)) - \frac{bc}{d} \tag{4.11}$$

where in Equation 4.10 $y = (1 - \gamma_{FC}\left(\frac{z}{L}\right))^{1/3}$ and $\gamma_{FC} = 10$ [126] and in Equation 4.11 $a = 1, b = \frac{2}{3}, c = 5,$ and $d = 0.35$ [127].

Temperature stability correction functions are taken from the landmark 1968 Kansas atmospheric surface layer experiment [128], [129] with a correction for highly stable ($z/L > 1$)

conditions, detailed further in Blumel (2000) [130]. For unstable, stable, and highly stable conditions respectively, these equations are:

$$\psi_{h,unstable}\left(\frac{z}{L}\right) = 2\ln\left(\frac{1+x^2}{2}\right) \tag{4.12}$$

$$\psi_{h,stable}\left(\frac{z}{L}\right) = -\beta\left(\frac{z}{L}\right) \tag{4.13}$$

$$\psi_{h,superstable}\left(\frac{z}{L}\right) = -\beta\ln\left(\frac{z}{L}\right) - \beta \tag{4.14}$$

For the unstable case temperature function (Equation 4.12), $x = (1 - \gamma\left(\frac{z}{L}\right))^{1/4}$. The standard empirical constants γ and β are set to 16 and 5 respectively [124], [129]. L can be solved for in Equation 4.8 using the RI_B calculated in Equation 4.5 along with the stability correction functions for both momentum (Equations 4.10 and 4.11) and temperature (Equations 4.12, 4.13, and 4.14). The stability-corrected logarithmic wind profile can then be obtained from Equation 4.2.

4.3 Machine Learning Techniques

The main ML prediction data considered in this study are generated from training each algorithm on one buoy location's data and testing on the other (i.e. Morro Bay predictions are made from an algorithm trained on Humboldt data and vice versa, unless otherwise stated). This is to eliminate model bias to the same train-test location and to simulate a realistic application of the model to a different spatial location.

4.3.1 Random Forest

Random forest (RF) is one of the most commonly-used ML algorithms across the field of data science. It is a supervised ensemble ML technique, meaning that it uses labeled data and aggregates multiple submodels to make predictions. In this case, the submodels are decision trees, which work by finding the optimal ways to make binary splits in input data to generate the most accurate possible prediction of the output. In the case of continuous input data, such as the meteorological and oceanic DOE buoy data, the decision trees are often called "regression trees". The RF algorithm constructs independent regression trees to subsets of the training data, then aggregates these weak predictions into a more accurate, overall prediction [95].

One example regression tree is shown in Figure 4.1, but tens to thousands of these are used within one RF model. The input feature with the smallest mean squared error (MSE)

from the split is used to create the division at each node. Figure 4.1 demonstrates how certain feature variables classify the output more effectively than others and thus appear more often in the tree. RF calculates feature variable importance based on the relative influence of each predictor variable on the output prediction.

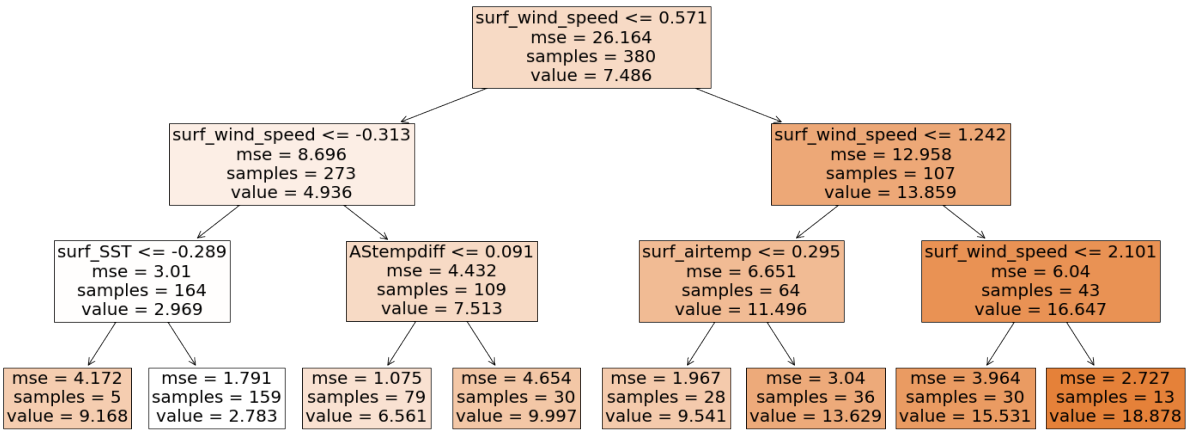

Figure 4.1: Example of one regression tree with depth = 3. The regression tree finds the optimal splits in a subset of normalized training data to reach a prediction (denoted as "value") as close as possible to the observed 100-meter AMSL wind speed. Darker coloration indicates a higher value prediction. The features used for splitting include 4-meter AMSL wind speed, sea surface temperature, air temperature, and air-sea temperature difference (denoted as "surf_wind_speed", "surf_SST", "surf_airtemp", and "AStempdiff" respectively).

The steps of random forest regression are shown in Figure 4.2.

1. T sample subsets are drawn randomly with replacement from the original training dataset. Within each training sample subset, a subset of feature variables are randomly selected. This selection process is based on the random subspace theory, also known as feature bagging [131], which is a learning method that uses random samples of feature variables instead of the whole feature set in order to reduce correlation between the predictions of the models in the ensemble [104].

2. Each regression tree is fit to its training data, wherein each node applies its own split function to the training data inputs in a process called recursive partitioning. This generates a unique prediction function for each decision tree, such as the example regression tree in Figure 4.1 above.

3. Each regression tree is then applied to the testing data to generate a prediction for each timestep.

4. By nature of the selection process, approximately one-third of the original training data are excluded from subsetting. This portion of unused data, called the Out-of-bag samples, is used as a reference for an unbiased estimate of the generalization error [132].

5. The predictions from each individual regression tree are averaged to produce the combined forecasted value(s).

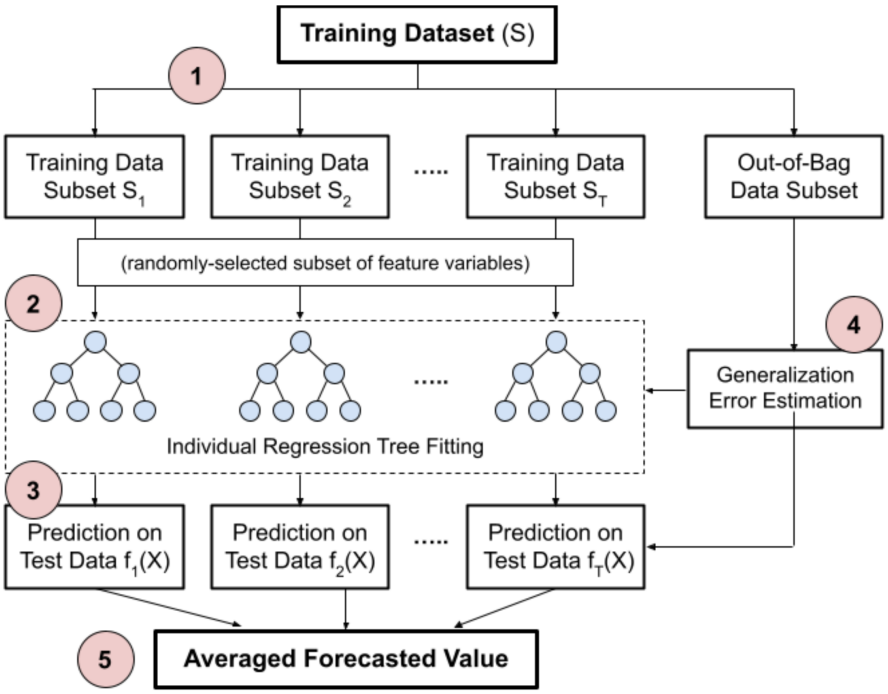

Figure 4.2: Schematic representation of random forest regression. The general steps of the algorithm are labeled in red circles.

Advantages: Random forest is advantageous for its small variance within the model, as noise in the training dataset ends up being smoothed out through averaging of the non-correlated trees (given a large enough set of decision trees). It also effectively handles missing data, requires minimal manual hyperparameter tuning, and because of the randomness in training subsetting and feature selection, does not overfit to training data. RF also contains

built-in cross-validation through calculating the Out-of-bag error (OOBE), also called generalization error, which is the average difference between a predicted and actual outcome from trees that were not trained on each observation. OOBE provides a preliminary example of model testing performance. Variable importance (VI) is also obtained by calculating OOBE before and after permuting a feature variable.

Disadvantages: A notable drawback of RF is the high computational cost of building a larger number of individual decision trees, which is generally a trade-off for increased prediction accuracy. The runtime of the algorithm is relatively high compared to the other algorithms in this study due to the large number of training subsets and combination of outputs. Additionally, because of the ensemble averaging, RF works as a "black box" algorithm, in that the calculations made by each tree to produce the prediction are not easily accessible or interpretable from the final output.

Implementation: Random forest regression is performed using the *RandomForestRegressor* module from the scikit-learn Python machine learning library. 3-fold grid search cross validation is used to determine the optimal hyperparameter values of the RF model for each train-test dataset. The range of values considered for each hyperparameter are shown below in Table 4.4.

RF Hyperparameter	**Range of Values**
No. of estimators (trees)	10 - 1000
Maximum depth	4 - 40
Max. No. features considered at a node	2 - 6
Min. No. samples to split a node	2 - 10
Min. No. samples at each end node	1 - 15

Table 4.4: Hyperparameters optimized in each training run of random forest regression.

4.3.2 Gaussian Process Regression

Similarly to RF, Gaussian process regression (GPR) is a supervised ML method used widely for both regression and classification problems. Unlike the traditional RF model, GPR is probabilistic in that it calculates a probability distribution over the prediction, lending GPR the powerful advantage of quantifying empirical uncertainty over all of its predictions [133].

In short, GPR defines a mathematical distribution over all the possible functions that can fit some observed data points. For visualization purposes, consider the univariate example of

predicting 100-meter wind speeds only from surface-level wind speeds, shown in Figure 4.3 below. Functions that best map the input to the outputs, called the "posterior distribution", are selected from a "prior distribution" of infinite possible functions. The posterior functions are averaged to create the mean prediction for any input value. Areas with less input data have larger confidence intervals due to the increased spread of possible underlying functions. For D input feature variables, this process is conducted in D-dimensional space.

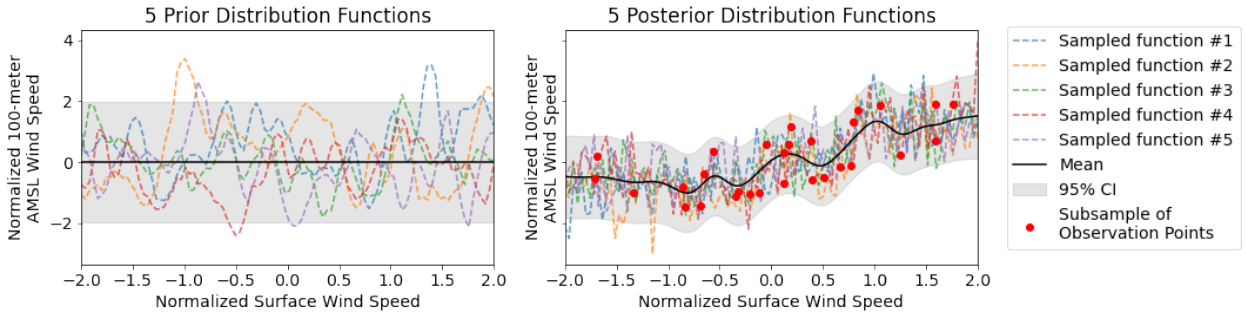

Figure 4.3: Univariate example of selecting functions from the prior distribution to fit the input and output observational data, then averaging to get an estimated mean function. Confidence intervals are determined by the probability distribution over these underlying functions. Only some sample functions and observational data points are shown for demonstration purposes.

The essential mathematical principles of GPR are the Gaussian distribution, multivariate normal distribution, kernels, and non-parametric models [134]. These concepts and their implementation into GPR will be briefly reviewed.

The probability density function (PDF) of the Gaussian (normal) distribution of a variable X with mean μ and variance σ^2 is defined in Equation 4.15:

$$P_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{4.15}$$

In the case where more than one correlated feature variable is used to describe the system, a multivariate normal distribution (MVN) is needed to combine these variables into one Gaussian model. For D feature variables, the PDF of the D-dimensional MVN is defined in Equation 4.16:

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \tag{4.16}$$

where μ is the vector of variable means and Σ is the symmetric DxD covariance matrix that stores the pairwise covariance of the jointly-modeled feature variables. The diagonal terms of Σ are the independent variance of each variable, and off-diagonal terms are the correlations between the two different feature variables. The purpose of this PDF is to generate smooth functional relationships between the feature variables (assuming non-identity covariance between the features). This is used to produce the prior distribution of possible functions, from which the functions with best fit to observational data are selected. A sample visual of the PDF of an MVN with 2 input features is given in Figure 4.4.

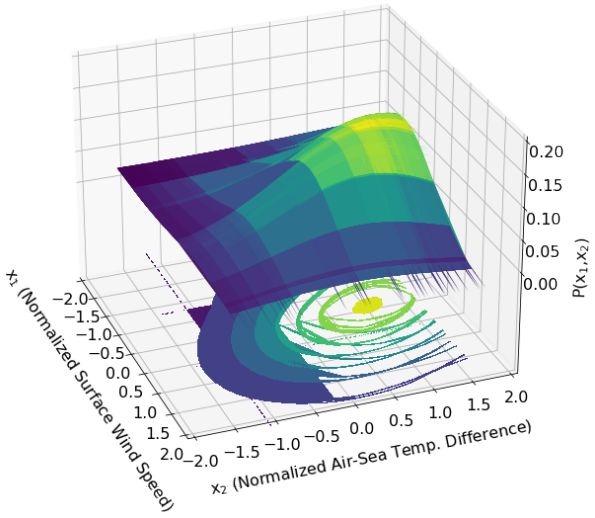

Figure 4.4: Sample visualization of the PDF of a bivariate (2-dimensional) normal distribution. This relationship can be described by Equation 4.16 with D = 2 and non-identity covariance between the predictor variables.

Covariance in regression is based on the idea of how similar input data is likely to produce similar target values. Kernel functions are used to quantify covariance between variables and generate the smoothed functional relationships between them. The squared exponential kernel function, also known as the radial basis function (RBF), is used as the default smoothing function for GPR and is defined as follows:

$$k(x_i, x_j) = cov(x_i, x_j) = exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right) \tag{4.17}$$

In Equation 4.17, ℓ is the length scale of correlation, an empirical constant that describes the general smoothness of the function.

It is important to note that the functions generated by the MVN are called "kernelized prior functions", as they are generated by the MVN model with kernel functions as prior knowledge before taking observed data points. When observational data points are inputted, only select functions that best fit these points are kept from the infinite number of functions to create the posterior distribution. GPR is nonparametric, meaning that it does not assume that a limited number of parameters can be defined to form the function over the output. It therefore obtains predictions from calculating the probability distribution over all of the posterior functions, then taking the mean of these selected functions. With every input of new training data, this posterior distribution is updated and averaged to best fit the observational data.

The above descriptions are to give a sense of the complex MVN and covariance operations occurring in the background of GPR. Overall, the practical steps of GPR can be summed up succinctly. Given the training data, GPR uses negative log marginalized likelihood to fit optimal hyperparameters to the covariance matrix Σ, employs this covariance matrix to obtain the predictive distribution (posterior) from the kernelized prior functions, averages the posterior distribution to obtain prediction results, then uses the covariance matrix and the prediction results to calculate the confidence interval.

Advantages: The hyperparameters of GPR can be tuned more specifically to an individual problem than RF, especially within the choice of the kernel/covariance function. Hyperparameters within the RBF include vertical scale (the values that the averaged function can span) and horizontal scale (the strength of correlation between two points as distance between them increases). Other hyperparameter specifications include the mean of the prior distribution (either zero or the training data mean) and allowable noise level in the targets. The empirical confidence intervals in areas of interest allow the user to decide if or how the prediction should be refit.

Disadvantages: The main constraint of GPR is its cubic growth of computational complexity with the number of feature variables (dimensions). Large losses of efficiency occur when the number of features exceeds a few dozen. Together with quadratic memory consumption [134], GPR can become inefficient for large datasets with numerous feature variables. This is because it is not sparse, meaning that it requires complete data from each sample point to perform predictions. Sparse Gaussian processes are a common workaround to this issue [135]. Similarly to RF, the internal processes of GPR make it a "black box" model,

meaning that the complex mathematical relationships between the variables are difficult to extract and interpret.

Implementation: Gaussian process regression is performed using the *GaussianProcessRegressor* module from the scikit-learn python machine learning library. All inputted values are normalized by mean removal and scaling to unit variance to ensure the Gaussian assumptions of GPR are met. An additive kernel consisting of RBF and noise kernels is used. The smoothness of the output function, defined by the RBF length scale ℓ, is optimized within the bounds (1e-2, 1e3) with the initialization value = 0.1. The noise level is optimized within the bounds (1e-15, 1e5) with the initialization value = 0.1^2.

4.3.3 Long Short-Term Memory Neural Network

Long Short-Term Memory (LSTM) neural networks were introduced in 1997 by Hochreiter and Schmidhuber [136] as an advanced type of recurrent neural network (RNN) to more effectively model long-term dependencies. RNNs are still a widely used machine-learning method for modeling of dynamic temporal data [137]. The basic principles of RNNs and LSTMs are the same; a short introduction to RNNs is given to contextualize the flow of information within LSTM.

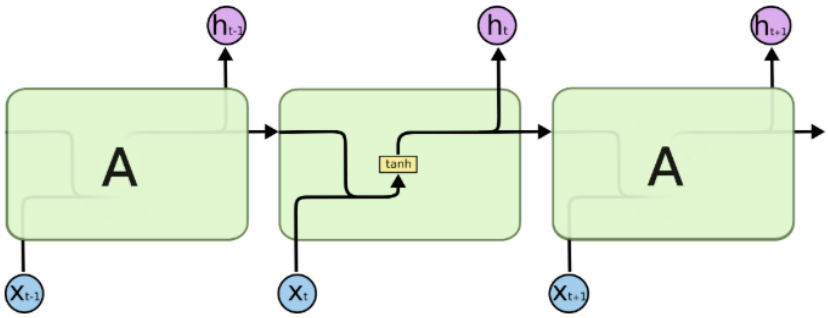

Figure 4.5: The repeating module of a standard RNN, which contains only a single tanh layer for transforming input data. (Image adapted from Olah, 2015 [138]).

The linearized flow of information in a traditional RNN is visualized above in Figure 4.5. In this diagram, A is a cell of the neural network, x_t is the input data at time t, and h_t is the output data. RNNs and LSTMs have a chain of repeating modules, each of which takes input from both previous and current timesteps, generates an output, and passes this information on to the next module. The RNN module contains one *tanh* layer that transforms values to [-1, 1] to regulate transformations performed on the data.

The actual training process of a neural network involves feeding information forward to make a prediction, then propagating the error of the prediction backward through the network to calculate the internal weights for the input(s) of each node. The calculated value to adjust the weights is called the gradient, and this part of the training process is called back-propagation through time (BPTT).

The main issue with RNNs is that they do not contain architecture that retains memory of long-term patterns. During BPTT, gradient values of distant modules can become extremely small, meaning that they exert little or no weight on the current output. This is known as the "vanishing gradient" problem.

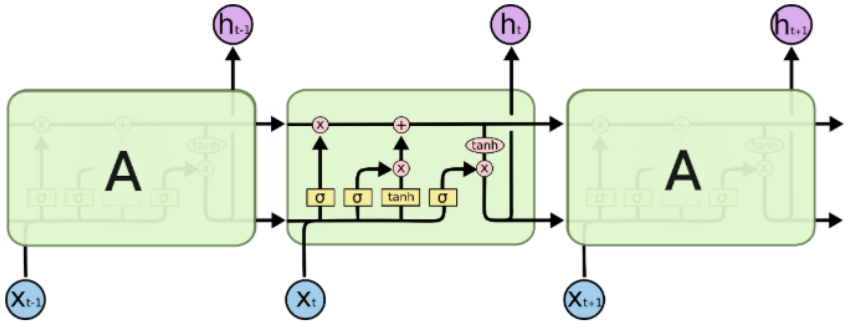

Figure 4.6: The repeating module of LSTM, which contains four layers that transform input data and retain long-term relationship patterns between predictors and targets. (Image adapted from Olah, 2015 [138]).

Long Short-Term Memory neural networks resolve these common issues faced with RNNs. In LSTM, visualized in Figure 4.6, the h_t output represents the hidden state, which is an encoded representation of the previous inputs. The hidden state at each timestep feeds information into a "cell state", which serves to aggregate the memory of all the previous weighted time-steps [139]. Additionally, LSTM resolves both the vanishing and exploding gradient issue by employing a more complex series of logistic layers in each LSTM cell, unlike the single *tanh* layer used by traditional RNNs. These layers of an LSTM cell consist of three logistic sigmoid (σ) gates and a *tanh* gate, which filter the information that determines the cell output and the next cell state.

The first sigmoid (σ) layer is known as the "forget gate", as it takes in the current data point and the previous cell's hidden state, and it outputs a number in [0,1] that is multiplied with the previous cell state. This output represents a range between rejecting all of the previous cell's information (0) versus keeping all of it (1). The second σ layer decides what new information is to be added to the cell, and the coinciding *tanh* layer outputs new

candidate values to store in the cell state. Multiplication of these two layers determines the amount of information added to the cell state. This result is then added with the previous cell state multiplied with the forget gate output, producing the current cell state. Lastly, the third σ layer decides what part of the cell state to include in the output, and uses a *tanh* function to scale the cell state into the range [-1,1]. Multiplication of these two layers' results produces the hidden state of the cell h_t, which is then carried over to the next timestep along with the new cell state.

Advantages: LSTM is state-of-the-art neural network architecture that captures patterns with long-term dependencies within time series data, which RF and GPR are not specifically designed to handle. The structure of this neural network is designed to handle noise and distributed representations. It is effective for a wide range of parameter specifications, such as learning rate and output bias, which allows the user to tune the network for a specific need. Due to its adaptability and robust handling of data inputs, LSTM is becoming increasingly popular for sequence-based learning tasks including text generation, speech recognition, and language translation [140].

Disadvantages: The LSTM neural network requires a high memory-bandwidth because of linear layers present in each cell. Out of the ML algorithms discussed, LSTM is the most prone to overfitting and often requires a dropout algorithm to curb this issue. Dropout is a regularization method where input and recurrent connections to LSTM units are probabilistically excluded from activation and weight updates while training a network. As is the case with RF and GPR, LSTM is also a "black box" model, in which the learning process of the algorithm is difficult to elucidate from all of the nonlinear combinations of neurons.

A notable disadvantage of LSTM is the need for complete training and testing datasets at regular time intervals. Adverse weather conditions and instrument failure make missing values a common occurrence in wind measurement data, meaning that missing data must be filled in order to employ LSTM.

Implementation: LSTM is implemented using modules from the Keras Python library using a Tensorflow backend. A stacked LSTM with four LSTM layers is used to increase levels of abstraction to the input data, essentially creating representations of the data at different timescales. A dropout algorithm with rate = 0.2 is applied to randomly omit 20% of each layer's information in order to help prevent overfitting. Mean absolute error (MAE) is used as the loss function, and the Adam optimization algorithm, which is an efficient method of stochastic gradient descent, is used as the optimizer. 200 epochs and a batch size of 72 are used for iteration over the dataset, which were determined by empirical observations of the loss function and runtime.

4.4 Error Metrics

The National Renewable Energy Laboratory (NREL) recommends to use bias, centered root-mean-square error (cRMSE), square of correlation coefficient (R^2), and Earth-mover's distance (EMD) as the performance metrics for offshore wind model validation [66]. The decomposition of typical error metrics, such as RMSE and MAE, into more distinct components is useful for identifying specific aspects of individual model error. RMSE is also taken into account as a metric of overall prediction quality. Each model's performance over the six train-test datasets is averaged to obtain these values. The purpose of using each performance metric is detailed in Table 4.5.

Error Metric	Description	Purpose
Bias	Summed difference between each model prediction and the actual value	Indicates if model is generally overestimating (positive bias) or underestimating (negative bias) the forecasted values
cRMSE	Differences in model variation around the mean after bias is removed	Quantifies the random error (noise) of model predictions, i.e. the spread of residuals around the actual values, without effects of model bias
$\mathbf{R^2}$	Proportion of the variance in the predictions explained by the actual values	Evaluates the general strength of the relationship between predicted values and actual values, scaled between 0 - 100%
EMD	Difference between the probability distributions of the predicted and actual values	Measures the "cost" of transforming one distribution to the other, considering the integral difference and distance to "move" data

Table 4.5: Error metrics used to assess the performance of the machine learning algorithms and the stability-corrected logarithmic law.

Chapter 5

Results & Analysis

5.1 Hub-height Wind Characteristics

Frequency distributions of hub-height (100 meters AMSL) wind speeds in both WEAs are shown in Figure 5.1. Wind speeds at hub-height in these areas average around 8-10 m/s, with strong gusts reaching 25-30 m/s. These wind speed attributes are desirable to wind energy developers, as the technically viable lower limit for offshore wind speeds is 7 m/s for the current performance standards of offshore wind turbines [141] and the cut-out speed is roughly 25 m/s.

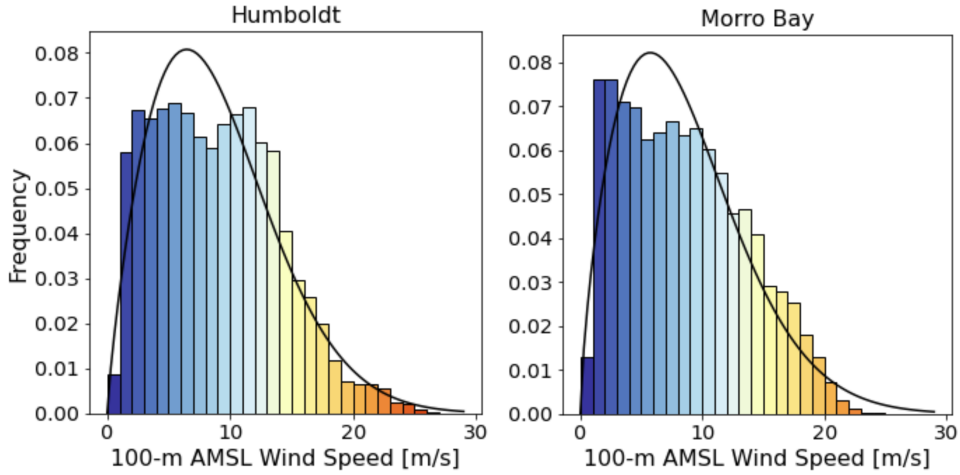

Figure 5.1: 100-meter AMSL wind speed frequency distributions in the Humboldt and Morro Bay WEAs. The data for Humboldt runs from October 2020 to January 2022 with data missing February through May. The data for Morro Bay runs from September 2020 to October 2021.

The curve over the data in Figure 5.1 is the Weibull distribution, which describes the theoretical wind speed probability distribution. Wind energy developers often use the Weibull distribution to inform the design and projected generation capacity and costs of their turbines. The curve is fit to each location's data using the maximum likelihood method. In both locations, the wind speed with highest projected frequency, approximately 6-7 m/s, does not match the expected distribution shape. This is to demonstrate that more sophisticated methods are necessary to accurately model even general hub-height wind characteristics in these areas.

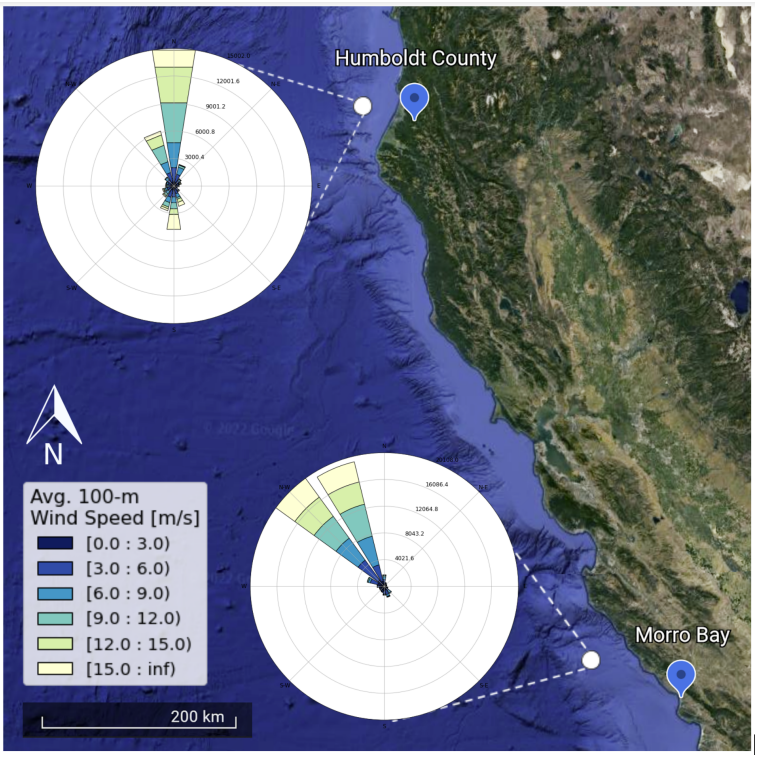

Figure 5.2: 100-meter AMSL wind direction distributions in the Humboldt and Morro Bay WEAs. The dates of the source data are the same as Figure 5.1 above. (Satellite imagery adapted from Google Earth).

Average hub-height wind direction is obtained from the LiDAR data and shown in Figure 5.2. Hub-height winds hail predominantly from the north in the Humboldt WEA and from the northwest in the Morro Bay WEA. The occurrences of wind direction reversal, visible in the Humboldt windrose, take place in October through December, though other aspects

of the seasonal wind direction pattern may be missing due to the data gap from February through May.

Prediction of hub-height wind direction is excluded from this analysis for two main reasons. The first reason, as shown above, is that hub-height wind directions in both locations demonstrate a generally unidirectional nature, whereas wind speed appears to be a more dynamic dependent variable for evaluating machine learning wind modeling capabilities. The second reason is the long-term miscalibration of the surface wind vane and surface anemometer on the Morro Bay buoy. Surface wind direction measurements from the Morro Bay buoy are yet to be corrected in the published data. Therefore, wind direction is not included as a predictor variable in this study. As surface wind direction would likely be vital predictor data for estimating hub-height wind direction, wind direction is also not a part of the dependent variable of this study (i.e. predictions are made for wind speed but not wind velocity). In any case, the exclusion of wind direction does not detract from the value of information gained in this study, as the yaw system built into wind turbines is designed to orient the rotor toward these generally unidirectional winds.

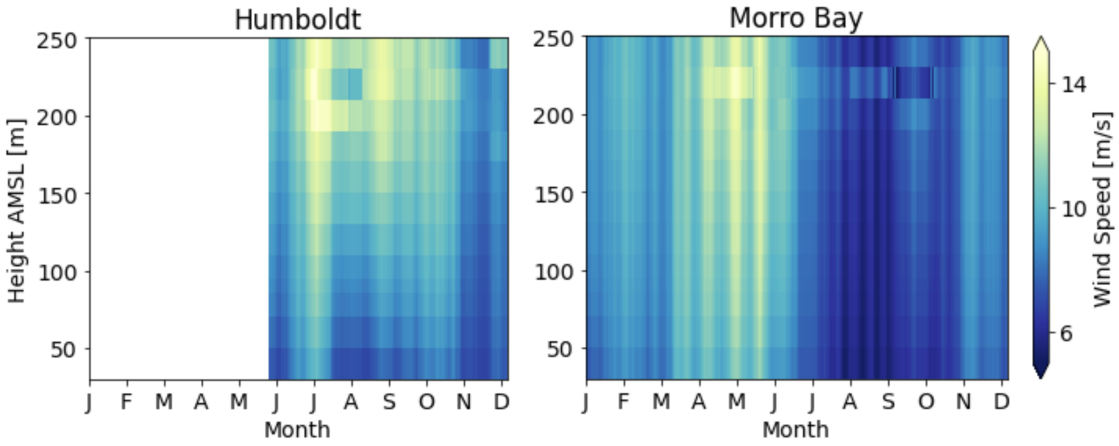

Figure 5.3: Average interseasonal wind speed profiles from 40 to 250 meters AMSL in the Humboldt and Morro Bay WEAs.

The LiDAR instrument data from 40 to 250 meters AMSL provide surface ABL wind speed profiles across seasonal timescales. The profiles displayed in Figure 5.3 were created using a 30-day window rolling average of wind speeds at each elevation. A 30-day window is chosen to best display the average wind speed trend over a month of data at a time, while also being short enough to exhibit differences between seasons. The irregular signatures appearing above 175 meters may be due to a higher ratio of missing data at these heights.

Less consistent data at higher elevations is an inherent drawback of LiDAR wind monitoring, as signal-to-noise ratios from these heights are more likely to be of lower quality and result in data infidelity.

Clear seasonal signatures of increased wind speed occur in springtime months at the Morro Bay site, with the subsequent profile minima occurring in the summertime months. The seasonal wind speed profile variation in Humboldt is less clear due to the large volume of missing data. However, from the available data, the annual wind speed pattern appears dissimilar between Humboldt and Morro Bay, as the magnitude of wind speeds in Humboldt are not nearly as low as those in Morro Bay in the summer and fall months.

The differing seasonal wind speed patterns between these two Pacific OCS locations pose a potential problem for accurate spatial extrapolation of machine learning algorithms from their training locations. This is because machine learning algorithms used for time-series forecasting and prediction may pick up on these long-term seasonal patterns and incorrectly apply these assumptions to the testing location. The ML predictions in this study are likely less prone to these types of errors, as they are trained on data no longer than a month instead of across seasonal timescales. This consideration is more important for future studies of the Pacific OCS wind resource that may use multiseasonal data from one location to perform wind prediction or forecasting at another location.

5.2 Selection of Predictor Variables

The purpose of this section is to demonstrate the correlations between certain surface variables and hub-height wind speeds. Establishing correlation is necessary for choosing predictor variables that will likely exert weight on the predictive decisions made by an ML algorithm.

5.2.1 Surface Wind Speed

For both WEAs in Figure 5.3, the positive vertical gradient of wind speeds present at nearly all timepoints in the collected data indicates some correlation between surface wind speed and hub-height wind speed. Basic characteristic surface wind speed frequencies are displayed in Figure 5.4, where the “surface” is measured at 4 meters above mean sea level. In both locations, these wind speeds are markedly lower on average than hub-height wind speeds and rarely exceed 20 m/s.

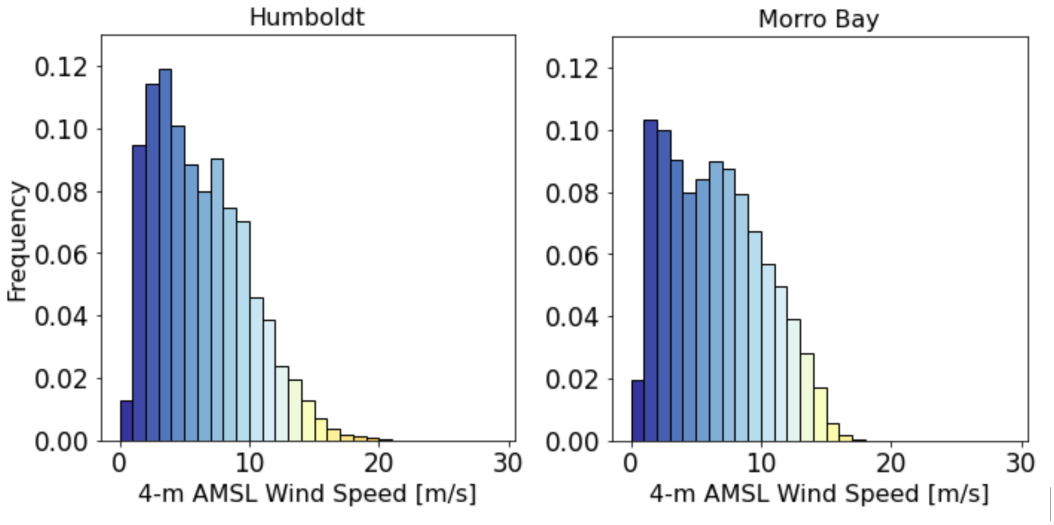

Figure 5.4: 4-meter AMSL ("surface") wind speed frequency distributions in the Humboldt and Morro Bay WEAs. The dates of the source data are the same as Figure 5.1 above.

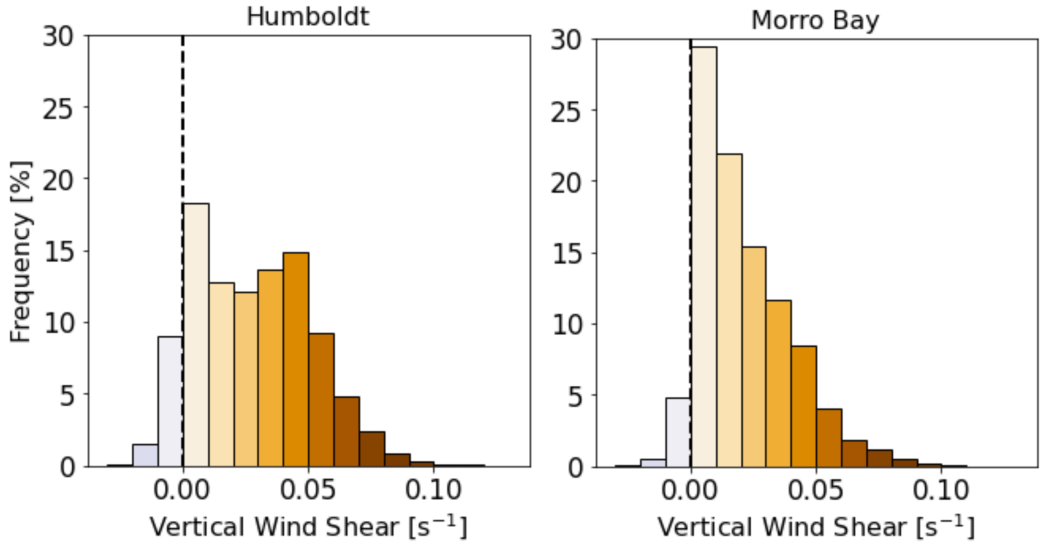

Figure 5.5: Vertical wind shear frequency distributions in the Humboldt and Morro Bay WEAs. Vertical wind shear quantifies the variation in average horizontal wind speed over a difference in altitude. Shear is measured between horizontal wind speeds at 4 and 100 meters AMSL. The dates of the source data are the same as Figure 5.1 above.

The connection between wind speeds at surface and hub-height is explored in Figure 5.5. Vertical wind shear measures the variation in wind speed over a difference in altitude, which in this case is 96 meters between measurement locations. Vertical wind shear can be used to characterize the wind profile across the rotor-swept area, and the degree of shear is important for wind energy developers to estimate average power density and rotor fatigue loads [142]. Equation 5.1 shows the simple calculation of average vertical wind shear at each timepoint.

$$s = \frac{v_2 - v_1}{h_2 - h_1} \tag{5.1}$$

The frequency distributions of vertical wind shear quantify the primarily positive wind shear in both locations, which ranges generally from -0.03 to 0.1 s^{-1}. The pronounced shear frequency profile, especially that of Morro Bay, suggests a relationship between surface and hub-height wind speeds, in which the variation with increasing altitude is most frequently positive and relatively small. In other words, the primarily positive values and distinct peak shape of the shear profile indicate that covariance between the two wind speed parameters is not random. Surface wind speed is thus considered a favorable input variable for training the ML algorithms to predict hub-height wind speed. The importance of surface wind speed for hub-height wind speed prediction can also be inferred from historical usage of the power law and logarithmic profiles, which in their simplest forms only rely on wind speed measurements from a near-surface reference height.

5.2.2 Air & Sea Surface Temperature

Air-sea temperature difference shows a clear seasonal diurnal signature in both WEAs, as can be seen in Figure 5.6. The periods from 3:30 to 15:30 UTC (8:30 PM - 8:30 AM PST during daylight savings time or 7:30 PM - 7:30 AM PST otherwise) are shaded in light blue to indicate approximate nighttime conditions, and approximate daytime conditions are shaded in light red from 15:30 to 3:30 UTC (8:30 AM - 8:30 PM PST during daylight savings time or 7:30 AM - 7:30 PM PST otherwise).

As the heat capacity of air is lower than that of water, air-sea temperature difference generally drops at nighttime as the air cools faster than the ocean, then increases again as air heats faster than the water in warmer daytime conditions. Air-sea temperature difference is primarily negative in wintertime months as the average air temperature generally drops below the ocean temperature in both offshore locations. Unstable and convective atmospheric conditions commonly occur when the water surface temperature is higher than the air temperature, though stability is also affected by other factors including wind speed and humidity [143].

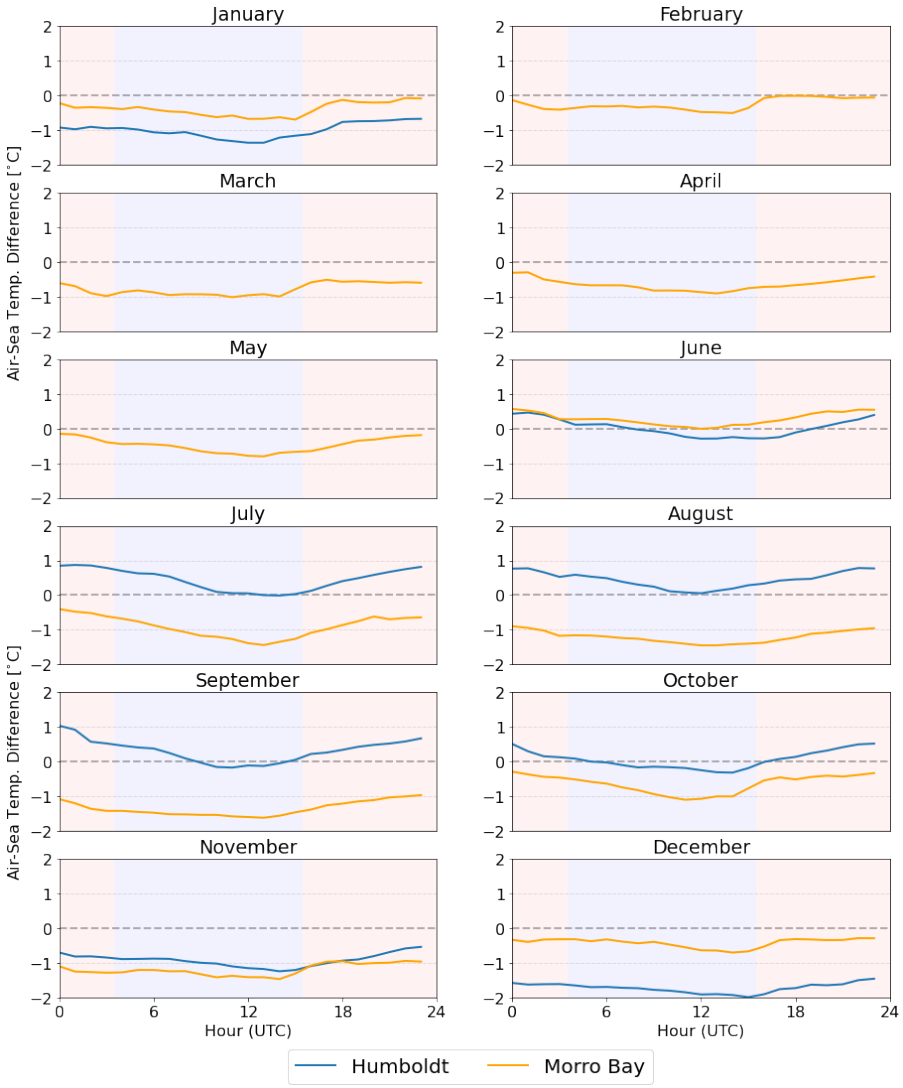

Figure 5.6: Average diurnal cycles of air-sea temperature difference by month in the Humboldt and Morro Bay WEAs. Red and blue background coloration refers to approximate local daytime and nighttime conditions. Data is missing February through May for the Humboldt WEA.

A weak positive correlation between air-sea temperature difference and 100-meter AMSL wind speeds can be observed in Figure 5.7. These correlations are generally stronger when data is separated on a monthly basis due to the different ABL stability and wind speed patterns occurring throughout the year, but month-by-month scatter plots are omitted for visual simplicity. This finding underscores the potential predictive power of sea surface

temperature (SST), air temperature, and difference between them, as well as the best practice to include time of year as an ML predictor variable if train-test datasets span multiple months. However, as the ML algorithms in this study are trained and tested on a single month of data at a time, these three variables are included as predictors of hub-height wind speed without a requisite month-of-year variable associated.

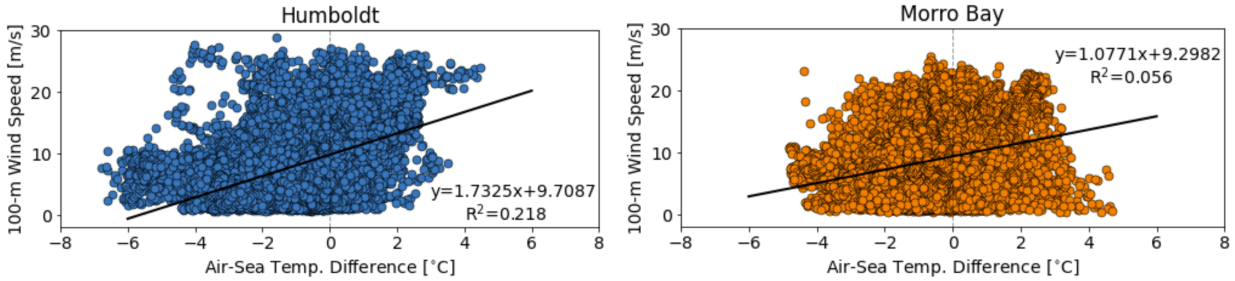

Figure 5.7: Correlation between 100-meter AMSL wind speeds and air-sea temperature difference values in the Humboldt and Morro Bay WEAs. Note that these correlations smooth over the generally stronger relationships exhibited on a monthly basis.

5.2.3 Surface Atmospheric Pressure

Winds are generated from differences in atmospheric pressure between two locations. Though the pressure differential between each WEA and nearby locations would be a more ideal parameter to correlate with wind speed, predictor variables in this study are constrained to local surface-level measurements. Thus, surface atmospheric pressure at the buoy locations is correlated to hub-height wind speed. Results are displayed in Figure 5.8.

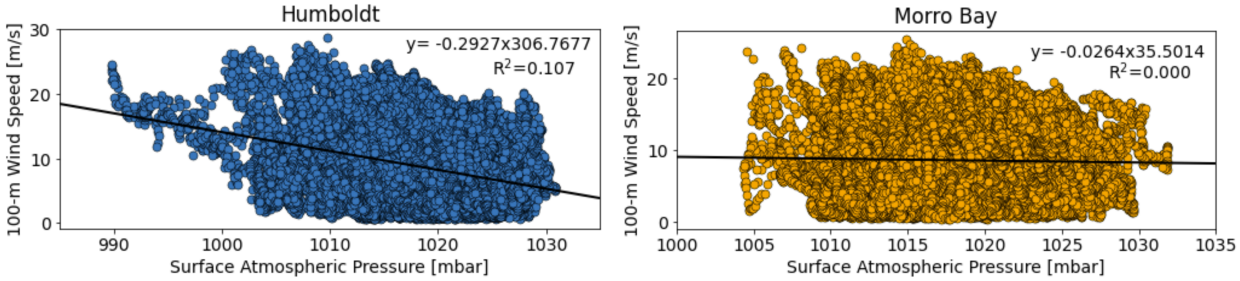

Figure 5.8: Correlation between 100-meter AMSL wind speeds and surface atmospheric pressure values in the Humboldt and Morro Bay WEAs. Note that these correlations smooth over the generally stronger relationships exhibited on a monthly basis.

A weak negative correlation between surface atmospheric pressure and 100-meter AMSL wind speeds can be observed, with near-zero overall correlation exhibited in the Morro Bay site. However, similar to the air-sea temperature difference patterns discussed above, in most cases correlation strength increases and sometimes even reverses upon evaluating month-by-month data. Surface atmospheric pressure is thus considered a potentially valuable micrometeorological component to include in the suite of predictor variables for ML training and testing.

5.2.4 Time of Day

Diurnal patterns of wind speeds at 100 meters AMSL in both WEAs are illustrated in Figure 5.9. Hub-height wind speeds exhibit clear diurnal signatures year-round in the Morro Bay location, while more stochastic variation is present at the Humboldt site between different months. The day and night coloration corresponds to the same times as Figure 5.6 above.

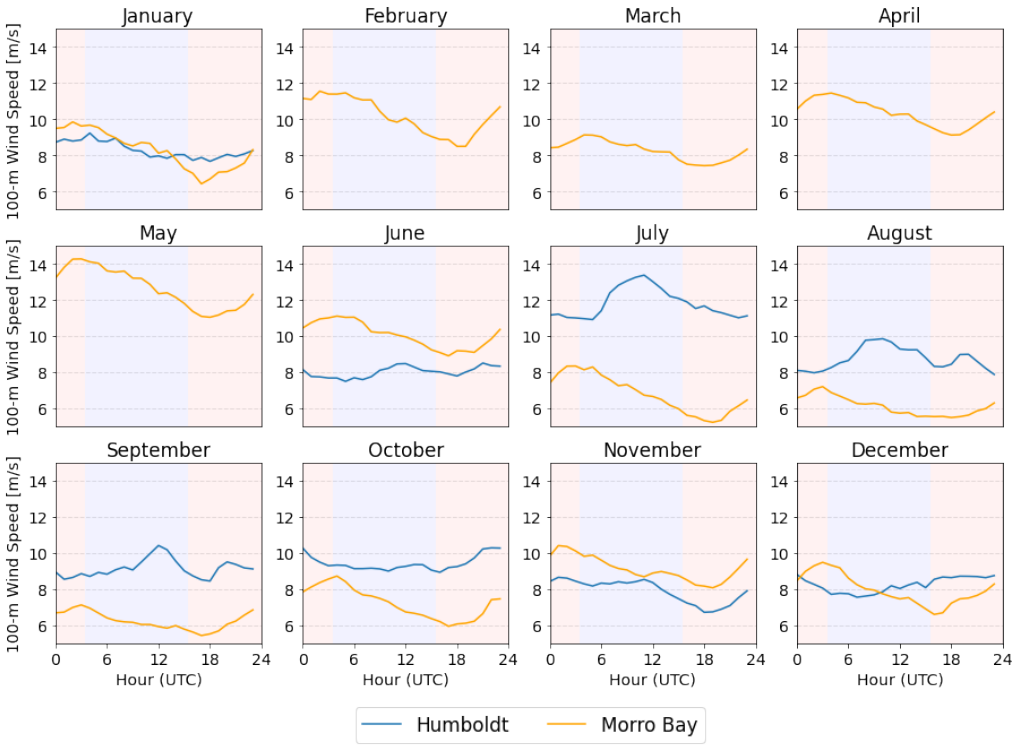

Figure 5.9: Average diurnal cycles of 100-meter AMSL wind speed by month in the Humboldt and Morro Bay WEAs. Red and blue background coloration refers to approximate local daytime and nighttime conditions. Data is missing for February through May for the Humboldt WEA.

The different diurnal wind speed patterns during July, August, and September in the Humboldt data may be due to larger shifts in the relative strengths of land and sea breezes across seasons. Without complete seasonal data, however, investigation of this phenomenon and its effects on offshore wind prediction is left for future study. The generally reliable diurnal patterns of hub-height wind speed across most parts of the year signify that the time of day is a relevant predictor variable for the ML algorithms.

5.3 Atmospheric Stability Calculations

5.3.1 Stability Classification

The dimensionless height parameter, z/L, was obtained from bulk Richardson number calculations and application of Monin-Obukhov stability theory. This methodology is hereafter referred to as the "profile method." z/L approaching zero indicates neutral atmospheric conditions, while positive and negative values indicate more stable or unstable, respectively.

Figure 5.10 shows the z/L probabilities from each 10-minute average datapoint over the entire deployment period. Both probability distributions exhibit a slight negative skew, indicating a relatively higher ratio of unstable conditions to stable conditions in both locations.

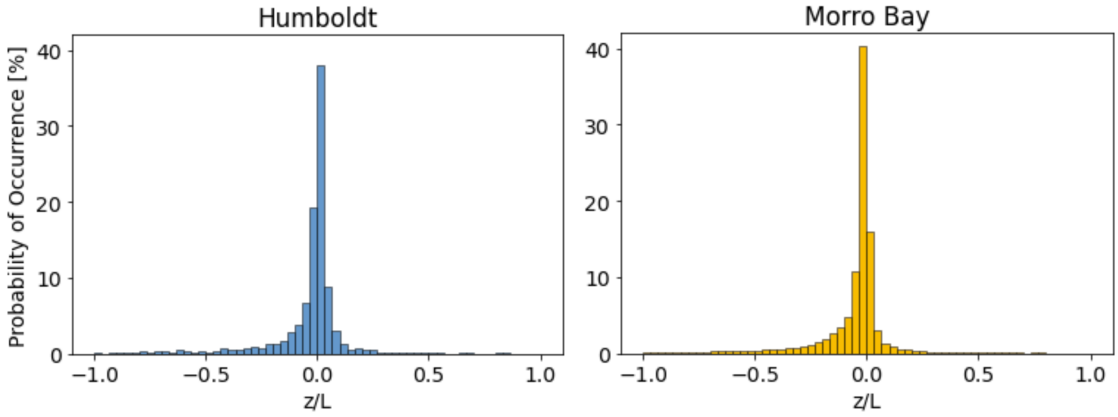

Figure 5.10: Probability distributions of calculated z/L using the "profile method". Calculations are based on all available, complete 10-minute data from October 2020 through January 2022.

Stability conditions can be labeled by respective Obukhov length L for each timepoint, as categorized in Table 4.3 in the Methods section. Percentages of data in each atmospheric

stability class are given in Table 5.1. Both the Humboldt and Morro Bay WEAs primarily experience unstable atmospheric conditions. The Morro Bay ratios are more meaningful as they represent a full year of near-complete data, while the Humboldt data may exert inherent bias toward summertime and autumnal conditions, considering the absence of data from February through May. These categorizations are important for identifying conditions which correlate with larger model uncertainty and error.

Study Location	**% Unstable**	**% Neutral**	**% Stable**
Humboldt	46.8	20.1	33.1
Morro Bay	59.4	24.1	16.5

Table 5.1: Percentage classification of atmospheric stability conditions in the Humboldt and Morro Bay WEAs. The classification scheme, based on the value of Obukhov length L, is detailed in Table 4.3.

5.3.2 Performance Against COARE 3.6 Algorithm

Calculations of z/L and atmospheric stability are cross-referenced with a more advanced algorithm to assess the validity of the approach. The more sophisticated method, known as the COARE 3.6 bulk algorithm, was performed by PNNL on hourly data from 10/01/2020 to 09/30/2021 to calculate z/L and published for public use. The COARE algorithm computes bulk air-sea fluxes and associated fluxes including wind stresses using an iterative approach with a large array of ambient physical parameters as inputs. Supplemental input data from outside the DOE LiDAR buoy project was sourced for application of the COARE algorithm to the Pacific OCS WEAs. The COARE bulk algorithm is described in further detail in Fairall et al. (2003) [144].

Direct comparisons between the calculated z/L of this study (obtained through the "profile method") and the COARE algorithm are shown in Figure 5.11. Comparison is limited to the hourly time points within the one-year date range of the available COARE algorithm data. There are relatively fewer data points for Humboldt due to the DOE buoy decommissioning for four months in early 2021.

For the z/L comparison, the Spearman correlation coefficient (ρ) is used instead of R^2 or normal linear regression since the data is highly concentrated around zero and non-normally distributed. The Spearman coefficient indicates how well a monotonic function can be applied to the relationship of the variables, ranging from -1 (perfect negative monotone function) to 1 (perfect positive monotone function). ρ in both cases is significant and p-values are zero, demonstrating good agreement between the two approaches.

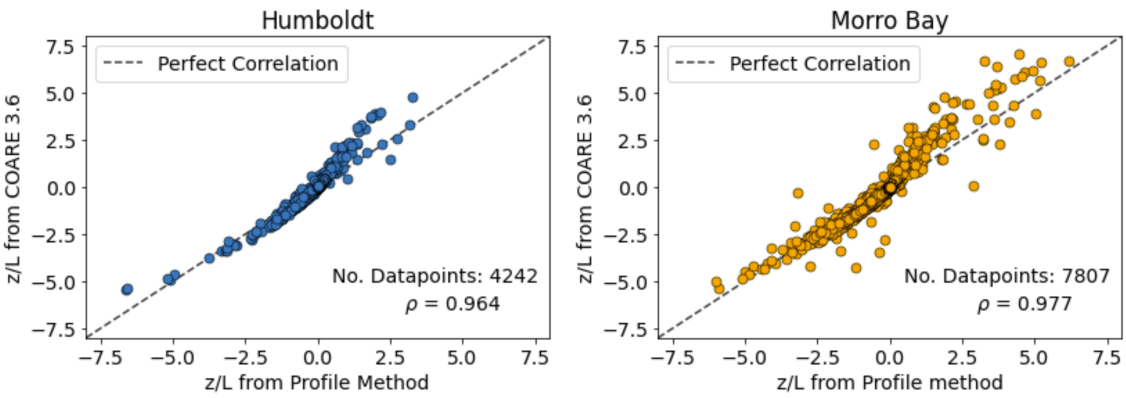

Figure 5.11: Comparison of z/L calculated using the profile method versus the COARE 3.6 algorithm run by PNNL. Only data with exact matching timestamps between the two resulting datasets are compared, which consists of hourly data from 10/01/2020 to 09/30/2021. A linear relationship with slope = 1 signifies identical results from the two methods. ρ is the Spearman correlation coefficient.

Discrepancies between the approaches are most concentrated under highly stable conditions, in which the profile method appears to slightly underestimate z/L compared to COARE 3.6. Assuming the COARE method provides a more reliable ground truth than the profile method, the profile method may have a slight bias of classifying stable conditions toward neutral.

Overlapping histograms of z/L in Figure 5.12 illustrate the small differences in atmospheric stability classifications between the two methods. The profile method leans slightly toward unstable classification compared to COARE and, as expected from Figure 5.11, slightly underestimates occurrences of stable conditions. The large number of inputs and equations involved in the COARE algorithm render it difficult to trace the main causes of these numerical discrepancies without more intensive analysis. It is likely that the handling of temperature data in COARE, which includes adjustments based on air temperature instrument height and corrections for interfacial SST, affects the calculated air-sea temperature difference and thus the predicted stability outcome. Overall, classification is in good agreement between the two methods, but the small skew of the profile method toward unstable conditions should be noted when analyzing further results.

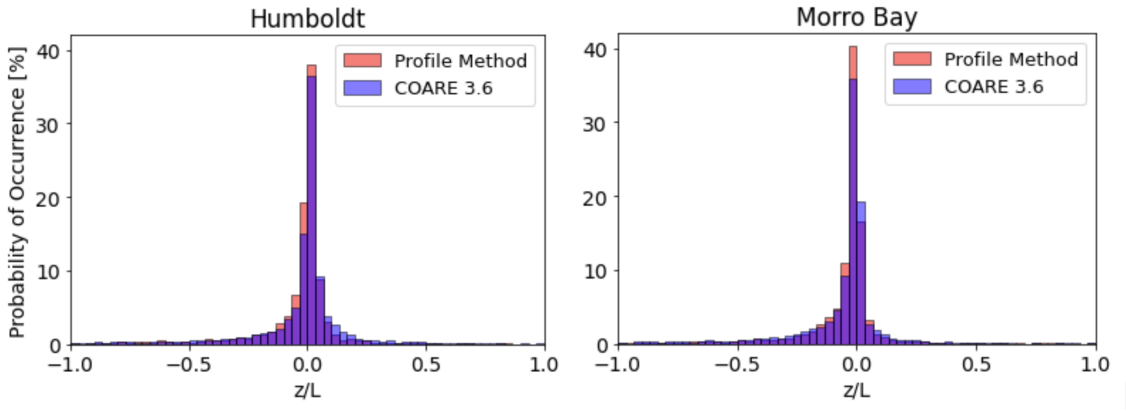

Figure 5.12: Probability distributions of z/L calculated using the profile method versus the COARE 3.6 algorithm run by PNNL. Only data with exact matching timestamps between the two resulting datasets are compared. Overlap is displayed by purple coloration.

	% Unstable		% Neutral		% Stable	
	Profile	**COARE**	**Profile**	**COARE**	**Profile**	**COARE**
Humboldt	37.2	33.1	20.5	21.9	42.3	45.0
Morro Bay	58.7	53.4	24.8	28.1	16.5	18.5

Table 5.2: Percentage classification of atmospheric stability conditions in the Humboldt and Morro Bay WEAs. Percentages differ slightly from above in Table 5.1 since only data with corresponding COARE calculations are considered. The classification scheme, based on the value of Obukhov length L, is detailed in Table 4.3.

5.4 Prediction Performance & Comparison

5.4.1 Variable Importance

Random forest provides a measure of variable importance (VI) by calculating the increase in mean squared error (MSE) when each input variable is randomly permuted. Figure 5.13 illustrates that the significant majority of prediction power is sourced from the 4-meter wind speed, exerting over 90% VI. Out of the three temperature parameters, air-sea temperature difference demonstrates more predictive power than the values of air temperature or SST alone.

Despite a clear diurnal signature of 100-meter wind speed for most months of the year, displayed in Figure 5.9, the combined sine and cosine components of the time of day exert the least predictive power out of all considered features. This result does not mean that time

of day is insignificant to predicting hub-height wind speed. It is likely that RF inherently reduces the importance of the time of day sine and cosine components (which encode unique times of the day only when considered together) since the regression trees split only one feature at a time. It is also possible that the 100-meter wind speed pattern correlated to the time of day may be more precisely encoded by the pattern of surface wind speed and/or air-sea temperature fluctuations.

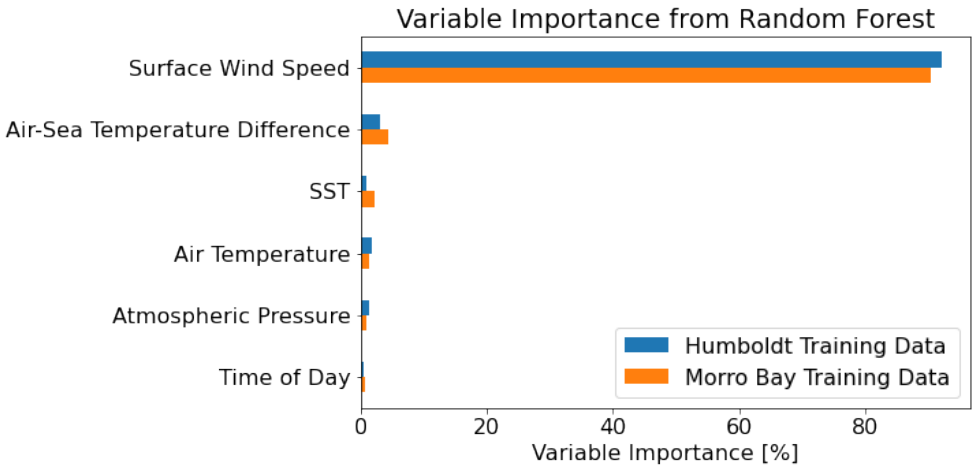

Figure 5.13: Variable importance (VI) of each input feature, calculated automatically via random forest regression.

The obtained VI implies that the S-C Log Law utilizes the two most important parameters (surface wind speed and air-sea temperature difference) for extrapolating hub-height wind speeds from surface measurements. This importance is relative to the seven predictors considered in this study and omits an exhaustive search of other possible predictor variables, such as significant wave height, surface current strength, and wind direction. Assessment and intercomparison of model performance will determine the sufficiency of using only these seven features to obtain accurate predictions of 100-meter AMSL wind speeds.

5.4.2 ML Prediction Samples

Figure 5.14 visualizes samples of the three ML algorithms' performance on the same subset of testing data. Predictions are shown in separate graphs to improve visual clarity. Across all three methods' predictions, the hourly fluctuations and multi-day trends of hub-height wind speed are well-represented. The relative accuracy of each model over these dates is not

easily inferrable due to the varying lengths and instances of prediction inaccuracies between the three methods.

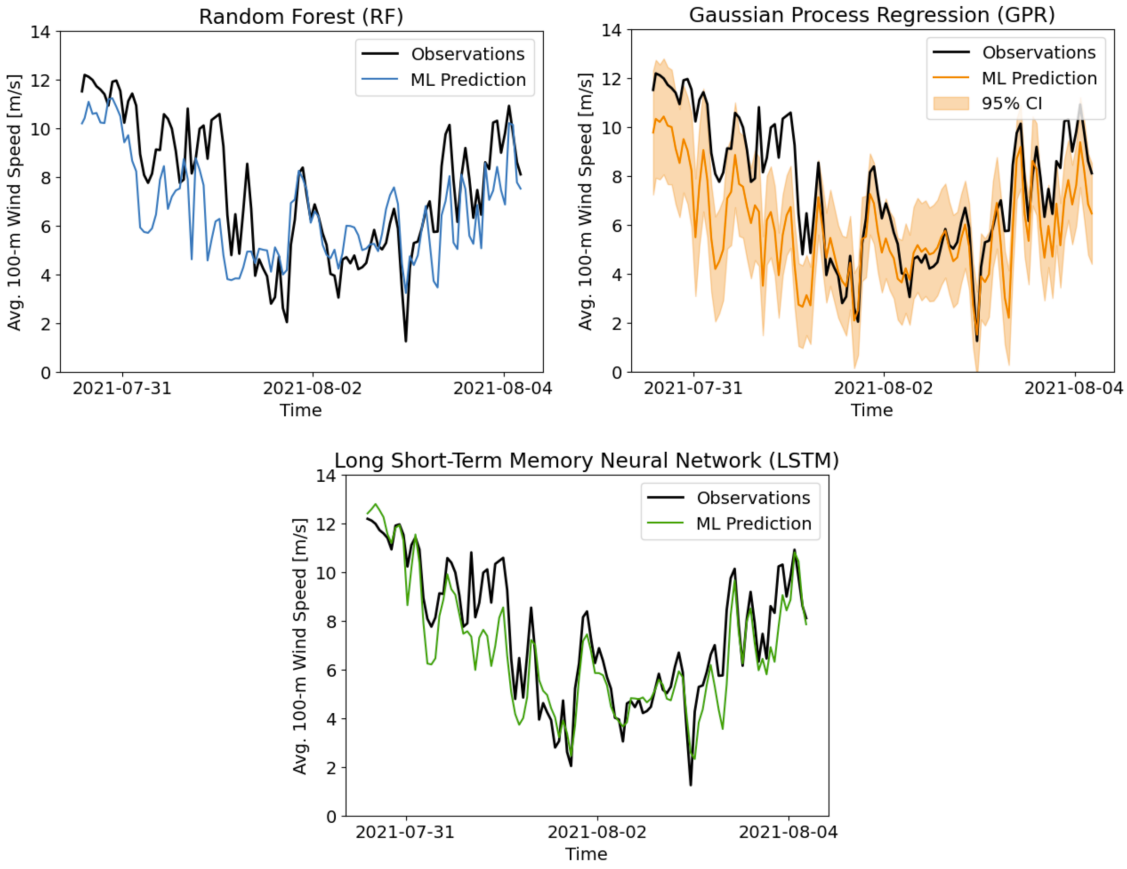

Figure 5.14: Samples of ML predictions of 100-meter AMSL wind speeds in the Humboldt WEA from 2021/07/30 - 2021/08/04. Algorithm training was conducted on roughly one month of preceding data from the Morro Bay WEA. The confidence interval computed by GPR is shaded around the mean prediction function.

The 95% confidence interval (CI) calculated over the underlying functions of GPR is displayed in the second graph of Figure 5.14. The CI narrows around higher-confidence predictions, but is always nonzero due to the multitude of possible underlying functions and the inclusion of a noise kernel. The reliability of the Gaussian process regression is highly dependent on the selection of the covariance function, which must be specifically tuned to the nature of patterns in the target variable.

The first half of the displayed data in the GPR sample, where some of the observational data falls outside of the confidence interval, highlights important considerations to make when using GPR. The likely cause of the large deviation from observational data is a model limitation around capturing longer-term temporal trends. This is inferred from the superior performance of LSTM in this particular time span, whose main advantage over the other algorithms is the "memory" component for long-term chronological patterns. It is possible that the performance of RF and GPR in this instance could be improved by longer training time, or for GPR specifically, more trial-and-error fine-tuning of the kernel.

The case-specific needs presented for RF and GPR prediction improvement decrease confidence in the spatial extrapolation capabilities of both methods for wind speed modeling. However, a relevant consideration is the vast spatial extrapolation between the training and testing sites performed in this study, which is 631 km (341 miles) between the Humboldt and Morro Bay WEAs. It is probable that model performance would improve using a relatively shorter distance between training and testing sites, which would likely be the case in practical offshore wind energy applications of these models. Further studies testing these models across shorter distances of the Pacific OCS would be helpful for understanding the impact of offshore spatial extrapolation extent on prediction accuracy, but this is currently not possible with the limited available data.

5.4.3 ML Performance vs. Stability-Corrected Logarithmic Law

A comparison of the ML methods against the S-C Log Law on a slice of hourly Morro Bay test data from 2020/12/25 - 2020/12/27 is displayed in Figure 5.15. The S-C Log Law demonstrates similar performance to the ML techniques, and in some instances provides the most accurate predictions to the LiDAR data. Most similar in performance to the S-C Log Law are LSTM and GPR, which generally exhibit inaccuracies in conjunction with those of the S-C Log Law.

Figure 5.15 illustrates one of the speculated shortcomings of employing RF for wind prediction, which is smoothed predictions of stochastic wind speed extremes. Though the RF predictions respond to patternal wind speed changes, the mean absolute error is relatively highest for the majority of time points. This limitation can be attributed to the inherent process of regression tree averaging in RF, which makes the method robust to outlier data, but also less capable of extrapolating anomalous wind events unforeseen in the training data.

The discussion thus far has focused only on model performance over small portions of test data. Generalized model performance is assessed through averaging the performance of each model across the six testing datasets. Figure 5.16 displays the average hourly 100-meter wind speeds across the diurnal cycle, with each prediction series normalized to the respective mean. The generalized offshore wind speed trend is again observable, with reductions induced by

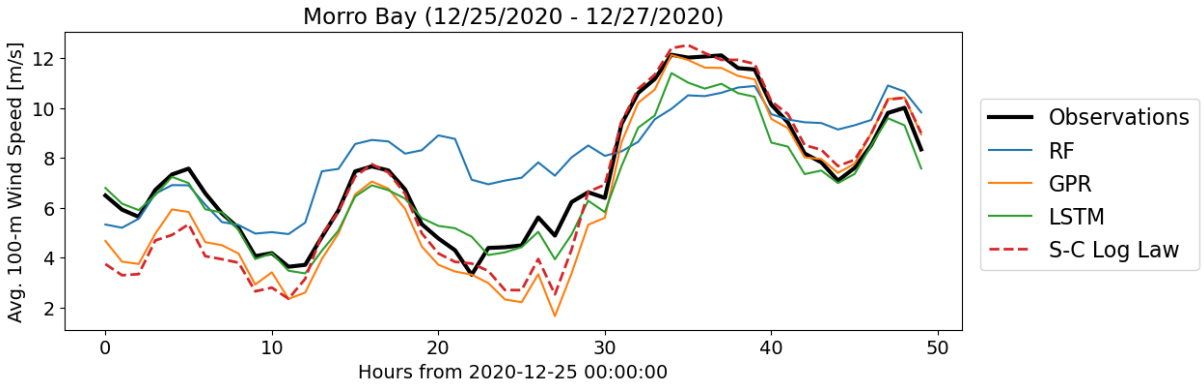

Figure 5.15: Sample comparison of ML predictions of 100-meter AMSL wind speeds versus values calculated from the stability-corrected logarithmic law. The testing data subset is from the Morro Bay WEA from 12/25/2020 to 12/27/2020. ML algorithm training was conducted on roughly one month of preceding data from the Humboldt WEA.

the more stable nighttime conditions beginning around 5 UTC and increased speeds brought about by convective daytime conditions beginning around 18 UTC.

The view of normalized wind speed per hour of the day elucidates some important aspects of model performance. RF demonstrates significant error across all parts of the diurnal cycle. In both locations, the RF model tends to underestimate hub-height wind speeds in nighttime conditions and overestimates speeds in the daytime. Besides some distinctions in bias at certain parts of the diurnal cycle, individual model performance between the two locations appears to be relatively similar. This result is promising in that it implies model performance is not hugely dependent on the offshore training location, but rather the inherent proficiency and applicability of individual model architecture.

The key performance metrics for each model are displayed by the bar graphs in Figure 5.17. The bars correspond to the "MOR-HUM", "HUM-MOR", and S-C Log Law columns in Table 5.3. Table 5.3 also shows the average runtime of training and testing each algorithm over the six datasets. RF has the longest average runtime due to the computationally-intensive process of training data subsetting and regression tree fitting. Though the number and depth of trees could be manually lowered to improve the runtime, prediction accuracy would likely degrade as a result. GPR delivers a moderate average runtime, likely on partial account of using only seven predictor variables and conducting dimensionality reduction on these variables with PCA. Training and testing of LSTM over the six datasets takes the shortest amount of time, averaging to just over two minutes. Computations were run on a 4-core processor with 1.5 GHz base speed and 32 GB RAM. The speed of the algorithms

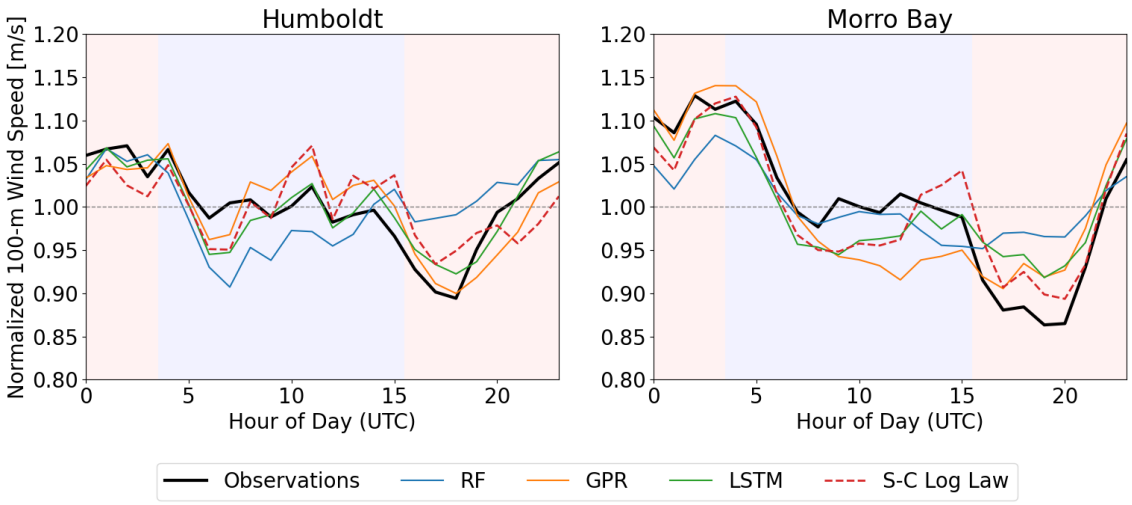

Figure 5.16: Comparison of observed and predicted diurnal averages of normalized 100-meter AMSL wind speed. The ML prediction data are only from runs using different train-test locations (i.e. the Humboldt predictions are made from algorithm training on Morro Bay data, and vice versa). All averages are drawn from the times corresponding to the six sets of testing data, as detailed in Table 4.2. Red and blue background coloration indicates approximate local daytime and nighttime conditions.

could be augmented by parallelizing computations or running a multinodal application.

Inferences made about RF performance from Figures 5.14, 5.15, and 5.16 are substantiated by the calculated error metrics. cRMSE, which represents unbiased model variation around the mean, is highest for RF in both locations and descends in the order of GPR, LSTM, and S-C Log Law. The previously discussed subsamples of testing performance demonstrated the relatively large inaccuracies of RF and its insufficiency for deriving stochastic wind speed variations. The R^2 values for RF are correspondingly the lowest, indicating the poorest relative fit to the observed data, and values for EMD are the highest, evidencing a large difference between the predicted and actual probability distributions of 100-meter AMSL wind speed. Moreover, most of these RF error metrics do not substantially improve even when training and testing are conducted in the same location. Based on its many aspects of large prediction error, RF is deduced to be the least effective method for extrapolating Pacific OCS offshore hub-height wind speeds out of the techniques considered.

Relative to the four methods included in the study, GPR demonstrates average performance across the board, though its error metrics are much more similar in magnitude to those of LSTM and S-C Log Law than RF. Though the MSE of GPR is larger than that of LSTM and S-C Log Law, this facet of prediction error is counterbalanced by comparatively

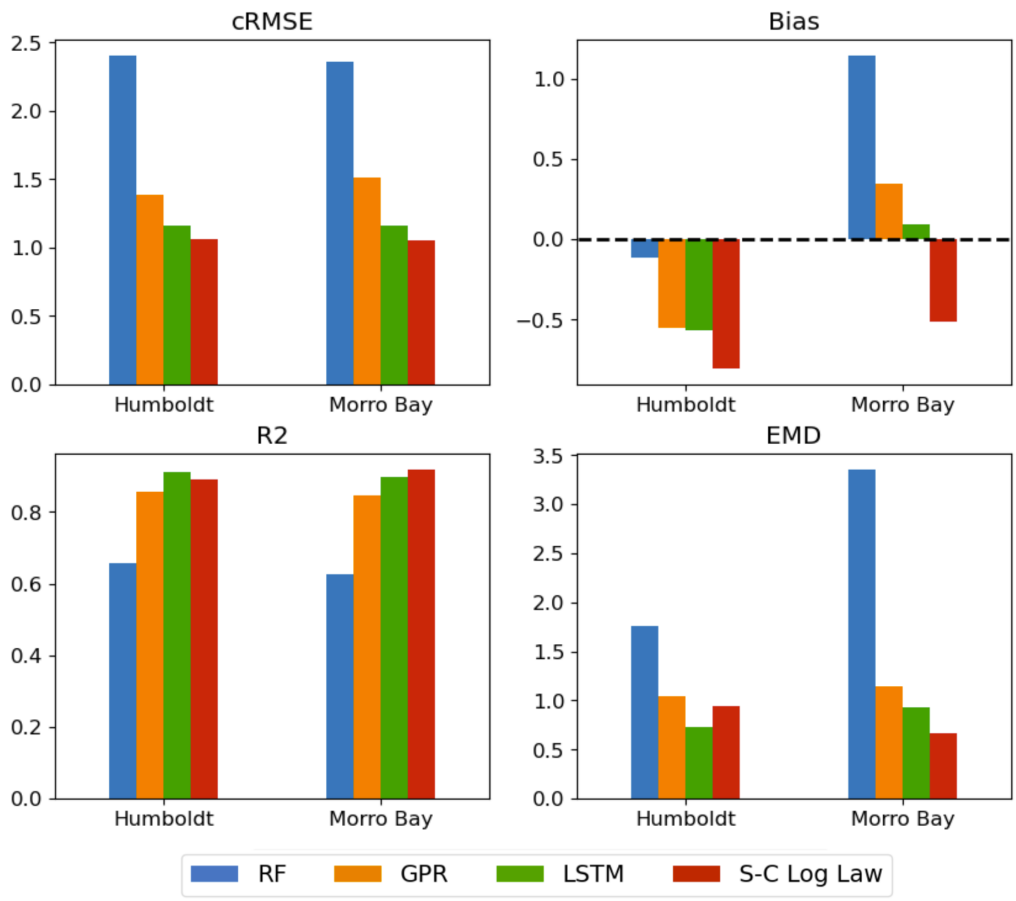

Figure 5.17: Site-specific error metrics of the ML methods' and the S-C Log Law's predictions of 100-meter AMSL wind speed. The ML prediction data used are only from runs using different train-test locations.

similar or lower prediction bias. GPR also demonstrates improvement in almost all performance metrics when trained on its same testing location and obtains an RMSE within ±0.1 m/s of the S-C Log Law predictions. GPR is considered a promising method for Pacific OCS wind prediction for offshore energy purposes, especially notable for its inclusion of confidence intervals for every timestep prediction.

LSTM is the most similar in cRMSE to S-C Log Law and demonstrates smaller magnitudes of bias than the S-C Log Law. LSTM attains extremely similar performance metrics with S-C Log Law in terms of cRMSE, R^2, and EMD. The results of LSTM are impressive when regarding the fact that training and testing took place over 631 kilometers of extrapolation distance. When trained and tested in the same location, LSTM produces superior

Test Location Target Variable	Humboldt WEA 100-meter AMSL Wind Speed [m/s]						
Method	RF		GPR		LSTM		S-C Log Law
TRAIN - TEST	MOR-HUM	HUM-HUM	MOR-HUM	HUM-HUM	MOR-HUM	HUM-HUM	HUM-HUM
RMSE	2.72	3.03	1.71	1.33	1.32	1.19	1.43
Bias	-0.12	1.12	-0.55	0.03	-0.57	-0.20	-0.81
cRMSE	2.40	2.36	1.39	1.26	1.16	1.08	1.06
R^2	0.66	0.63	0.86	0.85	0.91	0.92	0.89
EMD	1.76	2.09	1.04	0.58	0.73	0.56	0.94
Total Runtime [min:sec]	6:05	6:11	3:22	3:10	2:02	2:06	-

Test Location Target Variable	Morro Bay WEA 100-meter AMSL Wind Speed [m/s]						
Method	RF		GPR		LSTM		S-C Log Law
TRAIN - TEST	HUM-MOR	MOR-MOR	HUM-MOR	MOR-MOR	HUM-MOR	MOR-MOR	MOR-MOR
RMSE	4.05	2.75	1.92	1.33	1.44	0.89	1.23
Bias	1.14	-0.87	0.35	0.07	0.09	0.00	-0.52
cRMSE	2.36	2.13	1.52	1.19	1.16	0.85	1.06
R^2	0.62	0.67	0.85	0.87	0.90	0.95	0.92
EMD	3.35	1.89	1.14	0.64	0.93	0.42	0.67
Total Runtime [min:sec]	5:50	6:00	3:23	3:14	2:04	2:05	-

Table 5.3: Performance metrics of all 100-meter AMSL wind speed prediction methods. Metrics compared between methods in Figure 5.17 are highlighted in yellow. "MOR-HUM" denotes training on Morro Bay data and testing on Humboldt data, and vice versa for "HUM-MOR". The tables additionally list the performance metrics of the training and testing runs performed on the same location, denoted as "HUM-HUM" and "MOR-MOR"

performance over the S-C Log Law for almost all error metrics.

The LSTM results illuminate the importance of accounting for chronological wind sequence patterns and stochastic wind events within intelligent learning approaches to wind prediction. The performance of LSTM also highlights that additional training data or input variables are not necessarily imperative for attaining ML wind prediction results similar or better than those of the S-C Log Law, which are suppositions that could be drawn if analysis were limited to RF and GPR alone. LSTM demonstrates overall powerful performance for Pacific OCS hub-height wind prediction, and it is the best of all considered methods for local wind speed extrapolations.

5.4.4 ML Train-Test Data Considerations

The data from Table 5.3 indicate that training and testing in the same location generally improves ML prediction accuracy. A sample of this ML performance improvement is visualized in Figure 5.18. Average RMSE from all six datasets is displayed and correspond to Table 5.3, but only the slice of December testing data is shown for demonstration purposes. Training in the same location as test data (i.e. no horizontal spatial extrapolation) improves model RMSE in all cases except for the RF application to Humboldt data. Average RMSE decreases by 0.34 m/s for LSTM and by 0.49 m/s for GPR when the train-test location is the same. Without validation data available at a location between the Humboldt and Morro Bay sites, there remains uncertainty as to the rate of model performance degradation with distance from the training location.

In parallel to predictions made using different train-test locations, LSTM produces the most accurate results out of the three ML algorithms in the same train-test site case. These conditions produce RMSE of 1.19 and 0.89 for Humboldt and Morro Bay respectively. These are the lowest RMSE obtained from any extrapolation method used under the standard procedures of this study, including the S-C Log Law RMSE values of 1.43 and 1.23 for Humboldt and Morro Bay respectively. It is worth noting that the lowest RMSE obtained are still nontrivial; an RMSE of 1 m/s represents moderate error by wind energy standards, and thus it is not valid to treat these LSTM predictions as a validation metric for other techniques. The main takeaway from these results is that the LSTM architecture has proven to be capable of making more accurate offshore hub-height wind speed predictions than the popular and long-established S-C Log Law technique, and in this study demonstrates that only easily-obtainable surface variables are needed as inputs.

A constraint presented by the DOE buoy data source is the lack of continuous and complete LiDAR and surface variable data. The irregular data availability caused by short and long-term instrument failures inhibits analysis of each models' capabilities of incorporating interseasonal and interannual trends. The longest piece of continuous data in both locations

runs from the beginning of October to near the end of December. These three months of data provide one opportunity to test the effect of longer training data length. Figure 5.19 shows comparisons of approximately one month of training data (as is the case for all train-test datasets configured for this study) versus the approximate three months of training data, where both are tested on the same dates at the end of December. The RMSE listed on each graph are slightly different from those in Table 5.3 as they correspond only to the December test data.

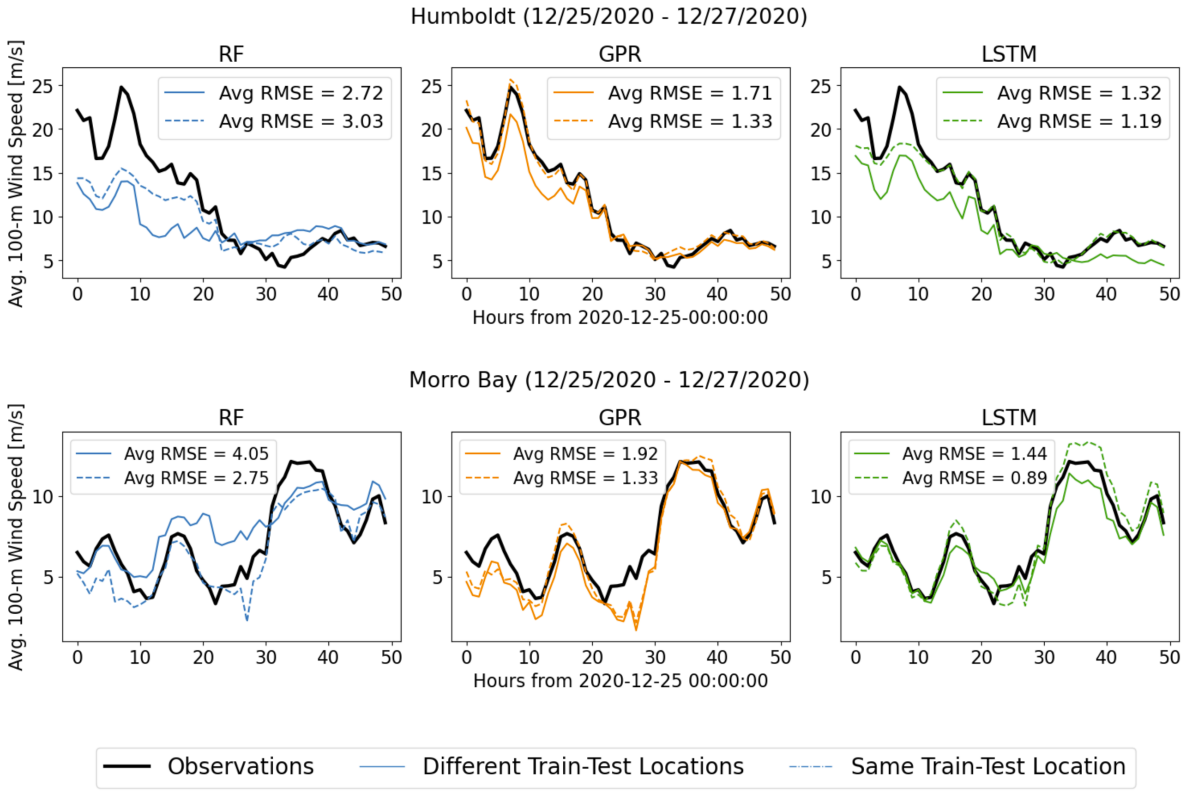

Figure 5.18: Differences in ML prediction performance of 100-meter AMSL wind speeds when using different train-test locations versus same train-test locations. "Avg RMSE" refers to the RMSE over all six testing datasets, not just the RMSE of the prediction subset shown.

In all cases except for the GPR-Humboldt predictions, longer training time effectuates moderate increases in predictive accuracy. Though it was noted earlier that additional training data is not imperative to obtaining results similar to the S-C Log Law, these results

demonstrate promise of LSTM's potential to significantly outperform the S-C Log Law given longer sequences of continuous training data.

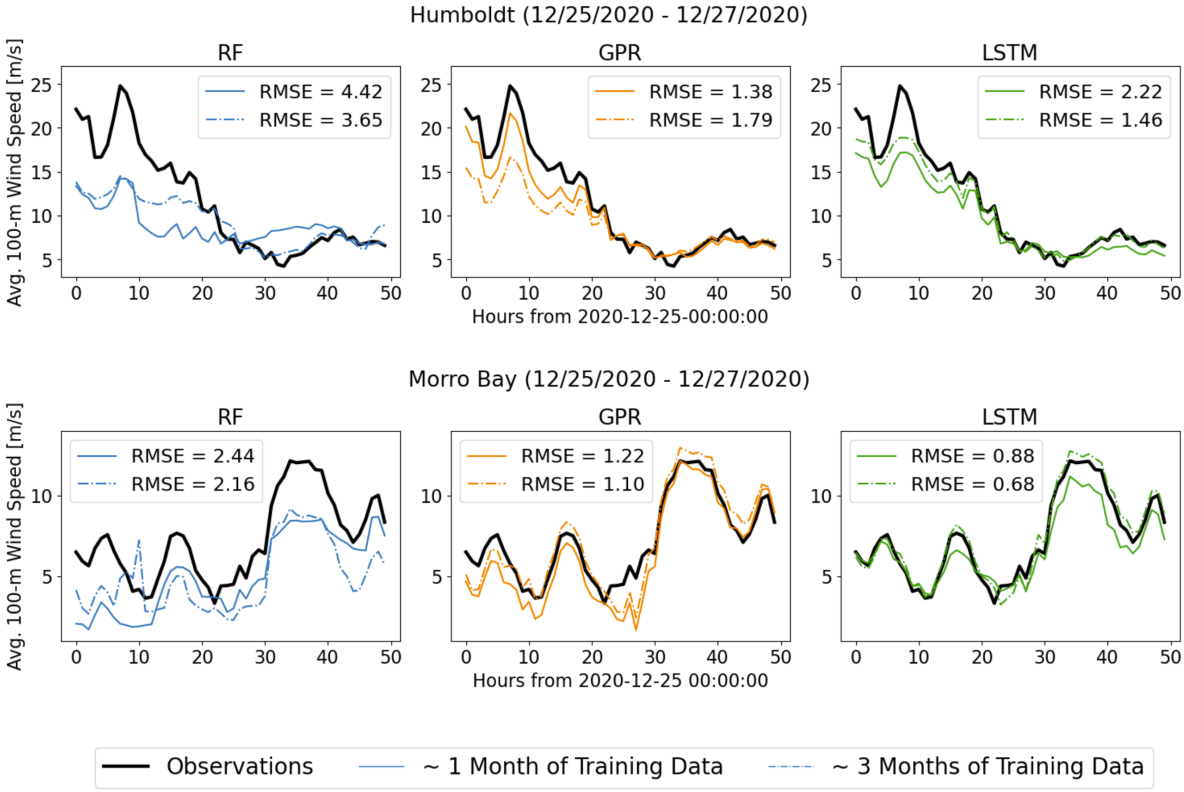

Figure 5.19: Differences in ML prediction performance of 100-meter AMSL wind speeds when using approximately one month of preceding training data versus approximately three months of preceding training data. The ML prediction data used are only from runs using different train-test locations.

5.4.5 Prediction Dependence on Atmospheric Stability

In Figure 5.20, model error metrics are categorized by the atmospheric stability condition during which each prediction was made. Across all prediction methods, stable conditions tend to produce the greatest prediction cRMSE and induce more model bias, especially evident in the Morro Bay test data. A strong negative bias for stable conditions in Humboldt and positive biases across all ML algorithms for stable Morro Bay conditions suggests that

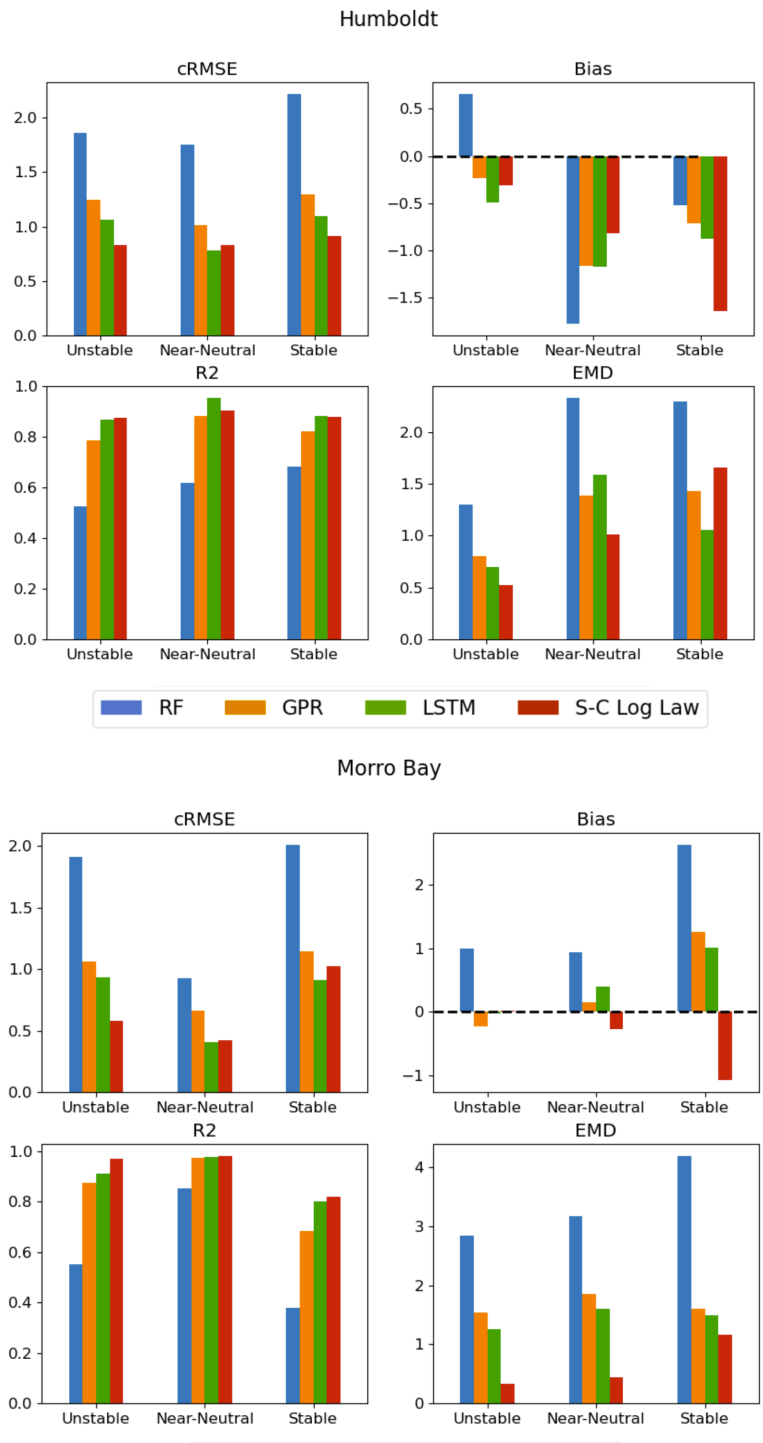

Figure 5.20: Error metrics of the ML methods' and the S-C Log Law's predictions of 100-meter AMSL wind speed, distinguished by the background atmospheric stability condition. The ML prediction data shown are only from runs using different train-test locations.

shear profiles between the two locations may diverge under a stable regime. This inference is corroborated by the significant reduction of bias across all methods in unstable conditions.

An important qualification of the obtained results is the bias in underlying stability categorization, which was drawn from comparisons between the profile method and COARE algorithm. Comparisons between the profile method and COARE algorithm demonstrated a slight bias toward neutral and unstable categorization using the profile method, assuming COARE data represents ground truth. It is possible that the statistical differences between unstable, near-neutral, and stable conditions would not be as stark if this bias were corrected. However, the bias was shown to be small, as only 1-5% of data categorization differs from the COARE algorithm results, so the trends displayed in Figure 5.20 are considered to be substantial.

The error magnitudes in RF and GPR lead to similar conclusions as before about the fidelity of each model's predictions. RF demonstrates the lowest R^2 and highest cRMSE and EMD across all regimes and thus remains the least robust method tested. GPR again performs between RF and LSTM across most error metrics, with exceptions of similar or lower bias compared to LSTM under unstable and near-neutral conditions. GPR is considered especially useful for its confidence estimates but is not recommended as a standalone model for Pacific OCS hub-height wind prediction.

The differences between Humboldt and Morro Bay metrics in Figure 5.20 present ambiguity in determining which method proves most robust to atmospheric stability conditions. LSTM and S-C Log Law both exhibit their lowest respective cRMSE under near-neutral conditions, in which both achieve values of less than 1 m/s in Humboldt and less than 0.5 m/s in Morro Bay. The cRMSE increases more for LSTM compared to S-C Log Law in unstable conditions. Comparative performance in terms of bias is inconclusive, as values do not show clear trends across the testing sites. S-C Log Law produces lower EMD values than LSTM (i.e. more similar probability distribution to the actual wind speeds) except for stable cases in Humboldt. In gross terms of performance metrics, the S-C Log Law demonstrates a slight advantage over LSTM, but differences are not significant or consistent enough to substantiate a clear recommendation.

5.4.6 Turbulence Intensity

Intelligent learning methods for wind prediction provide two inherent advantages over classic physical law extrapolation techniques. Firstly, many of these intelligent methods are able to provide short-term forecasts. In other words, once trained, learned patterns can be extrapolated forward in time without additional inputs of training data. Forecasting is outside of the current scope of this study, but is a pertinent topic of future research for Pacific OCS wind energy purposes. Secondly, these ML methods are not restricted to pre-

dicting vertically-extrapolated wind speeds, as are the S-C Log Law and power law. Any environmental feature of interest can be predicted, with accuracy dependent on the skill of the model, the selection of predictor variables, and the natural correlation or stochasticity of the feature compared to the selected predictors.

For example, another important aspect of the wind resource to wind energy developers is turbulence intensity (TI), which represents the strength of wind velocity fluctuations. TI is important for assessing potential mechanical fatigue on wind turbines. TI of 100-meter winds is gathered from the LiDAR data by calculating the ratio of standard deviation of fluctuating wind velocity to the mean wind velocity. Calculation of TI is shown in Equation 5.2, where $u^{'}$ is the root-mean-square of the turbulent velocity fluctuations and U is the mean velocity.

$$TI = \frac{\sqrt{\frac{1}{3}({u'_x}^2 + {u'_y}^2 + {u'_z}^2)}}{\sqrt{U_x^2 + U_y^2 + U_z^2}} \tag{5.2}$$

TI at 100 meters AMSL is predicted using the same inputs and ML model configurations as for 100-meter wind speed prediction. Results for normalized TI are displayed in Figure 5.21, and performance metrics are listed in Table 5.4. Normalized results are shown instead of individual predictions to demonstrate whether the models capture the trend in TI over the diurnal cycle; further input and model tuning would be appropriate before conducting more detailed assessment of prediction results. The procedures used for modeling TI in this case were tuned for prediction of wind speed and were not tested for relevance to patterns in TI. Thus, results are expected to have more prediction error than those for 100-meter wind speeds. Note that TI is a percentage and the range of values (between 0 and 1) is more limited than that of wind speed.

Patterns in TI are predicted to varying degrees of success between the ML algorithms and test locations. However, based on the generally low R^2 values, none of the ML methods attain a significant degree of accuracy. This result is expected because of the lack of feature correlation analysis for TI and demonstrates the importance of examining a wider range of possibly relevant surface variable predictors.

Despite low prediction accuracy, especially to that of the Morro Bay test data trained on Humboldt data, similar trends to the 100-meter wind speed predictions emerge. LSTM predictions achieve the lowest RMSE and cRMSE for both cases of same train-test location and different train-test location. Using the same train-test location again lowers the majority of prediction error metrics across all models. The smallest RMSE value attained is from employing LSTM on the same train-test location data, for which it produces an RMSE of 6% in both sites. Overall prediction accuracy of the three ML algorithms, akin to the

wind speed predictions, can be hierarchized as LSTM, GPR, and RF from highest to lowest proficiency. With further tuning of the models and input features, it is possible that one or more of these models could be capable of achieving accurate real-time and forecasted TI estimates at turbine heights. This exploration of TI prediction emphasizes a primary advantage of using intelligent learning methods: they can be adaptable toward estimating parameters not otherwise obtainable by conventional physical law approaches.

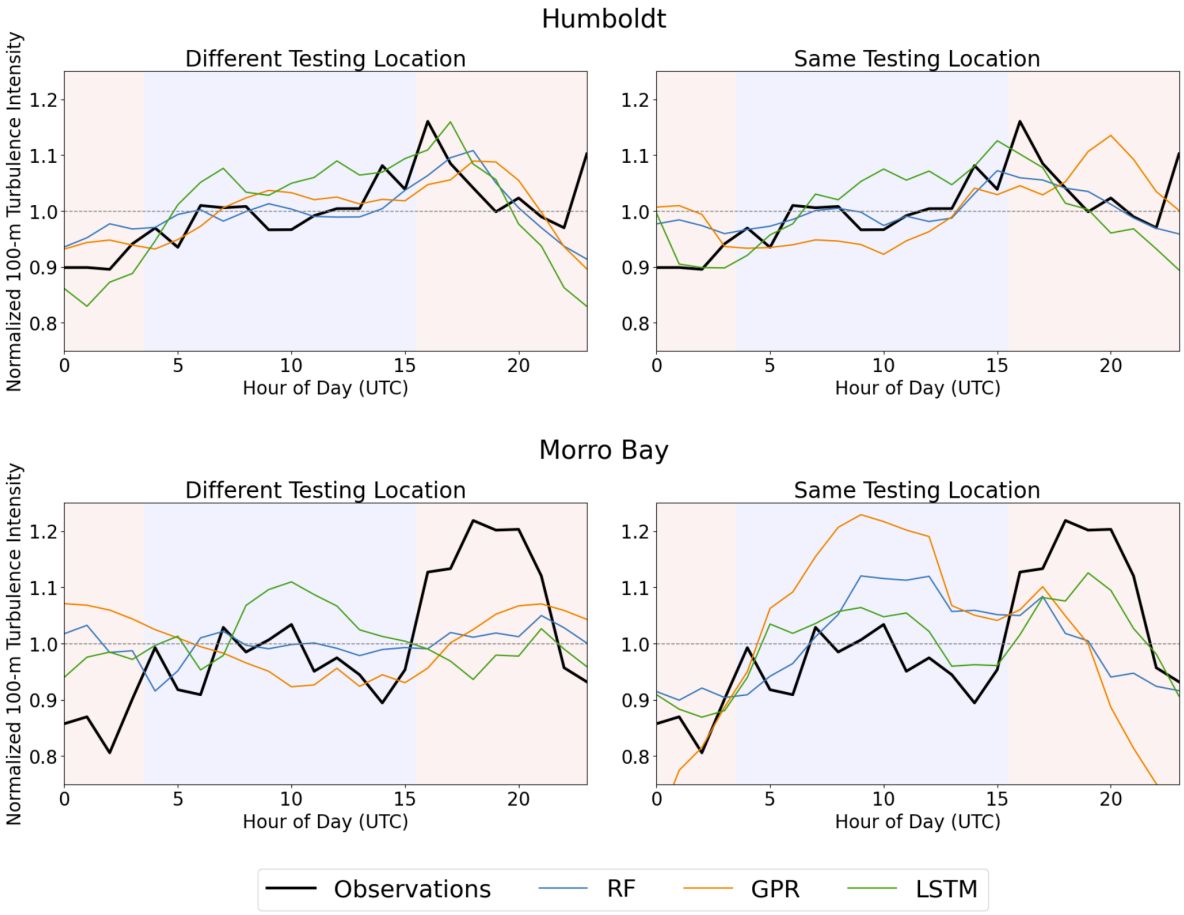

Figure 5.21: Comparison of observed and predicted diurnal averages of normalized 100-meter AMSL turbulence intensity (TI). All averages are drawn from the times corresponding to the six sets of testing data, as detailed in Table 4.3. Red and blue background coloration indicates approximate daytime and nighttime conditions.

Test Location	Humboldt WEA					
Target Variable	Turbulence Intensity (TI)					
Method	RF		GPR		LSTM	
TRAIN - TEST	MOR-HUM	HUM-HUM	MOR-HUM	HUM-HUM	MOR-HUM	HUM-HUM
RMSE	0.09	0.09	0.09	0.08	0.08	0.06
Bias	-0.01	0.01	-0.01	-0.02	0.00	-0.02
cRMSE	0.08	0.08	0.09	0.08	0.07	0.06
R2	0.17	0.15	0.16	0.22	0.43	0.48
EMD	0.06	0.05	0.05	0.03	0.05	0.02

Test Location	Morro Bay WEA					
Target Variable	Turbulence Intensity (TI)					
Method	RF		GPR		LSTM	
TRAIN - TEST	HUM-MOR	MOR-MOR	HUM-MOR	MOR-MOR	HUM-MOR	MOR-MOR
RMSE	0.13	0.08	0.10	0.08	0.09	0.06
Bias	0.05	0.01	0.02	0.00	-0.02	0.00
cRMSE	0.10	0.08	0.08	0.08	0.07	0.06
R2	0.11	0.20	0.18	0.30	0.44	0.62
EMD	0.09	0.04	0.07	0.04	0.05	0.02

Table 5.4: Performance metrics of all 100-meter AMSL turbulence intensity prediction methods. "MOR-HUM" denotes training on Morro Bay data and testing on Humboldt data, and vice versa for "HUM-MOR". The tables additionally list the performance metrics of the training and testing runs performed on the same location, denoted as "HUM-HUM" and "MOR-MOR".

Chapter 6

Conclusion

As the offshore wind industry continues to gain momentum in the United States, it is imperative to test new methods for turbine-height wind characterization that are specific to the unique offshore environment and fit the needs of energy developers. Validation of predictive models for the U.S. West Coast offshore wind arena has been enabled only recently by the vertically-profiled wind data collected by the two DOE/PNNL scientific buoys. Physical law approaches and NWP models are presently the standard techniques for wind prediction in this region, but novel intelligent learning methods offer potential enhancements in site-specific adaptability, prediction accuracy, and user accessibility. These improvements are highly valuable to offshore energy developers, who require accurate, efficient, and flexible wind forecasting methods to provide dependable energy without incurring extraneous costs.

The three ML methods of random forest, Gaussian process regression, and Long Short-Term Memory neural network were compared against physical law calculations of the stability-corrected logarithmic law to assess predictive accuracy of hub-height (100 meters AMSL) wind speeds in two potential offshore wind sites. The input predictor variables for ML training were constrained to measurements taken at the surface in each offshore location. Comparisons against the stability-corrected logarithmic law are primarily focused on ML predictions made at the opposite location of where the algorithm was trained (i.e. algorithm training on Morro Bay WEA data and testing at the Humboldt WEA, and vice versa). Using different training and testing locations ensures that model predictions are not overfit to the training location. However, in this case, the distance between the two locations (631 km) is likely much larger than the practical spatial extrapolation needed for model applications in the wind energy industry. The model predictions using different train-test locations in this study can thus be contextualized as a relative "lower limit" of ML performance for Pacific OCS offshore wind prediction.

Training and testing was conducted over six datasets, each containing approximately one month of 10-minute meteorological and oceanic data. Predictor variables included in these datasets were analyzed for relevance to hub-height wind speed patterns and decided to be 4-

meter AMSL wind speed, air temperature, SST, air-sea temperature difference, atmospheric pressure, and time of day. This study omits an exhaustive search of additional variables that may improve ML predictive accuracy, such as turbulence [145] and wave characteristics [102].

Performance metrics for each method's 100-meter AMSL wind speed predictions were obtained from average performance across the six train-test datasets. The best performances, generally evaluated from assessing the error metrics of bias, cRMSE, R^2, and EMD, were exhibited by LSTM and the stability-corrected logarithmic law. Random forest produced the least accurate results and demonstrated notable insufficiencies in predicting stochastic or extreme wind speed events, likely due to its inherent prediction smoothing from regression tree averaging. GPR demonstrated slightly lower accuracy than LSTM and the stability-corrected logarithmic law across most of the error metrics, though predictions were much more similar to these two methods than those of random forest. Across all methods, stable atmospheric conditions tended to induce greater magnitudes of cRMSE and bias compared to neutral and unstable conditions. This result may be due to atypical shifts in the vertical wind shear profile influenced by stable stratification in the surface ABL.

It is important to note an advantageous aspect of GPR over the other methods, which is an empirical confidence interval calculated over all predictions. The wind energy industry often requires uncertainty estimates to forecast different energy production scenarios. Thus, GPR is considered an especially promising method for offshore wind prediction but requires further study and model tuning.

It is worth reiterating that the comparative ML performances discussed above are only from predictions made over a 631 km train-test extrapolation distance. When trained and tested in the same location, LSTM surpasses the performance of the stability-corrected logarithmic law for every error metric, and GPR predictions become relatively similar to the stability-corrected logarithmic law in terms of general accuracy. Using a three-month-long training time instead of one month similarly augments prediction accuracy for almost all ML testing cases. LSTM is evidenced to be the most accurate and adaptable method for offshore wind prediction out of the techniques considered. Overall, LSTM and GPR present similar or better wind prediction performance than the stability-corrected logarithmic law, which is a widely-accepted physical law wind extrapolation technique in use by wind researchers and energy developers. However, these methods still present low-to-moderate prediction error by wind energy standards, and thus should not be used as validation data for other methods.

ML offers greater flexibility for wind characterization than conventional physical law methods, as it can be used for predicting other wind parameters and making short-term forecasts. To demonstrate the former advantage, the 100-meter AMSL wind speed target variable was replaced with turbulence intensity (TI) at the same height. LSTM again demonstrated the highest prediction accuracy and RF the lowest, though overall prediction

accuracy was poorer than predictions made for hub-height wind speeds. It is probable that the TI prediction results could be significantly improved if model hyperparameters were tuned to best capture TI patterns and if more predictor variables with specific relevance to TI patterns were included, but these methodological refinements are left to future study.

Many possibilities exist for optimization of ML offshore wind speed prediction performance in this specific study. To amend the disparities in prediction performance across different atmospheric stability regimes, relevant predictor variables such as the Obukhov length L and surface TI could be included and compared against the original model runs. Other potentially relevant surface variables, including wind direction and relative humidity, were excluded from this study due to data irregularities or infidelity. Given continued data availability, more surface variables should be included in the train-test datasets and assessed for importance to model predictions. Additionally, though many ML hyperparameters were optimized through grid search tuning, some other hyperparameter values and functions were default selections and not tested for best suitability to wind speed prediction. A more thorough study of optimal ML hyperparameters in the context of offshore wind characterization would likely enhance the quality of the wind speed predictions.

The data gathered by LiDAR buoys offer novel opportunities for studying and modeling Pacific OCS offshore winds, but observational data is still extremely limited compared to most other wind energy development sites. Future deployment of these scientific buoys on the Pacific OCS will help to overcome the spatial and temporal constraints of the currently available data. With longer spans of continuous wind data, the ML techniques could be trained across longer timescales and tested on their ability to capture seasonal wind patterns. More spatial availability of observed vertical wind profiles between the Humboldt and Morro Bay WEAs would provide vital information for assessing the degradation of ML performance over shorter train-test extrapolation distances. Many more opportunities for wind-related ML research are available through the possibilities of short-term forecasting and prediction of other wind parameters, such as wind power density and turbulence properties. Other continuations of this research could be to draw comparisons with NWP models estimates, such as WRF, and to potentially combine the individual advantages of each ML technique through use of hybrid modeling. The research presented here determines that intelligent learning methods, namely LSTM and GPR, can provide more accurate and adaptable predictions of offshore winds in comparison to a conventional physical law approach. Thus, these state-of-the-art ML techniques demonstrate great promise for informing the development of Pacific OCS offshore wind energy and supplementing the future research needed for other potential offshore wind areas.

Bibliography

[1] N. Bodini, J. K. Lundquist, and A. Kirincich, "U.S. East Coast Lidar Measurements Show Offshore Wind Turbines Will Encounter Very Low Atmospheric Turbulence," *Geophysical Research Letters*, vol. 46, no. 10, pp. 5582–5591, 2019, ISSN: 1944-8007. DOI: 10.1029/2019GL082636.

[2] *Wind Market Reports: 2021 Edition*. [Online]. Available: https://www.energy.gov/eere/wind/wind-market-reports-2021-edition (visited on 01/20/2022).

[3] G. R. Timilsina, "Are renewable energy technologies cost competitive for electricity generation?" *Renewable Energy*, vol. 180, pp. 658–672, Dec. 2021, ISSN: 0960-1481. DOI: 10.1016/j.renene.2021.08.088.

[4] *Wind has surpassed hydro as most-used renewable electricity generation source in U.S. - Today in Energy - U.S. Energy Information Administration (EIA)*. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=42955 (visited on 01/20/2022).

[5] *Global Wind Report 2021*, Mar. 2021. [Online]. Available: https://gwec.net/global-wind-report-2021/ (visited on 01/20/2022).

[6] S. Wang and S. Wang, "Impacts of wind energy on environment: A review," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 437–443, Sep. 2015, ISSN: 1364-0321. DOI: 10.1016/j.rser.2015.04.137.

[7] *Haliade 150-6MW Offshore Wind Turbine — GE Renewable Energy*. [Online]. Available: https://www.ge.com/renewableenergy/wind-energy/offshore-wind/offshore-turbine-haliade-150-6mw (visited on 05/02/2022).

[8] *Dominion Energy — Coastal Virginia Offshore Wind*. [Online]. Available: https://coastalvawind.com/ (visited on 05/02/2022).

[9] *Biden-Harris Administration Advances Offshore Wind in the Pacific*, May 2021. [Online]. Available: https://www.doi.gov/pressreleases/biden-%20harris-administration-advances-offshore-wind-pacific (visited on 01/27/2022).

[10] *Humboldt Wind Energy Area — Bureau of Ocean Energy Management*. [Online]. Available: https://www.boem.gov/renewable-energy/state-activities/humboldt-wind-energy-area (visited on 04/19/2022).

[11] B. Shao, D. Song, G. Bian, and Y. Zhao, "Wind Speed Forecast Based on the LSTM Neural Network Optimized by the Firework Algorithm," *Advances in Materials Science and Engineering*, vol. 2021, e4874757, Sep. 2021, Publisher: Hindawi, ISSN: 1687-8434. DOI: 10.1155/2021/4874757.

[12] C. Ortega, A. Younes, M. Severy, C. Chamberlin, and A. Jacobson, "Resource and Load Compatibility Assessment of Wind Energy Offshore of Humboldt County, California," *Energies*, vol. 13, no. 21, p. 5707, Jan. 2020. DOI: 10.3390/en13215707.

[13] W. P. Mahoney, K. Parks, G. Wiener, *et al.*, "A Wind Power Forecasting System to Optimize Grid Integration," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 670–682, 2012, ISSN: 1949-3029.

[14] C. B. Hasager, P. H. Madsen, G. Giebel, *et al.*, "Design tool for offshore wind farm cluster planning: EWEA Annual Conference and Exhibition 2015," *Proceedings of the EWEA Annual Event and Exhibition 2015*, 2015, Publisher: European Wind Energy Association (EWEA).

[15] P. d. Jong, R. Dargaville, J. Silver, S. Utembe, A. Kiperstok, and E. A. Torres, "Forecasting high proportions of wind energy supplying the Brazilian Northeast electricity grid," *Applied Energy*, vol. 195, pp. 538–555, 2017, Publisher: Elsevier, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2017.03.058.

[16] S. Hsu, "Models for estimating offshore winds from onshore meteorological measurements," *Boundary-Layer Meteorology*, vol. 20, no. 3, pp. 341–351, 1981, ISSN: 1573-1472. DOI: 10.1007/BF00121378.

[17] S. SethuRaman and G. S. Raynor, "Comparison of Mean Wind Speeds and Turbulence at a Coastal Site and an Offshore Location," *Journal of Applied Meteorology (1962-1982)*, vol. 19, no. 1, pp. 15–21, 1980, Publisher: American Meteorological Society, ISSN: 0021-8952.

[18] Z. R. Shu, Q. S. Li, Y. C. He, and P. W. Chan, "Observations of offshore wind characteristics by Doppler-LiDAR for wind energy applications," *Applied Energy*, vol. 169, pp. 150–163, May 2016, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2016.01.135.

[19] A. Possner and K. Caldeira, "Geophysical potential for wind energy over the open oceans," *Proceedings of the National Academy of Sciences*, vol. 114, no. 43, pp. 11 338–11 343, Oct. 2017, Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.1705710114.

[20] *BOEM Advances Offshore Wind Leasing Process in California — Bureau of Ocean Energy Management*. [Online]. Available: https://www.boem.gov/newsroom/press-releases/boem-advances-offshore-wind-leasing-process-california (visited on 01/27/2022).

[21] H. Díaz, J. M. Rodrigues, and C. G. Soares, "Preliminary cost assessment of an offshore floating wind farm installation on the Galician coast," in *Progress in Renewable Energies Offshore : Proceedings of the 2nd International Conference on Renewable Energies, 2016 (RENEW2016)*, Taylor & Francis Books Ltd, 2016, pp. 843–850, ISBN: 978-1-138-62627-0.

[22] M. Satir, F. Murphy, and K. McDonnell, "Feasibility study of an offshore wind farm in the Aegean Sea, Turkey," *Renewable and Sustainable Energy Reviews*, vol. 81, no. 2, pp. 2552–2562, Jan. 2018. DOI: 10.1016/j.rser.2017.06.063.

[23] J. Bosch, I. Staffell, and A. D. Hawkes, "Global levelised cost of electricity from offshore wind," *Energy*, vol. 189, p. 116 357, Dec. 2019, ISSN: 0360-5442. DOI: 10.1016/j.energy.2019.116357.

[24] E. Lantz, G. Barter, P. Gilman, *et al.*, "Power Sector, Supply Chain, Jobs, and Emissions Implications of 30 Gigawatts of Offshore Wind Power by 2030," National Renewable Energy Lab. (NREL), Golden, CO (United States), Tech. Rep. NREL/TP-5000-80031, Aug. 2021. DOI: 10.2172/1814139.

[25] *Interior Department Takes Action to Advance Offshore Wind in the Atlantic and Gulf of Mexico*, Oct. 2021. [Online]. Available: https://www.doi.gov/pressreleases/interior-department-takes-action-advance-offshore-wind-atlantic-and-gulf-mexico (visited on 01/21/2022).

[26] K. L. McGrath, A. J. Turner, and A. Hamilton, *Five Things You Should Know About Offshore Wind Development Right Now*. [Online]. Available: https://www.natlawreview.com/article/five-things-you-should-know-about-offshore-wind-development-right-now (visited on 04/19/2022).

[27] D.-E. Choe, G. Talor, and C. Kim, "Prediction of Wind Speed, Potential Wind Power, and the Associated Uncertainties for Offshore Wind Farm Using Deep Learning," American Society of Mechanical Engineers Digital Collection, Oct. 2020. DOI: 10.1115/POWER2020-16557.

[28] W. Musial, P. Beiter, S. Tegen, and A. Smith, "Potential Offshore Wind Energy Areas in California: An Assessment of Locations, Technology, and Costs," Tech. Rep. NREL/TP–5000-67414, 1338174, Dec. 2016. DOI: 10.2172/1338174.

[29] *UC Berkeley GeoData Repository*. [Online]. Available: https://geodata.lib.berkeley.edu/catalog/stanford-qh592xp2377 (visited on 01/27/2022).

[30] M. J. Dvorak, C. L. Archer, and M. Z. Jacobson, "California offshore wind energy potential," *Renewable Energy*, vol. 35, no. 6, pp. 1244–1254, Jun. 2010, ISSN: 0960-1481. DOI: 10.1016/j.renene.2009.11.022.

[31] N. Bento and M. Fontes, "Emergence of floating offshore wind energy: Technology and industry," *Renewable and Sustainable Energy Reviews*, vol. 99, pp. 66–82, Jan. 2019, ISSN: 1364-0321. DOI: 10.1016/j.rser.2018.09.035.

[32] W. Musial and B. Ram, "Large-Scale Offshore Wind Power in the United States: Assessment of Opportunities and Barriers," National Renewable Energy Lab. (NREL), Golden, CO (United States), Tech. Rep. NREL/TP-500-40745, Sep. 2010. DOI: 10.2172/990101.

[33] P. Beiter, W. Musial, L. Kilcher, M. Maness, and A. Smith, "An Assessment of the Economic Potential of Offshore Wind in the United States from 2015 to 2030," National Renewable Energy Lab. (NREL), Golden, CO (United States), Tech. Rep. NREL/TP-6A20-67675, Mar. 2017. DOI: 10.2172/1349721.

[34] J. K. Kaldellis and M. Kapsali, "Shifting towards offshore wind energy—Recent activity and future development," *Energy Policy*, vol. 53, pp. 136–148, Feb. 2013, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2012.10.032.

[35] T. Tajalli Bakhsh, K. Simpson, T. LaPierre, *et al.*, "Potential Geo-Hazards to Floating Offshore Wind Farms in the US Pacific," American Society of Mechanical Engineers Digital Collection, Apr. 2021. DOI: 10.1115/IOWTC2021-3564.

[36] M. Schwartz, D. Heimiller, S. Haymes, and W. Musial, "Assessment of Offshore Wind Energy Resources for the United States," Tech. Rep. NREL/TP-500-45889, 983415, Jun. 2010. DOI: 10.2172/983415.

[37] *Offshore wind energy – Schatz Energy Research Center*. [Online]. Available: http://schatzcenter.org/wind/ (visited on 01/22/2022).

[38] M. Debnath, P. Doubrawa, M. Optis, P. Hawbecker, and N. Bodini, "Extreme wind shear events in US offshore wind energy areas and the role of induced stratification," *Wind Energy Science*, vol. 6, no. 4, pp. 1043–1059, Aug. 2021, Publisher: Copernicus GmbH, ISSN: 2366-7443. DOI: 10.5194/wes-6-1043-2021.

[39] *Interior Department Announces Environmental Review of Proposed Wind Energy Projects Offshore New Jersey*, Sep. 2021. [Online]. Available: https://www.doi.gov/pressreleases/interior-department-announces-environmental-review-proposed-wind-energy-projects (visited on 01/27/2022).

[40] M. Optis, O. Rybchuk, N. Bodini, M. Rossol, and W. Musial, "2020 Offshore Wind Resource Assessment for the California Pacific Outer Continental Shelf," Tech. Rep. NREL/TP-5000-77642, 1677466, MainId:29568, Oct. 2020. DOI: 10.2172/1677466.

[41] R. K. Newsom, L. M. Sheridan, B. J. Gaudet, *et al.*, "A Study on Modeled Wind Speed Errors Using the U.S. Department of Energy Buoys," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Tech. Rep. PNNL-30117, Jul. 2020. DOI: 10.2172/1713065.

[42] *Deployments and Analysis — PNNL*. [Online]. Available: https://www.pnnl.gov/projects/lidar-buoy-program/deployments-and-analysis (visited on 01/24/2022).

[43] M. A. Severy, A. M. Gorton, R. Krishnamurthy, and M. S. Levin, *Lidar Buoy Data Dictionary*, May 2021. [Online]. Available: https://a2e.energy.gov/data/buoy/lidar.z05.a0/attach/pnnl-30937-datadictionary.pdf.

[44] M. Wächter, J. Gottschall, A. Rettenmeier, and J. Peinke, "Power curve estimation using LIDAR measurements," *European Wind Energy Conference and Exhibition 2009, EWEC 2009*, vol. 7, Jan. 2009.

[45] *Technical Specifications — PNNL*. [Online]. Available: https://www.pnnl.gov/projects/lidar-buoy-program/technical-specifications (visited on 01/24/2022).

[46] M. J. Kavaya, J. Y. Beyon, G. J. Koch, *et al.*, "The Doppler Aerosol Wind (DAWN) Airborne, Wind-Profiling Coherent-Detection Lidar System: Overview and Preliminary Flight Results," *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 4, pp. 826–842, Apr. 2014, Publisher: American Meteorological Society Section: Journal of Atmospheric and Oceanic Technology, ISSN: 0739-0572, 1520-0426. DOI: 10.1175/JTECH-D-12-00274.1.

[47] S.-E. Gryning, T. Mikkelsen, C. Baehr, *et al.*, "2 - Measurement methodologies for wind energy based on ground-level remote sensing," in *Renewable Energy Forecasting*, ser. Woodhead Publishing Series in Energy, G. Kariniotakis, Ed., Woodhead Publishing, Jan. 2017, pp. 29–56, ISBN: 978-0-08-100504-0. DOI: 10.1016/B978-0-08-100504-0.00002-0.

[48] Z. Liu, J. F. Barlow, P.-W. Chan, *et al.*, "A Review of Progress and Applications of Pulsed Doppler Wind LiDARs," *Remote Sensing*, vol. 11, no. 21, p. 2522, Jan. 2019, Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-4292. DOI: 10.3390/rs11212522.

[49] M. Hill, R. Calhoun, H. J. S. Fernando, *et al.*, "Coplanar Doppler Lidar Retrieval of Rotors from T-REX," *Journal of the Atmospheric Sciences*, vol. 67, no. 3, pp. 713–729, Sep. 2010, ISSN: 0022-4928. DOI: 10.1175/2009jas3016.1.

[50] R. Krishnamurthy, R. Calhoun, and H. Fernando, "Large-eddy simulation-based retrieval of dissipation from coherent Doppler lidar data," *Boundary-Layer Meteorology*, vol. 136, no. 1, pp. 45–57, 2010, ISSN: 0006-8314. DOI: 10.1007/s10546-010-9495-y.

[51] M. Sanchez Gomez, J. K. Lundquist, P. M. Klein, and T. M. Bell, "Turbulence Dissipation Rate Estimated from Lidar ObservationsDuring the LAPSE-RATE Field Campaign," Atmosphere – Meteorology, preprint, Jan. 2021. DOI: 10.5194/essd-2020-406.

[52] C. Chen, J. Zhou, X. Li, *et al.*, "Ultrashort-Term Forecasting of Mid–Low-Altitude Wind Based on LSTM," in *Proceedings of the 5th China High Resolution Earth Observation Conference (CHREOC 2018)*, L. Wang, Y. Wu, and J. Gong, Eds., ser. Lecture Notes in Electrical Engineering, Springer, 2019, pp. 104–111, ISBN: 9789811365539. DOI: 10.1007/978-981-13-6553-9_12.

[53] L. Thobois, J. P. Cariou, and I. Gultepe, "Review of Lidar-Based Applications for Aviation Weather," *Pure and Applied Geophysics*, vol. 176, no. 5, pp. 1959–1976, May 2019, ISSN: 1420-9136. DOI: 10.1007/s00024-018-2058-8.

[54] R. Dang, Y. Yang, X.-M. Hu, Z. Wang, and S. Zhang, "A Review of Techniques for Diagnosing the Atmospheric Boundary Layer Height (ABLH) Using Aerosol Lidar Data," *Remote Sensing*, vol. 11, no. 13, p. 1590, Jan. 2019. DOI: 10.3390/rs11131590.

[55] W. E. Baker, R. Atlas, C. Cardinali, *et al.*, "Lidar-Measured Wind Profiles: The Missing Link in the Global Observing System," *Bulletin of the American Meteorological Society*, vol. 95, no. 4, pp. 543–564, Apr. 2014, Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-12-00164.1.

[56] N. Bodini, J. K. Lundquist, and R. K. Newsom, "Estimation of turbulence dissipation rate and its variability from sonic anemometer and wind Doppler lidar during the XPIA field campaign," *Atmospheric Measurement Techniques*, vol. 11, no. 7, pp. 4291–4308, Jul. 2018, Publisher: Copernicus GmbH, ISSN: 1867-1381. DOI: 10.5194/amt-11-4291-2018.

[57] V.-M. Kumer, J. Reuder, and B. R. Furevik, "A Comparison of LiDAR and Radiosonde Wind Measurements," *Energy Procedia*, EERA DeepWind' 2014, 11th Deep Sea Offshore Wind R&D Conference, vol. 53, pp. 214–220, Jan. 2014, ISSN: 1876-6102. DOI: 10.1016/j.egypro.2014.07.230.

[58] D. A. Smith, M. Harris, A. S. Coffey, *et al.*, "Wind lidar evaluation at the Danish wind test site in Høvsøre," *Wind Energy*, vol. 9, no. 1-2, pp. 87–93, 2006, ISSN: 1099-1824. DOI: 10.1002/we.193.

[59] D. Kindler, A. Oldroyd, A. MacAskill, and D. Finch, "An eight month test campaign of the Qinetiq ZephIR system: Preliminary results," *Meteorologische Zeitschrift*, pp. 479–489, Oct. 2007, Publisher: Schweizerbart'sche Verlagsbuchhandlung. DOI: 10.1127/0941-2948/2007/0226.

[60] A. Peña, C. B. Hasager, S.-E. Gryning, M. Courtney, I. Antoniou, and T. Mikkelsen, "Offshore wind profiling using light detection and ranging measurements," *Wind Energy*, vol. 12, no. 2, pp. 105–124, 2009, ISSN: 1099-1824. DOI: 10.1002/we.283.

[61] D. G. Ortiz-Suslow and Q. Wang, "An Evaluation of Kolmogorov's 5/3 Power Law Observed Within the Turbulent Airflow Above the Ocean," *Geophysical Research Letters*, vol. 46, no. 24, pp. 14 901–14 911, 2019, ISSN: 1944-8007. DOI: 10.1029/2019GL085083.

[62] S. Kucukali and Ç. Dinçkal, "Wind energy resource assessment of Izmit in the West Black Sea Coastal Region of Turkey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 790–795, Feb. 2014, ISSN: 1364-0321. DOI: 10.1016/j.rser.2013.11.018.

[63] Y. L. Pichugina, R. M. Banta, W. A. Brewer, S. P. Sandberg, and R. M. Hardesty, "Doppler Lidar–Based Wind-Profile Measurement System for Offshore Wind-Energy and Other Marine Boundary Layer Applications," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 2, pp. 327–349, Feb. 2012, Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, ISSN: 1558-8424, 1558-8432. DOI: 10.1175/JAMC-D-11-040.1.

[64] I. Antoniou, H. Jørgensen, T. Mikkelsen, *et al.*, "Offshore wind profile measurements from remote sensing instruments," *European Wind Energy Conference and Exhibition 2006, EWEC 2006*, vol. 1, pp. 471–480, Jan. 2006.

[65] M. Optis, N. Bodini, M. Debnath, and P. Doubrawa, "New methods to improve the vertical extrapolation of near-surface offshore wind speeds," *Wind Energy Science*, vol. 6, no. 3, pp. 935–948, Jun. 2021, Publisher: Copernicus GmbH, ISSN: 2366-7443. DOI: 10.5194/wes-6-935-2021.

[66] ——, "Best Practices for the Validation of U.S. Offshore Wind Resource Models," Tech. Rep. NREL/TP-5000-78375, 1755697, MainId:32292, Dec. 2020. DOI: 10.2172/1755697.

[67] O. Bischoff, I. Wuerth, P. Cheng, J. Tiana Alsina, and M. Gutierrez, "Motion effects on lidar wind measurement data of the EOLOS buoy," in *Renewable Energies Offshore*, Google-Books-ID: nNCYCgAAQBAJ, CRC Press, Sep. 2015, pp. 197–204, ISBN: 978-1-315-64785-2.

[68] G. Wolken-Möhlmann, H. Lilov, and B. Lange, "Simulation of motion induced measurement errors for wind measurements using LIDAR on floating platforms," *Proceedings 15th International Symposium for the Advancement of Boundary-layer Remote Sensing (ISARS)*, pp. 28–30, Jan. 2010.

[69] T. Désert, G. Knapp, and S. Aubrun, "Quantification and Correction of Wave-Induced Turbulence Intensity Bias for a Floating LIDAR System," *Remote Sensing*, vol. 13, no. 15, p. 2973, Jan. 2021, Number: 15 Publisher: Multidisciplinary Digital Publishing Institute. DOI: 10.3390/rs13152973.

[70] J. Tiana-Alsina, M. A. Gutiérrez, I. Würth, J. Puigdefábregas, and F. Rocadenbosch, "Motion compensation study for a floating Doppler wind LiDAR," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2015, pp. 5379–5382. DOI: 10.1109/IGARSS.2015.7327051.

[71] W. J. Shaw, J. Draher, G. Garcia Medina, *et al.*, "General Analysis of Data Collected from DOE Lidar Buoy Deployments Off Virginia and New Jersey," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Tech. Rep. PNNL-29823, Apr. 2020. DOI: 10.2172/1632348.

[72] L. M. Sheridan, R. Krishnamurthy, and B. J. Gaudet, "Assessment of Model Hub Height Wind Speed Performance Using DOE Lidar Buoy Data," Pacific Northwest National Lab. (PNNL), Richland, WA (United States), Tech. Rep. PNNL-30840, Apr. 2021. DOI: 10.2172/1779495.

[73] *Numerical Weather Prediction — National Centers for Environmental Information (NCEI)*. [Online]. Available: https://www.ncei.noaa.gov/products/weather-climate-models/numerical-weather-prediction (visited on 01/22/2022).

[74] H. Hersbach, B. Bell, P. Berrisford, *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020, ISSN: 1477-870X. DOI: 10.1002/qj.3803.

[75] S. G. Benjamin, S. S. Weygandt, J. M. Brown, *et al.*, "A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh," *Monthly Weather Review*, vol. 144, no. 4, pp. 1669–1694, Apr. 2016, Publisher: American Meteorological Society Section: Monthly Weather Review, ISSN: 1520-0493, 0027-0644. DOI: 10.1175/MWR-D-15-0242.1.

[76] C. Draxl, B. M. Hodge, A. Clifton, and J. McCaa, "Overview and Meteorological Validation of the Wind Integration National Dataset toolkit," National Renewable Energy Lab. (NREL), Golden, CO (United States), Tech. Rep. NREL/TP-5000-61740, Apr. 2015. DOI: 10.2172/1214985.

[77] *Meteorological and oceanographic data collected from the National Data Buoy Center Coastal-Marine Automated Network (C-MAN) and moored (weather) buoys*, Last Modified: 2022-04-19, 2021. [Online]. Available: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:NDBC-CMANWx (visited on 04/19/2022).

[78] S. Samadianfard, S. Hashemi, K. Kargar, *et al.*, "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm," *Energy Reports*, vol. 6, pp. 1147–1159, Nov. 2020, ISSN: 2352-4847. DOI: 10.1016/j.egyr.2020.05.001.

[79] E. W. Peterson and J. P. Hennessey, "On the Use of Power Laws for Estimates of Wind Power Potential," *Journal of Applied Meteorology and Climatology*, vol. 17, no. 3, pp. 390–394, Mar. 1978, Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, ISSN: 1520-0450. DOI: 10.1175/1520-0450(1978)017<0390:OTUOPL>2.0.CO;2.

[80] A. S. Monin and A. M. Obukhov, "Basic Laws of Turbulent Mixing in the Surface Layer of the Atmosphere," *Geophys. Inst. Acad. Sci. USSR*, vol. 24, pp. 163–187, 1954.

[81] F. Onea and E. Rusu, "An Evaluation of the Wind Energy in the North-West of the Black Sea," *International Journal of Green Energy*, vol. 11, no. 5, pp. 465–487, May 2014, ISSN: 1543-5075. DOI: 10.1080/15435075.2013.773513.

[82] J.-K. Jang, B.-M. Yu, K.-W. Ryu, and J.-S. Lee, "Offshore Wind Resource Assessment around Korean Peninsula by using QuikSCAT Satellite Data," *Journal of the Korean Society for Aeronautical & Space Sciences*, vol. 37, no. 11, pp. 1121–1130, 2009, Publisher: The Korean Society for Aeronautical and Space Sciences, ISSN: 1225-1348. DOI: 10.5139/JKSAS.2009.37.11.1121.

[83] B. Bethel, "Caribbean Sea Offshore Wind Energy Assessment and Forecasting," *Journal of Marine Science and Application*, vol. 20, Jul. 2021. DOI: 10.1007/s11804-021-00216-z.

[84] L. Alblas, W. Bierbooms, and D. Veldkamp, "Power output of offshore wind farms in relation to atmospheric stability," *Journal of Physics: Conference Series*, vol. 555, p. 012 004, Dec. 2014, Publisher: IOP Publishing, ISSN: 1742-6596. DOI: 10.1088/1742-6596/555/1/012004.

[85] J. S. Rodrigo, E. Cantero, B. García, *et al.*, "Atmospheric stability assessment for the characterization of offshore wind conditions," *Journal of Physics: Conference Series*, vol. 625, p. 012 044, Jun. 2015, ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/625/1/012044.

[86] U. Högström, A.-S. Smedman, and H. Bergstrom, "Calculation of Wind Speed Variation With Height Over the Sea," *Wind Engineering*, vol. 30, pp. 269–286, May 2006. DOI: 10.1260/030952406779295480.

[87] A. Sathe, S.-E. Gryning, and A. Peña, "Comparison of the atmospheric stability and wind profiles at two wind farm sites over a long marine fetch in the North Sea," *Wind Energy*, vol. 14, pp. 767–780, Sep. 2011. DOI: 10.1002/we.456.

[88] A. J. M. Van Wijk, A. C. M. Beljaars, A. A. M. Holtslag, and W. C. Turkenburg, "Evaluation of stability corrections in wind speed profiles over the North Sea," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 33, no. 3, pp. 551–566, Apr. 1990, ISSN: 0167-6105. DOI: 10.1016/0167-6105(90)90007-Y.

[89] P. Domagalski, L. M. Bardal, and L. R. Sætran, "Vertical Wind Profiles in Non-Neutral Conditions: Comparison of Models and Measurements From Frøya," *Journal of Offshore Mechanics and Arctic Engineering*, vol. 141, no. 4, Jan. 2019, ISSN: 0892-7219. DOI: 10.1115/1.4041816.

[90] M. Richmond, A. Sobey, R. Pandit, and A. Kolios, "Stochastic assessment of aerodynamics within offshore wind farms based on machine-learning," *Renewable Energy*, vol. 161, pp. 650–661, Dec. 2020, ISSN: 0960-1481. DOI: 10.1016/j.renene.2020.07.083.

[91] Z. Sun, H. Sun, and J. Zhang, "Multistep Wind Speed and Wind Power Prediction Based on a Predictive Deep Belief Network and an Optimized Random Forest," *Mathematical Problems in Engineering*, vol. 2018, e6231745, Jul. 2018, Publisher: Hindawi, ISSN: 1024-123X. DOI: 10.1155/2018/6231745.

[92] N. Bodini and M. Optis, "The importance of round-robin validation when assessing machine-learning-based vertical extrapolation of wind speeds," *Wind Energy Science*, vol. 5, no. 2, pp. 489–501, Apr. 2020, Publisher: Copernicus GmbH, ISSN: 2366-7443. DOI: 10.5194/wes-5-489-2020.

[93] T. Liu, H. Wei, and K. Zhang, "Wind power prediction with missing data using Gaussian process regression and multiple imputation," *Applied Soft Computing*, vol. 71, pp. 905–916, Oct. 2018, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2018.07.027.

[94] G. Chen, B. Tang, X. Zeng, P. Zhou, P. Kang, and H. Long, "Short-term wind speed forecasting based on long short-term memory and improved BP neural network," *International Journal of Electrical Power & Energy Systems*, vol. 134, Jan. 2022, ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2021.107365.

[95] D. Vassallo, R. Krishnamurthy, T. Sherman, and H. J. S. Fernando, "Analysis of Random Forest Modeling Strategies for Multi-Step Wind Speed Forecasting," *Energies*, vol. 13, no. 20, p. 5488, Jan. 2020, Number: 20 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1996-1073. DOI: 10.3390/en13205488.

[96] D. Lee and R. Baldick, "Short-Term Wind Power Ensemble Prediction Based on Gaussian Processes and Neural Networks," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 501–510, Jan. 2014, Conference Name: IEEE Transactions on Smart Grid, ISSN: 1949-3061. DOI: 10.1109/TSG.2013.2280649.

[97] N. Chen, Z. Qian, I. Nabney, and X. Meng, "Wind power forecasts using gaussian processes and numerical weather prediction," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 656–665, 2014, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2013.2282366.

[98] J. Hu and J. Wang, "Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression," *Energy*, vol. 93, pp. 1456–1466, Dec. 2015, ISSN: 0360-5442. DOI: 10.1016/j.energy.2015.10.041.

[99] J. Yu, K. Chen, J. Mori, and M. M. Rashid, "A Gaussian mixture copula model based localized Gaussian process regression approach for long-term wind speed prediction," *Energy*, vol. 61, pp. 673–686, Nov. 2013, ISSN: 0360-5442. DOI: 10.1016/j.energy.2013.09.013.

[100] A. Banik, C. Behera, T. V. Sarathkumar, and A. K. Goswami, "Uncertain wind power forecasting using LSTM-based prediction interval," *IET Renewable Power Generation*, vol. 14, no. 14, pp. 2657–2667, 2020, ISSN: 1752-1424. DOI: 10.1049/iet-rpg.2019.1238.

[101] B. K. Saxena, S. Mishra, and K. V. S. Rao, "Offshore wind speed forecasting at different heights by using ensemble empirical mode decomposition and deep learning models," *Applied Ocean Research*, vol. 117, p. 102 937, Dec. 2021, ISSN: 0141-1187. DOI: 10.1016/j.apor.2021.102937.

[102] C. Serrano-Barreto, C. Leonard, and M. Santos, "LSTM Neural Network Modeling of Wind Speed and Correlation Analysis of Wind and Waves," in *Intelligent Data Engineering and Automated Learning – IDEAL 2021*, H. Yin, D. Camacho, P. Tino, *et al.*, Eds., ser. Lecture Notes in Computer Science, Springer International Publishing, 2021, pp. 458–466, ISBN: 978-3-030-91608-4. DOI: 10.1007/978-3-030-91608-4_45.

[103] C. Yuan, Y. Tang, R. Mei, F. Mo, and H. Wang, "A PSO-LSTM Model of Offshore Wind Power Forecast considering the Variation of Wind Speed in Second-Level Time Scale," *Mathematical Problems in Engineering*, vol. 2021, e2009062, Nov. 2021, Publisher: Hindawi, ISSN: 1024-123X. DOI: 10.1155/2021/2009062.

[104] K. Shi, Y. Qiao, W. Zhao, Q. Wang, M. Liu, and Z. Lu, "An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness," *Wind Energy*, vol. 21, no. 12, pp. 1383–1394, 2018, ISSN: 1099-1824. DOI: 10.1002/we.2261.

[105] *Atmosphere to Electrons (A2e)*, 2021. DOI: 10.21947/1783809. (visited on 01/22/2022).

[106] *A2e: Atmosphere to Electrons*. [Online]. Available: https://a2e.energy.gov/data#AreaFilter=%2248.1074311884803 9%2C-127.35351562500001%20to%2027.52775820686191%2C-67.76367187500001%22&ProjectFilter=%5B%22buoy%22%5D (visited on 04/20/2022).

[107] *Wind Energy WindCube*. [Online]. Available: https://www.vaisala.com/en/wind-lidars/wind-energy/windcube (visited on 01/24/2022).

[108] G. Geernaert, "BOUNDARY LAYERS — Surface Layer," in *Encyclopedia of Atmospheric Sciences*, Elsevier, 2003, pp. 305–311, ISBN: 978-0-12-227090-1. DOI: 10.1016/B0-12-227090-8/00092-0.

[109] A. K. Resen, A. B. Khamees, and S. F. Yaseen, "Determination of Wind Shear Coefficients and Conditions of Atmospheric Stability for Three Iraqi Sites," *IOP Conference Series: Materials Science and Engineering*, vol. 881, no. 1, p. 012 161, Jul. 2020, ISSN: 1757-8981, 1757-899X. DOI: 10.1088/1757-899X/881/1/012161.

[110] C. Fairall, E. Bradley, D. Rogers, J. Edson, and G. Young, "Bulk parameterization of air-sea fluxes for the Tropical Ocean Global Atmosphere Coupled Ocean-Atmosphere Response Experiment (TOGA-COARE)," *J. Geophys. Res.*, vol. 101, pp. 3747–3764, Feb. 1996. DOI: 10.1029/95JC03205.

[111] A. A. Grachev and C. W. Fairall, "Dependence of the Monin–Obukhov Stability Parameter on the Bulk Richardson Number over the Ocean," *Journal of Applied Meteorology and Climatology*, vol. 36, no. 4, pp. 406–414, Apr. 1997, Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, ISSN: 1520-0450, 0894-8763. DOI: 10.1175/1520-0450(1997)036<0406:DOTMOS>2.0.CO;2.

[112] M. Golbazi and C. L. Archer, "Surface roughness for offshore wind energy," *Journal of Physics: Conference Series*, vol. 1452, no. 1, p. 012 024, Jan. 2020, ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/1452/1/012024.

[113] J. Wieringa, "Updating the Davenport roughness classification," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 41, no. 1, pp. 357–368, Oct. 1992, ISSN: 0167-6105. DOI: 10.1016/0167-6105(92)90434-C.

[114] R. Barthelmie, J. Badger, S. Pryor, C. Hasager, M. Christiansen, and B. Jørgensen, "Offshore Coastal Wind Speed Gradients: Issues for the Design and Development of Large Offshore Windfarms," *Wind Engineering*, vol. 31, no. 6, pp. 369–382, Dec. 2007, Publisher: SAGE Publications, ISSN: 0309-524X. DOI: 10.1260/030952407784079762.

[115] B. Lange, S. Larsen, J. Højstrup, and R. Barthelmie, "Importance of thermal effects and sea surface roughness for offshore wind resource assessment," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 92, no. 11, pp. 959–988, Sep. 2004, ISSN: 0167-6105. DOI: 10.1016/j.jweia.2004.05.005.

[116] H. S. Dhiman and D. Deb, "Machine Intelligent Techniques for Ramp Event Prediction in Offshore and Onshore Wind Farms," *arXiv:2011.14220 [cs, eess]*, Nov. 2020, arXiv: 2011.14220.

[117] A. Obukhov, "Turbulentnost'v temperaturnoj-neodnorodnoj atmosfere," *Trudy Inst. Theor. Geofiz. AN SSSR*, vol. 1, pp. 95–115, 1946.

[118] A. A. M. Holtslag, "Estimates of diabatic wind speed profiles from near-surface weather observations," *Boundary-Layer Meteorology*, vol. 29, no. 3, pp. 225–250, Jul. 1984, ISSN: 1573-1472. DOI: 10.1007/BF00119790.

[119] M. C. Holtslag, W. A. A. M. Bierbooms, and G. J. W. van Bussel, "Validation of surface layer similarity theory to describe far offshore marine conditions in the Dutch North Sea in scope of wind energy research," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 136, pp. 180–191, Jan. 2015, ISSN: 0167-6105. DOI: 10.1016/j.jweia.2014.10.013.

[120] S.-E. Gryning, E. Batchvarova, B. Brümmer, H. Jørgensen, and S. Larsen, "On the extension of the wind profile over homogeneous terrain beyond the surface boundary layer," *Boundary-Layer Meteorology*, vol. 124, no. 2, pp. 251–268, Aug. 2007, ISSN: 1573-1472. DOI: 10.1007/s10546-007-9166-9.

[121] C. L. Archer, B. A. Colle, D. L. Veron, F. Veron, and M. J. Sienkiewicz, "On the predominance of unstable atmospheric conditions in the marine boundary layer offshore of the U.S. northeastern coast," *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 15, pp. 8869–8885, 2016, ISSN: 2169-8996. DOI: 10.1002/2016JD024896.

[122] J. H. Evans, "Dimensional Analysis and the Buckingham Pi Theorem," *American Journal of Physics*, vol. 40, no. 12, pp. 1815–1822, Dec. 1972, Publisher: American Association of Physics Teachers, ISSN: 0002-9505. DOI: 10.1119/1.1987069.

[123] J. A. Businger, J. C. Wyngaard, Y. Izumi, and E. F. Bradley, "Flux-Profile Relationships in the Atmospheric Surface Layer," *Journal of the Atmospheric Sciences*, vol. 28, no. 2, pp. 181–189, Mar. 1971, Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, ISSN: 0022-4928, 1520-0469. DOI: 10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2.

[124] A. J. Dyer, "A review of flux-profile relationships," *Boundary-Layer Meteorology*, vol. 7, no. 3, pp. 363–372, Nov. 1974, ISSN: 1573-1472. DOI: 10.1007/BF00240838.

[125] Y. Delage and C. Girard, "Stability functions correct at the free convection limit and consistent for for both the surface and Ekman layers," *Boundary-Layer Meteorology*, vol. 58, no. 1, pp. 19–31, Jan. 1992, ISSN: 1573-1472. DOI: 10.1007/BF00120749.

[126] A. A. Grachev, C. W. Fairall, and E. F. Bradley, "Convective Profile Constants Revisited," *Boundary-Layer Meteorology*, vol. 94, no. 3, pp. 495–515, Mar. 2000, ISSN: 1573-1472. DOI: 10.1023/A:1002452529672.

[127] A. C. M. Beljaars and A. a. M. Holtslag, "Flux Parameterization over Land Surfaces for Atmospheric Models," *Journal of Applied Meteorology and Climatology*, vol. 30, no. 3, pp. 327–341, Mar. 1991, Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, ISSN: 1520-0450, 0894-8763. DOI: 10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2.

[128] C. A. Paulson, "The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer," *Journal of Applied Meteorology and Climatology*, vol. 9, no. 6, pp. 857–861, Dec. 1970, Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, ISSN: 1520-0450. DOI: 10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2.

[129] W. Brutsaert, "Mean Profiles and Similarity in a Stationary and Horizontally-Uniform ABL," in *Evaporation into the Atmosphere: Theory, History and Applications*, ser. Environmental Fluid Mechanics, Springer Netherlands, 1982, pp. 57–112, ISBN: 978-94-017-1497-6. DOI: 10.1007/978-94-017-1497-6_4.

[130] K. Blümel, "An Approximate Analytical Solution Of Flux-Profile Relationships For The Atmospheric Surface Layer With Different Momentum And Heat Roughness Lengths," *Boundary-Layer Meteorology*, vol. 97, no. 2, pp. 251–271, Nov. 2000, ISSN: 1573-1472. DOI: 10.1023/A:1002708313017.

[131] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, Aug. 1998, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, ISSN: 1939-3539. DOI: 10.1109/34.709601.

[132] A. Lahouar and J. Ben Hadj Slama, "Hour-ahead wind power forecast based on random forests," *Renewable Energy*, vol. 109, pp. 529–541, Aug. 2017, ISSN: 0960-1481. DOI: 10.1016/j.renene.2017.03.064.

[133] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, 3. print, ser. Adaptive computation and machine learning. MIT Press, 2008, ISBN: 978-0-262-18253-9.

[134] J. Wang, "An Intuitive Tutorial to Gaussian Processes Regression," *arXiv:2009.10862 [cs, stat]*, Feb. 2021, arXiv: 2009.10862.

[135] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian Process Meets Big Data: A Review of Scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4405–4423, Nov. 2020, Conference Name: IEEE Transactions on Neural Networks and Learning Systems, ISSN: 2162-2388. DOI: 10.1109/TNNLS.2019.2957109.

[136] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

[137] J. Moon, S. Park, S. Rho, and E. Hwang, "A comparative analysis of artificial neural network architectures for building energy consumption forecasting," *International Journal of Distributed Sensor Networks*, vol. 15, no. 9, Sep. 2019, Publisher: SAGE Publications, ISSN: 1550-1329. DOI: 10.1177/1550147719877616.

[138] C. Olah, *Understanding LSTM Networks*, 2015. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (visited on 04/12/2022).

[139] H. Zheng, J. Yuan, and L. Chen, "Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation," *Energies*, vol. 10, no. 8, p. 1168, Aug. 2017, Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1996-1073. DOI: 10.3390/en10081168.

[140] S. Santhanam, "Context based Text-generation using LSTM networks," *arXiv:2005.00048 [cs]*, Apr. 2020.

[141] *Offshore Wind Technical Potential — Analysis and Maps — ESMAP*. [Online]. Available: https://esmap.org/esmap_offshorewind_techpotential_analysis_maps (visited on 04/06/2022).

[142] D. Borvarán, A. Peña, and R. Gandoin, "Characterization of offshore vertical wind shear conditions in Southern New England," *Wind Energy*, vol. 24, no. 5, pp. 465–480, 2021, ISSN: 1099-1824. DOI: 10.1002/we.2583.

[143] A. Abbasi, F. O. Annor, and N. van de Giesen, "Effects of atmospheric stability conditions on heat fluxes from small water surfaces in (semi-)arid regions," *Hydrological Sciences Journal*, vol. 62, no. 9, pp. 1422–1439, Jul. 2017, ISSN: 0262-6667. DOI: 10.1080/02626667.2017.1329587.

[144] C. W. Fairall, E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson, "Bulk Parameterization of Air–Sea Fluxes: Updates and Verification for the COARE Algorithm," *Journal of Climate*, vol. 16, no. 4, pp. 571–591, Feb. 2003, Publisher: American Meteorological Society Section: Journal of Climate, ISSN: 0894-8755, 1520-0442. DOI: 10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

[145] M. Optis and J. Perr-Sauer, "The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production," *Renewable and Sustainable Energy Reviews*, vol. 112, pp. 27–41, Sep. 2019, ISSN: 1364-0321. DOI: 10.1016/j.rser.2019.05.031.