Learning Spatio-Temporal LES Corrections using Wind Field Information in Wind Resource Assessment

A.C. Sonneveld

Learning Spatio-Temporal LES Corrections using Wind Field Information in Wind Resource Assessment

by

A.C. Sonneveld

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday December 15th, 2025 at 11:00 AM.

Student number:	4854578	
Project duration:	April 23, 2025 – December 15, 2025	
Thesis committee:	Prof. dr. ir. M.J.T. Reinders,	TU Delft, supervisor
	Dr. J. Sun,	TU Delft, daily supervisor
	Dr. M. Viljanen,	Whiffle, company supervisor
	Dr. J. Urbano,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Abstract

Accurate Wind Resource Assessment (WRA) requires the correction of systematic errors in modeled wind fields using sparse and temporally limited on-site measurements. With the emergence of microscale Large-Eddy Simulation (LES) as a high-resolution alternative to traditional mesoscale models, current correction practices, such as Measure-Correlate-Predict (MCP) for temporal extrapolation and Inverse Distance Weighting (IDW) for spatial extrapolation, need to be re-evaluated and adapted to the microscale context. This thesis investigates how data-driven methods can improve LES wind-speed error correction in both time and space, addressing three research gaps: the lack of validation of MCP on microscale data, the absence of temporal and spatial context in standard MCP-style formulations, and the limited use of flow information in spatial extrapolation.

The temporal component evaluates a set of linear and nonlinear regression models as flexible MCP-style baselines and extends them with additional contextual information. Temporal structure is incorporated through multi-step time windows, and local spatial structure is added by including neighboring LES grid cells. To represent full-field atmospheric patterns, the models are further enriched with latent encodings of the LES wind-speed field obtained through a Convolutional Autoencoder. The results show that MCP generalizes well to the microscale and that adding temporal and spatial context improves time-series accuracy across all observation locations, with combined strategies outperforming industry-standard methods.

The spatial component introduces Wind Speed-enhanced IDW (WS-IDW), which augments traditional IDW by weighting observation locations not only by geographic distance but also by similarity in LES wind speed. WS-IDW produces consistent improvements over the baseline, particularly when more observation locations are available. Analysis of the correction maps reveals that WS-IDW partially smooths misplaced fine-scale streaks in the LES wind field, supporting the hypothesis that LES is prone to slight spatial misalignment of coherent structures. The proposed method generalizes reasonably across sites and across different numbers of source masts.

Together, the temporal and spatial results demonstrate that incorporating LES-derived spatial and temporal information yields systematic improvements in microscale correction performance. The thesis provides a refined understanding of how LES behavior interacts with data-driven correction methods and offers a foundation for developing more robust microscale WRA correction frameworks in future work.

Contents

1

Introduction

The transition to sustainable energy sources is one of the defining challenges of the 21st century, with wind energy playing a crucial part in global efforts to mitigate climate change and reduce reliance on fossil fuels. Projections indicate that the expansive deployment of wind power could mitigate global warming by approximately 0.3 to 0.8°C by the end of the century, contingent on the scale of adoption and socio-economic pathways [1]. This potential is vital for achieving international climate targets, such as those outlined in the Paris Agreement. To realize this potential and ensure energy security, wind projects must move from theoretical concepts to operational reality, a process in which Wind Resource Assessment (WRA) is an important step.

WRA is the systematic process of evaluating and quantifying the wind energy potential at a specific location or over a larger region. This process relies on hindcasting: a retrospective analysis in which numerical models use historical data to simulate past wind conditions. WRA is critical, as it provides the insights into wind characteristics required to predict the potential energy output. The central aim of WRA is to produce a comprehensive reconstruction of the wind resource in both space and time. This involves mapping the wind field across the entire area of interest and characterizing how the resource evolves over time, from short-term turbulent fluctuations to seasonal patterns and inter-annual variability. Given that the typical operational lifetime of a wind farm is approximately 20 years [2], it is crucial that WRA provides a robust insight into the potential wind energy over such an extensive period.

A precise and comprehensive WRA is of paramount importance for multiple reasons. First, wind farm profitability depends strongly on the local wind resource [3] and therefore pre-construction energy yield assessments underpin project financing [2]. Even a 1% reduction in energy uncertainty can save millions of dollars, and a 1% change in wind speed uncertainty can shift project net present value by 3% to 5% [2]. Furthermore, WRA is essential for identifying suitable locations and for supporting informed decisions about turbine type and scale, which helps maximise energy output and ensures long-term performance [3]. Additionally, it provides essential data for national energy planning, guiding infrastructure siting and ensuring that sustainability goals align with realistic resource potential [3].

The WRA relies on a nested modeling hierarchy in which each tier is initialized by the level above [4]. This process begins with coarse global reanalysis data (typically ~ 30 km resolution, e.g., the Fifth-generation ECMWF atmospheric reanalysis dataset, ERA5), which feeds into regional mesoscale models (~ 2-15 km resolution) for dynamic downscaling. The final step involves microscale models. While mesoscale models remain the industry standard, this research focuses on the cutting-edge use of Large-Eddy Simulation (LES) (~ 50-250 m resolution), a technology made economically viable by companies like Whiffle due to the utilization of GPU-optimized atmospheric solvers [5]. LES models explicitly resolve turbulent structures and terrain interactions on grids of tens to hundreds of meters.

As each tier uses the boundary conditions provided by the one above, any existing errors or biases are carried through the entire hierarchy. For reanalysis data the main causes of bias are coarse horizontal resolution and incomplete assimilation, which prevent the model from resolving speed-up over ridges or roughness transitions [6, 4]. Mesoscale models introduce their own deterministic errors through imperfect parameterizations (planetary-boundary-layer schemes, land-surface models, microphysics), sub-optimal data assimilation and misspecified lateral forcing [7, 4]. LES is typically limited by simplified surface parameters and by its inability to resolve eddies below the grid scale. A post-processing step is therefore required to correct the simulated wind conditions.

1.1. Problem Statement and Research Gaps

The critical challenge lies in performing the required correction of the simulation data when the only source of truth is scarce. While the LES provides a full dataset that is spatially complete (a grid covering the entire site) and temporally continuous (a long-term time series), the validation data from met-masts and LiDAR is severely limited. Physical measurement campaigns provide the ”ground truth” but are spatially sparse, covering only a few specific locations across a potentially large site, and temporally limited, often spanning a short period, sometimes less than one year. The fundamental task becomes a complex spatio-temporal mapping problem: determining how to use a minimal, localized, and short-term ground-truth dataset to globally and temporally error-correct the entire high-resolution LES wind field. This dual challenge requires novel approaches to extrapolate corrections accurately both through space and time.

WRA relies on error-correction techniques to bridge the gap between modeled and measured wind conditions. Two standard approaches dominate current industry practice: Measure-Correlate-Predict (MCP) for temporal correction and Inverse Distance Weighting (IDW) for spatial extrapolation when multiple observation locations are available. While both methods are straightforward and robust, their simplifying assumptions limit their accuracy.

MCP is the standard approach for extending short-term on-site measurements to long-term reference datasets. It relies on the assumption of a stationary linear relationship between the local (measured) and reference (modeled or nearby) wind speeds. Although originally formulated as a simple linear regression, MCP has been extended to include more expressive transfer functions using models such as Artificial Neural Networks (ANN), Support Vector Regression (SVR) and Random Forests (RF). However, these methods still rely only on pointwise reference inputs and do not attempt to exploit the spatial or temporal flow structure present in high-resolution models such as LES.

IDW is a widely used interpolation/extrapolation method for estimating wind conditions at unmeasured locations based on nearby observations. It assumes that similarity decreases purely with geographical distance, ignoring terrain effects and the spatial coherence of the wind field. In practice, distance is the only prior used when weighting the contribution of each observation point. While this simplicity ensures robustness and ease of application, it also leaves room for improvement, especially when high-resolution flow information from LES is available.

While MCP and IDW remain central to current workflows, they have been developed and applied primarily using reanalysis data, mesoscale NWP data and conventional meteorological observations as reference sources. With the emergence of high-resolution LES, it is necessary to verify their validity at this scale and to explore how they can be adapted to exploit the richer spatial and temporal information now available. The following research gaps outline where current practice falls short and where this thesis aims to contribute.

Gap 1 – Lack of validation of MCP on microscale data.
Despite its central role in WRA, MCP has not been evaluated using microscale (LES-based) reference data. This is likely because LES is not commonly used for long-term WRA, due to the increased cost compared to coarser alternatives. Existing MCP studies rely exclusively on reanalysis, mesoscale NWP data or nearby meteorological stations. As a result, it remains unknown whether the standard MCP assumptions hold at microscale resolution, whether LES time series exhibit different correlation behavior, and whether MCP can correct the systematic discrepancies between LES output and observations. Establishing this baseline is essential before designing improved temporal correction methods for LES.

Gap 2 – Incorporating temporal and spatial context from the LES into MCP-based temporal correction.
The standard temporal correction approach depends only on the modeled wind speed at the specific location and specific time, ignoring valuable temporal and spatial dependencies in the LES data. In reality, wind evolution is continuous and structured: neighboring time steps and the surrounding area carry information about the same atmospheric processes. Leveraging this context could enable a more

informed and accurate temporal correction, but current WRA practices do not exploit this potential.

Gap 3 – Exploiting instantaneous LES flow structure for IDW-based spatial extrapolation. The standard spatial extrapolation method, IDW, uses only distance as a similarity measure and ignores the flow structure resolved by the LES. The availability of a high resolution wind field at every timestep, given by the LES, offers a unique opportunity to inform spatial correction using modeled flow features, such as wind speed. Incorporating these spatial structures could yield more realistic and effective correction maps that follow the actual flow dynamics rather than geometric proximity alone.

The gaps identified above highlight the limitations of current correction practices when applied to high-resolution LES data. Together, they point to the need for a systematic investigation into how both temporal and spatial correction methods can be adapted and improved for use at the microscale. Addressing these gaps forms the basis of this thesis and leads to the following central research question:

> *How can data-driven methods improve LES wind-speed error correction in both the spatial and temporal dimensions within the context of Wind Resource Assessment?*

This overarching question combines two complementary challenges. Firstly, correcting systematic LES errors in time at locations with measurements (temporal extrapolation). Secondly, extending these corrections to the rest of the domain (spatial extrapolation). Together, they form the foundation of a complete LES correction framework.

1.2. Research Contributions

The temporal extrapolation component focuses on improving the correction of LES wind-speed predictions at the observation locations. A collection of linear and non-linear Machine Learning models, namely Linear Regression (LR), Ridge Regression (Ridge), Multi-Layer Perceptron (MLP), Support Vector Regression (SVR), and Extreme Gradient Boosting Regressor (XGB), are employed as flexible baseline models, forming the foundation for further methodological extensions. The performance of these models when given mesoscale and LES data as input is compared, validating the existing MCP-style approach for the microscale LES.

Building on these baselines, additional context is introduced to enhance correction accuracy. Temporal context is incorporated through multi-step time windows, allowing the model to capture short-term wind evolution, while local spatial context is introduced by including data from neighboring LES grid cells. To capture larger-scale atmospheric structures, the temporal correction model is further extended with encodings of the full LES wind-speed field, obtained through a Convolutional Autoencoder (CAE). This enables the model to represent site-wide flow patterns in a compact latent space. Each of these strategies is evaluated across multiple models and observation locations to quantify their individual and combined contributions to performance and generalizability. The most effective strategies are subsequently integrated into a single combined correction strategy, demonstrating a significant improvement over industry-standard methods.

The spatial extrapolation component extends the corrected wind speeds from the observation locations to unmeasured areas across the domain. To improve upon the traditional IDW method, a Wind-Speed Enhanced IDW (WS-IDW) approach is introduced, which incorporates flow-aware weighting by including LES wind-speed similarity in addition to geographic distance. This modification allows the extrapolation process to follow physical flow structures more closely rather than relying solely on spatial proximity. The WS-IDW method is evaluated against the baseline IDW across multiple sites and for varying numbers of source locations to assess its generalizability and robustness. Complementary variants are also tested, incorporating additional features such as wind direction and terrain characteristics (elevation and tree canopy height) to evaluate their potential in improving spatial extrapolation performance.

Throughout both the temporal and spatial analyses, standard industry methods (MCP and IDW) serve as empirical baselines. Their performance on LES-scale data establishes a necessary reference point against which all proposed enhancements are evaluated.

1.3. Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 surveys the data sources and modeling frameworks essential to Wind Resource Assessment, detailing the operation of Large-Eddy Simulation (LES) and established correction techniques such as MCP and IDW. Chapter 3 outlines the methodology, detailing the datasets, experimental setup, and implementation of both baseline and proposed methods for temporal and spatial extrapolation. Chapter 4 presents the experimental results and quantitative performance comparisons. Chapter 5 discusses the findings and addresses key limitations. Finally, Chapter 6 concludes the work, summarizing the main outcomes and identifying limitations and opportunities for future research.

2

Data and Methods for Wind Resource Assessment

Wind resource assessment in practice is constrained by the quality and coverage of the available data. A potential site can cover a large area, with the biggest offshore wind park (Hornsea 2) and the biggest onshore wind park (Gansu Wind Farm) covering 472 km^2 [8] and 700 km^2 [9] respectively, while measurements are usually collected in only a few locations. Measurements are thus unavailable for the majority of the area being considered for development. Compounding this is the common limitation of temporal data. Long-term, high-quality measurement campaigns, ideally spanning multiple years, are crucial for capturing the full spectrum of wind conditions, including both seasonal patterns and inter-annual variability. However, such extensive campaigns are not always feasible due to cost and logistical constraints. Relying on data from shorter periods, particularly those less than a full year, can introduce significant biases [6]. The specific season during which these abbreviated measurements are taken can lead to a substantial over- or under-estimation of the long-term wind energy potential [6]. Moreover, the wind resource itself is subject to considerable inter-annual variability, often driven by larger climatic oscillations. This can cause the actual annual energy production (AEP) to fluctuate, commonly by ±10-20% around the long-term average, and potentially more in specific years [2, 3].

To mitigate the significant uncertainties introduced by these inherent limitations, effective WRA relies on reconciling two complementary data categories: numerical models, which are spatially and temporally complete but inherently imperfect, and physical measurements, which serve as the ground truth but are scarce. Section 2.1 details the characteristics of these diverse data sources and modeling frameworks. Subsequently, Section 2.2 discusses the correction and extrapolation techniques designed to bridge the gap between them, utilizing sparse observations to learn and rectify systematic errors within the complete model simulations.

2.1. Atmospheric Data Sources and Model Frameworks

The estimation of wind resources relies on a hierarchy of data sources and modeling techniques, ranging from direct on-site measurements to sophisticated numerical simulations of atmospheric flow. Each approach offers distinct advantages in terms of accuracy, spatial-temporal coverage, and cost, and they are often used in conjunction to build a comprehensive understanding of the wind climatology at a potential site.

2.1.1. Observational Data

Direct wind measurements collected at or near the site provide reliable, site-specific information. This observational data is essential for validating models, training forecasting systems, and conducting detailed WRA [10, 11, 12, 13, 14, 15, 16, 7, 17].

Meteorological (Met.) Masts

A meteorological mast is a physical tower-like structure erected at a relevant location to provide stable platforms for mounting various meteorological sensors at multiple heights above ground level [11, 13].

These sensors are designed to collect point measurements of crucial atmospheric variables, primarily wind speed and direction, but often including temperature, atmospheric pressure, and relative humidity [11, 13, 18, 19]. Measurement campaigns typically involve deploying met masts for a period sufficient to capture site-specific wind conditions, often around one year [10, 6, 14, 7]. Data is commonly collected at a high temporal resolution, such as 10-minute or 15-minute intervals, before being averaged to hourly values for use in analysis and modeling [10, 11, 6, 13, 19, 20, 7]. High data availability, ideally exceeding 90%, is generally considered necessary for the time series to be representative [10, 7].

Various types of sensors are mounted on these masts [11, 19]. Cup anemometers [19, 21, 14, 17] and sonic anemometers [19, 21, 17] are commonly used to measure wind speed. Sonic anemometers, in particular, can offer very fine temporal resolution and are suitable for long-term use, even in challenging conditions, as they have no moving parts [19]. Wind vanes measure wind direction [19, 17]. Measurements are taken at different heights [10, 11, 21, 7, 17]. Historically, a standard height was around 10 meters above ground level [19, 21, 14], but with wind turbines becoming larger [19] measurements closer to typical turbine hub heights (e.g., between 100 and 140 meters [10]) are increasingly relevant [6, 19]. Measurements taken at much lower heights require extrapolation, which introduces additional uncertainties [19]. Despite the challenges and costs associated with their construction and maintenance [13], especially in remote areas [11, 7], and the fact that they only provide data at one location, met masts are considered the most direct and effective approach for characterizing the wind field of a site [12]. They are indispensable for training and validating model simulation results [11, 13, 21].

LiDAR

LiDAR is a remote sensing technique that uses laser beams to measure wind speed and direction by detecting the Doppler shift from backscattered light off aerosols [13, 19, 21, 16, 22]. It can provide high vertical resolution profiles of the wind across the turbine rotor-swept area [21, 16, 22] and is increasingly used as an alternative or supplement to met masts. LiDAR systems are highly sensitive and can measure wind speeds up to several kilometers away, depending on atmospheric clarity and instrument configuration [13]. LiDAR data is valuable for Measure-Correlate-Predict (MCP; Section 2.2.1) methods and model validation [21, 16, 22, 23]. However, raw data requires extensive filtering which can significantly reduce the amount of recovered data finally used [22]. Performance in complex terrain can be particularly challenging [23], potentially leading to larger errors in wind speeds than over flat terrain [21, 22]. Studies have reported LiDAR error rates ranging approximately between 2% and 7% depending on terrain complexity evaluated by the Ruggedness Index (RIX), which can stem from the curvature of streamlines over terrain [22]. Despite this, wind speed and direction data from LiDAR can match well with data from met masts, especially on flat terrain [22].

SCADA

Installed on operational wind turbines, SCADA systems record numerous parameters, including nacelle-measured wind speed and direction, power output, and other operational data, typically as 10-minute averages [24, 25, 26, 13, 27]. While offering a cost-effective source of long-term operational wind data [7], SCADA data quality can be compromised by noise, missing values, and sensor inaccuracies, often requiring significant preprocessing and correction (e.g., Nacelle Wind Speed Transfer Functions) to approximate free-stream conditions [25, 7, 27].

2.1.2. Mesoscale Numerical Weather Prediction Models

Numerical Weather Prediction (NWP) models constitute a fundamental class of physical models employed in atmospheric forecasting [24]. These models operate by computationally solving complex physical equations that describe atmospheric processes. In the context of wind power NWP models serve as a crucial and indispensable data source for forecasting applications, particularly for short-term horizons [11, 24, 26, 28, 29]. NWP models provide forecasts for a range of meteorological variables essential for wind energy applications, such as wind speed, wind direction, temperature, humidity and barometric pressure [29]. This data is typically available at defined spatial and temporal resolutions, often ranging from kilometers (e.g., 1 km, 3 km, 4 km, 18 km; these are called mesoscale models) [30, 11, 6] to tens of kilometers (e.g., 23 km, 31 km, 55 km) [19] spatially, and with temporal outputs ranging from 10 minutes [11, 6] to hourly [30, 6]. NWP models are often computationally expensive, especially when aiming for higher resolutions [12, 26, 30].

A characteristic of NWP data is the presence of inherent uncertainties and biases in the forecast outputs [31, 11, 26, 32, 29]. These errors can stem from various factors including simplifications within the model formulation, inaccuracies in initial measurements, or the representation of surface characteristics. The accuracy of NWP forecasts is notably affected by aspects such as the specific type of forecasting system used, the forecast lead time, the spatial scale of the prediction, and the occurrence of extreme transitional weather events [11, 29]. Overall, the resolution and accuracy of NWP data are considered limited, and substantial technical breakthroughs in these areas are not anticipated in the short term [24]. Studies indicate that NWP errors exhibit patterns, often linked to the season and specific wind speed and direction conditions [10, 11, 26]. An example of this is given by [11], where certain months, like June, exhibit higher Mean Absolute Error (MAE) in raw NWP data than the other months. NWP wind speed errors are identified as a key factor impacting the accuracy of short-term wind power production (STWPP) [29]. The residual series, representing the error between NWP forecasts and measured wind speeds, also demonstrates time series continuity and predictability [29].

Despite these inherent limitations, NWP models remain an essential component for wind forecasting [31, 11, 12, 24, 26, 29] and hindcasting [6, 33, 13]. Given the acknowledged limitations and the critical role NWP plays, improving wind power prediction accuracy often necessitates addressing the quality and inherent errors of the NWP data [24]. Consequently, NWP wind speed correction is recognized as an effective strategy to enhance Wind Power Forecast (WPF) accuracy, focusing on the data itself rather than solely refining the prediction model [24, 29]. Furthermore, [29] suggests that using corrected NWP data can significantly improve prediction accuracy compared to using uncorrected data, demonstrating the positive impact of NWP correction, leading to improvements in evaluation metrics such as R^2 and reducing the spread of data distribution.

2.1.3. Global and Regional Reanalysis Products

Reanalysis products are retrospective gridded reconstructions of the atmosphere, generated by rerunning a frozen Numerical Weather Prediction system while continuously assimilating the full archive of historical observations [4]. Flagship data sets [7] such as ERA5 [34] and MERRA-2 [35] now span multiple decades (ERA5 from 1950, MERRA-2 from 1980) at hourly cadence, providing globally consistent [19, 4] meteorological variables, including wind speed, at multiple heights. Horizontal resolution for ERA5 is 0.25° x 0.25° (approximately 31 km), while for MERRA-2 it is 0.5° x 0.625° (approximately 55 km). ERA5 uses 137 vertical model levels, and MERRA-2 uses 72 vertical levels. These resolutions are substantially finer than earlier ERA-Interim-class products [19, 4] (ERA-Interim is 0.75° (79 km) with 60 vertical levels) but still coarse compared with turbine-scale flow.

Reanalysis datasets are valued for their long-term temporal consistency, global coverage, and completeness, often spanning more than 30 years with hourly datapoints [4]. Their construction scheme, using a fixed NWP system and data assimilation, ensures homogeneity over time, making them well suited for studies requiring climatic stability [10, 4]. Some reanalyses, such as ERA5, also provide wind speed estimates at or near turbine hub heights. However, these advantages come with inherent limitations. The spatial resolution, though improving, remains too coarse to resolve fine-scale terrain effects or local atmospheric variability, particularly in complex or mountainous regions [19, 7, 17, 4]. This smoothing leads to biases in near-surface wind speeds, especially where land-sea contrasts or steep topography are present. Errors in reanalysis data arise from model assumptions and assimilation limitations, and they can vary by region, season, and time of day [4]. Additionally, reanalysis products often lack explicit uncertainty estimates [4], and their accuracy compared to local observations is still not fully understood [19, 4], particularly for inter-annual variability or extremes. As a result, while reanalysis provides a consistent and accessible baseline, it often requires bias correction [36, 24] or downscaling [13, 4] before it can be applied with confidence to site-specific wind assessments.

Within wind-energy workflows, reanalysis datasets serve a range of complementary functions. They supply the long-term reference series used in MCP and related climatological transfer methods, where month- to decade-scale ERA5 distributions are mapped to short on-site campaigns to produce bankable 20- or 30-year wind statistics [6, 19]. They also provide boundary and initial conditions for regional or microscale model nests, which inherit large-scale atmospheric patterns while resolving site-specific terrain effects [4]. Furthermore, reanalysis data has been used as input for machine-learning downscaling frameworks, alongside high-resolution WRF simulations, to train deep learning models that perform spatial super-resolution, enabling the reconstruction of fine-scale wind fields over extended historical periods [13]. WRF stands for Weather Research and Forecasting and it is a mesoscale NWP system. In addition, the temporal consistency of reanalysis products makes them valuable for detecting long-term climatic variability [6, 4].

Emerging high-resolution regional reanalyses (e.g. COSMO-REA6 [37] at 6 km) promise incremental gains, yet remain computationally intensive and still require site-level down-scaling. Consequently, most modern pipelines treat reanalysis as a baseline to be corrected, either statistically, through methods discussed in Section 2.2 or dynamically via nested NWP runs, before the data can drive reliable energy-yield assessments.

2.1.4. Microscale Flow Modeling and LES

Microscale flow models occupy the scale between mesoscale weather forecasts and point measurements, solving the three-dimensional Navier-Stokes equations on grids with resolutions of tens to a few hundred metres in order to translate large-scale wind climates into turbine-level fields. They range from steady Reynolds-Averaged Navier-Stokes solvers, which use turbulence closures to approximate the unresolved eddies and can deliver site-wide speed-up and roughness corrections within minutes, to time-resolved Large-Eddy Simulations that explicitly resolve the energy-carrying turbulence and thus reproduce shear, wake interaction and stability effects with much higher fidelity [38, 39]. Hybrid frameworks such as PALM couple an LES core to land-surface, radiation and urban canopy schemes so that terrain, forest or building heterogeneity is represented consistently across diurnal cycles [40]. All of these models depend on accurate boundary conditions and surface data, often sourced from reanalysis or mesoscale nests, and their practical value is determined by the trade-off between computational cost and the level of turbulence detail required for the engineering task at hand.

Whiffle employs a high-resolution Large Eddy Simulation (LES) model called GRASP (GPU-Resident Atmospheric Simulation Platform) for simulating detailed wind fields. GRASP is derived from the Dutch Atmospheric Large Eddy Simulation (DALES) framework [41], optimized to run efficiently on Graphics Processing Units (GPUs) [42]. This allows LES to be applied to multi-year simulations at spatial resolutions of approximately 100m × 100m horizontally and 25-64 m vertically, producing 4D wind field outputs across the full rotor-swept area [5].

The main advantage of LES over traditional mesoscale or reanalysis models lies in its ability to resolve turbulent structures directly, rather than relying on parameterizations. This yields more accurate representations of wind shear, turbulence intensity and wake effects. These are critical inputs for wind energy applications, especially in complex terrain or offshore environments.

To reduce computational costs, GRASP uses a two-domain system: a precursor LES driven by coarse-scale ERA5 weather data provides turbulent inflow conditions to a finer-resolution cursor LES over the area of interest [5, 43]. Despite these optimizations, LES remains computationally expensive compared to mesoscale models, requiring powerful hardware and significant simulation time. Nonetheless, the accuracy gains make it a valuable tool for high-fidelity wind resource assessments, turbine siting, and wake modeling.

2.2. Correction and Extrapolation Techniques in WRA

There are 3 spatial and 1 temporal dimensions in which to correct the 4D high resolution LES model output. The observation data, which is spatially extremely sparse and temporally limited to a short period, serves as a target for the correction and can be used in evaluating and validating correction methods.

Spatial extrapolation and temporal extrapolation are usually treated as separate challenges and tackled sequentially. Spatial extrapolation techniques take as input biases for a limited number of locations, and extrapolate this bias to all locations of a site. Temporal extrapolation techniques are location-specific, meaning they learn a bias correction for one location and are subsequently able to bias correct unseen forecasts for that location. A schematic of this sequential workflow is illustrated in Figure 2.1, assuming a one-year measurement campaign and a 20-year hindcast period.

Some methods aim to address both dimensions simultaneously, although this is rare. Section 2.2.1 considers site-based methods that can be used for temporal extrapolation. Section 2.2.2 considers methods for spatial extrapolation in the horizontal plane while section 2.2.3 discusses methods used for joint spatio-temporal extrapolation.

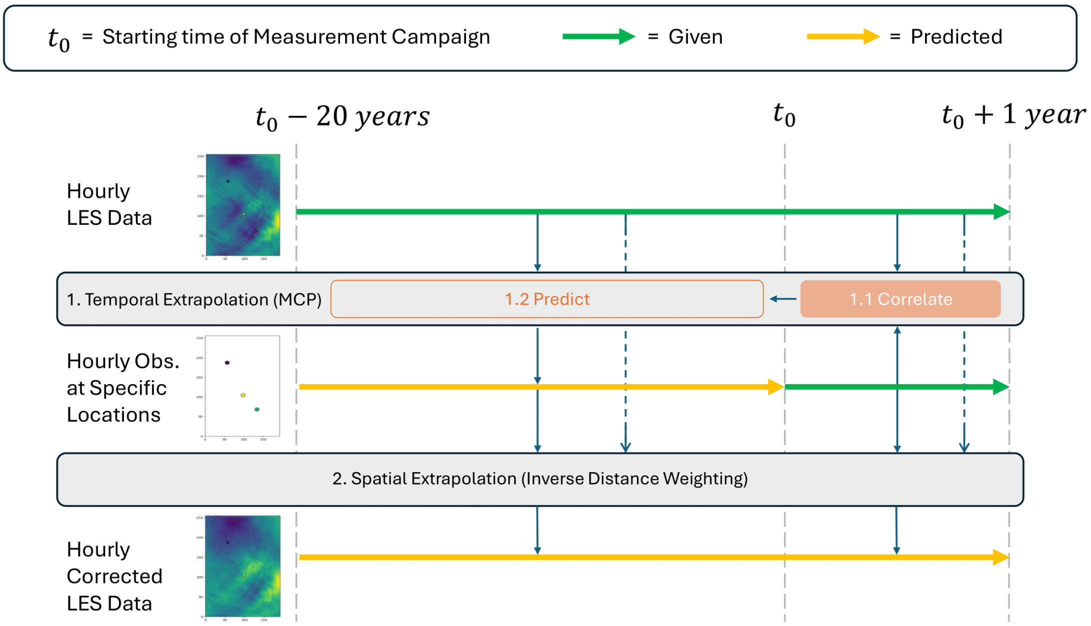


Figure 2.1: Schematic representation of the sequential two-step correction framework. The upper section illustrates Temporal Extrapolation via Measure-Correlate-Predict (MCP), where a transfer function is derived from the concurrent measurement and reference periods (Box 1.1) and subsequently applied to hindcast the long-term wind speed time series (Box 1.2). The lower section details Spatial Extrapolation, in which the temporally corrected observations serve as the input to distribute local bias corrections across the spatially complete LES domain.

2.2.1. Temporal Extrapolation Methods

Temporal extrapolation addresses the gap between short on-site measurements and the multi-year period required for WRA. Its purpose is to relate limited observations to long-term atmospheric behavior in a consistent and reliable way. Several approaches exist, each making different assumptions about how short-term data can represent long-term conditions.

Temporal extrapolation can serve two distinct objectives within WRA. The traditional objective is to reconstruct the long-term wind climate at a site, typically expressed through aggregated statistics such as the mean wind speed, Weibull distribution parameters, wind roses, and turbulence characteristics. Measure-Correlate-Predict (MCP) was initially developed for this purpose. In contrast, a second and increasingly relevant objective is to reconstruct the historical wind-speed time series itself with high timestep accuracy, evaluated using metrics such as MAE, RMSE and R^2. Time-resolved correction provides insight into when specific atmospheric conditions occur, which is essential for applications that depend on temporal dynamics, including diurnal electricity pricing, curtailment during bird migration periods, and hybrid wind-battery optimization. Advances in high-resolution modeling and machine learning have made this time-series-focused approach more feasible, although it remains the less common approach in industry practice. The present thesis adopts this second perspective and evaluates temporal correction methods in terms of their ability to produce accurate, time-resolved wind-speed predictions.

Measure-Correlate-Predict (MCP)

MCP is the standard remedy when developers only have data from a brief on-site measurement campaign but information for a longer time period is required, and is thus widely studied and applied [10, 30, 6, 19, 20, 14, 15, 16, 7]. Typically, a few months to several years of measurements (M - measure) collected at the target site are paired with a much longer archive, often covering decades, from a reference data source. The reference data source can be a nearby meteorological mast or station [30], turbine nacelle, reanalysis data [10, 30, 19, 7], or mesoscale simulation [30, 7]. A transfer function is learned (C - correlate) from the period for which both measurements and reference data is available. Subsequently, the learned transfer function is applied to the full reference data period, stretching the short measurement campaign into a multi decade hindcast (P-predict). This newly acquired long term data at the target site can be used for the EYA and the P-value analysis required by lenders [30, 19].

Correlation Methods
MCP's long history has produced a broad array of correlation methods, and the selected correlation method can strongly influence the performance. The idea of stretching a short on-site record to climatological scale can be traced to [44] published in 1948, which already suggested rescaling local observations with statistics from a longer off-site series. Almost thirty years later [45] set the first explicit cross-correlation frame by showing that target-site anomalies could be reproduced by multiplying concurrent reference anomalies by the ratio of their long-term standard deviations, an elegant linear correction that quickly became a workhorse for early MCP applications. [46] introduced sector-split ordinary least squares after which it was popularized by [47]. Its slope and intercept have intuitive physical meaning and its residuals are easy to audit, which explains why it still serves as the benchmark in commercial software [48].

Ordinary least squares performs robustly when the overlap spans a full seasonal cycle, but its variance dampening motivated the Variance-Ratio (VR) model introduced by [49] in 2005. By forcing the fit through the origin and scaling only the variance, they preserved turbulence intensity and cut root mean square error by up to 15% at several North American sites [49]. Linear-with-residuals soon followed [50], re-injecting stored residuals into the hindcast so that high-frequency variance is retained, while power-law and low-order polynomial transfers were tested at sites where stability or flow curvature makes a single slope inadequate, although gains proved inconsistent and the risk of distorting extremes remained. At the same time practitioners refined the direction-sector concept, merging under-populated bins or letting clustering algorithms decide sector boundaries directly from the wind vector itself [51], an approach that is now embedded in several industrial tools, among which WindPRO and WindoGrapher.

The next substantive leap was the introduction of ML techniques. Addison's neural-network prototype in 2000 [52] was among the first to apply ML to this problem. Many followed, applying ANNs [53, 21, 20, 14, 23], Support Vector Regression (SVR) [54, 20, 23], Random Forest (RF) [20, 55, 23, 7] and others [6]. These reliably trim root mean square error by 5-40% relative to LR or VR fits. Multiple studies have compared a number of ML methods used in MCP [20, 7, 55, 23, 16] but results vary and no single learner dominates, likely because performance still hinges on terrain complexity, overlap length and hyper-parameter tuning. Research into feature selection methods when performing ML-based MCP using multiple reference stations found that a feature subset lowered mean errors in at least 80% of cases, although statistically significant improvements were observed in only up to 10% of cases analyzed [20, 14].

The correlation step in MCP thus spans a spectrum from simple to complex: at one end there is the transparent ordinary least squares, which remains the decent default; at the other, data-driven ML methods which automatically select features and return full predictive distributions.

Choice of reference data
The choice of reference data is equally consequential. A well-maintained mast within 20-30 km and at a comparable height still delivers the tightest correlations, but such masts are not always available and often contain measurement gaps [7]. Multiple reference stations can sometimes be used to capture additional information, compared to using a single station, but no best-practice has emerged for choosing which reference sources to combine [14]. Reanalysis products such as ERA5 or MERRA-2 provide global, hourly coverage back to 1950 or 1980, making them attractive MCP anchors once seasonal orographic errors have been corrected [10, 19, 4].

However, recent studies indicate that utilizing high-resolution mesoscale models as long-term references can significantly improve MCP accuracy compared to global reanalysis data, particularly when capturing local terrain effects or coastal gradients [30, 7].

The logical extension of this trend is moving from mesoscale to microscale models, but this remains undocumented in literature, likely due to the prohibitive computational cost of running such models over decadal timescales. This computational challenge is partly solved by Whiffle's exploitation of GPUs and can be further dealt with by only running a selection of representative days for each year [56]. As using microscale models in MCP becomes feasible, this might unlock a next tier of MCP accuracy.

Limitations
The simplicity of MCP comes with several limitations. A concurrent period shorter than six months can

produce a poor fit, and a campaign confined to a season with high or low wind speed projects that seasonal bias forward into the 20-year mean unless explicitly corrected [10]. The core assumption that the transfer function is time-invariant is threatened by climate trends and teleconnections [6], yet routine workflows still treat it as static. Contrary evidence also exists, with [13, 57] investigating the effect of climate change on wind conditions in the coming decades and finding a subtle or even irrelevant impact.

Furthermore, error embedded in the reference that is not completely captured during the measurement period, whether from reanalysis bias or mast maintenance drift, propagates unfiltered into the hindcast. According to [6] future research for MCP methods should primarily focus on developing methods for uncertainty quantification and addressing the non-stationarity of wind speed time series, which challenges the fundamental assumptions of the MCP approach. Further advances aim to shrink the persisting ca. 6-10% uncertainty band that still envelops long-term mean wind-speed estimates at many projects [6, 19, 16].

Bias-correction techniques

Various of the methods developed for bias-correction produce transfer functions that can be used in the context of MCP. Although from a technical point of view they are closely related, the goal of both methods is different, with bias-correction also focusing on forecasting situations and altering the in-sample distribution while MCP is a purely hindcasting method and generalizes to out-of-sample timesteps.

Bias-correction techniques obtain bias estimates by comparing model outputs with corresponding observations for a chosen location. These bias estimates can subsequently be used to correct for the systematic bias contained in the model. A wide variety of post-processing schemes has been proposed for doing this, of which a selection will now be presented.

The simplest approaches impose deterministic adjustments to the forecasts so that their mean, and, in some cases their variance, match those of the observations. Linear scaling, for example, multiplies every forecast by the ratio between a reference observational dataset and the model mean, and was shown to improve ERA5s long-term means over complex terrain while leaving sub-daily variability untouched [6].

A related but more comprehensive transformation is the average-variance-trend (AVT) method, which first detrends the series, then rescales its anomalies so that mean, variance and linear trend match the observations before re-introducing the trend [36]. At three Chinese wind-farm sites AVT reduced percentage bias by roughly 20% relative to raw 3 km WRF output.

Ordinary least-squares Model Output Statistics (MOS) delivers similar one-time coefficients and was among the four correction methods (short-term bias, diurnal bias, directional bias and linear regression) assessed at thirteen Indian masts, where each individual method cut root-mean-square error by up to 55% compared to the numerical prediction alone [11]. Diurnal bias correction assigns distinct hourly offsets that remove systematic intra-day errors, while directional bias correction applies sector-specific adjustments based on wind direction, tailoring the correction to flow orientation [11, 58]. Rolling thirty-day means or hourly and wind-sector tables outperformed static offsets in seasons with marked regime changes, indicating the benefit of recent measurements being available, as in a forecasting context.

A similar set of corrections, including short-term, diurnal, mean-variance scaling, orthogonal regression, Kalman filtering and artificial neural networks (ANN) was evaluated by [58] in Ireland. Mean-variance scaling rescales forecasts so that their mean and variance coincide with the observations but, unlike AVT, leaves the linear trend unchanged, providing a simpler two-parameter correction.

[59] proposes a method for creating a correction for a location by subtracting the model's climatological mean from the predicted variable and adding the real climatological mean. The real climatological mean can be derived from historical observation or reanalysis data. This method requires low computational effort.

Quantile mapping aligns the full cumulative distributions of forecasts and observations by mapping each model quantile to its observational counterpart, thereby correcting biases across all probability levels [57].

Machine learning approaches have been applied to this challenge. An ANN, as used in [11] and [58], learns a nonlinear mapping from recent forecast-observation pairs and can correct complex systematic flow-dependent errors. [60] applied Random Forest (RF) and MOML, a ML based alternative of MOS.

Apart from using increasingly clever and complex bias correction methods, gains were obtained by blending the competing corrections. Inverse RME weighted average, multi-linear regression, combinatorial random forest, combinatorial (extreme) gradient boosting and combinatorial K-nearest neighbors were all explored in [11]. In [58] a mean-squared-error weighted combination surpassed every constituent stream at almost all station-resolution pairs while eliminating mean bias entirely.

However, only a subset of the foregoing techniques is directly transferable to the hindcasting context examined in this thesis, where a limited period of overlap is available for training and the subsequent target period lack observations. Corrections that can be frozen after the calibration stage, among which static linear scaling, AVT and fixed-coefficient MOS, remain valid for every out-of-distribution timestep because they apply an invariant algebraic mapping. Methods that depend on sliding windows or recursive filters require contemporaneous observations to update their coefficients and therefore cannot be operated in their original form as presented in [11] and [58]. A conceivable workaround is to retain the parameters estimated on the final day of the overlap and apply them unchanged throughout the hindcast, but earlier studies indicate that most of the skill of short-term, diurnal or directional bias schemes derives from continual adaptation, so such freezing would forfeit a substantial part of the expected error reduction [11, 58].

Analog Ensemble (AnEn)

The Analog Ensemble (AnEn) method adopts a local, non-parametric approach that contrasts with the formulation of MCP. Instead of learning a single regression rule, AnEn treats correction as a dynamic nearest-neighbor search within the historical archive, reconstructing the time series by identifying past atmospheric states that resemble current conditions. It thus provides a distinct, albeit less standardized, alternative to MCP.

First introduced by [61], AnEn converts a coarse forecast into a probabilistic prediction by retrieving analogous forecasts from a fixed archive [33]. For any new forecast, the algorithm identifies the most similar historical model states utilizing a weighted Euclidean distance. The observations associated with these selected analogs form the ensemble prediction, ensuring that the result is automatically bias-adjusted and preserves naturally occurring spatial-temporal error correlations [33].

The method retains exceptional computational economy because it relies on multivariate distance calculations rather than additional complex model integrations [62]. However, its performance depends on the availability of a sufficiently long and representative archive. To mitigate issues arising from data scarcity, recent innovations include patch-based analog construction to expand the effective search space [63] and autoencoder-based synthesis of realistic ensemble perturbations to populate under-represented portions of the state space [64].

Comparing the two approaches, a benchmark of thirty-five sites found that while MCP maintains a slight advantage in strongly correlated regimes, AnEn offers greater adaptability. After optimizing analog-ranking weights, AnEn achieved the highest R^2 at 60% of the sites, particularly outperforming MCP in offshore conditions or when the target-to-reference correlation is lower [19].

2.2.2. Spatial Extrapolation Methods

The task of wind resource assessment also contains the challenge of spatial extrapolation from the locations for which measurements are available, which are few and sometimes far between, to the entire site under consideration. This spatial extrapolation is vital in performing a WRA, and is often used alongside a distinct temporal extrapolation step. It plays a central role in blending high-quality but sparse ground-truth data with dense but biased numerical model output.

Horizontal Extrapolation

When multiple measurement locations are available they each provide a local correction, and for any unseen target location these corrections must be combined into a single estimate. Horizontal extrapolation methods perform this spatial weighting by constructing a continuous correction field from the sparse pointwise information.

A widely used and straightforward method is inverse-distance weighting (IDW). Each grid point is assigned a weighted average of the values observed at nearby mast locations, where the weights decrease with distance [65]. Despite its simplicity and lack of statistical optimization, IDW remains an industry standard due to its transparency, speed and ease of implementation, and is featured in widely used software such as OpenWIND [66].

A precursor to any spatial weighting step is often the application of a static flow model, such as a linearised flow solver or RANS-CFD, which provides sector-wise transfer functions from each measurement location to the target point. When multiple masts are available this produces several mast-based predictions for the same target location, and IDW is then used to blend these predictions into a single estimate rather than combining the correction factors directly.

The challenge of spatially extrapolating sparse measurement is a challenge that appears in many domains, and for which many methods have been developed. A selection of such existing methods, some of which have previously been applied to the relevant domain, will now be given.

Kriging is another long-standing method, originating in 1978 [67], and remains widely applied. It treats the field as a realisation of a spatial random process with a defined covariance structure, and derives weights that minimise the expected extrapolation error under that model. Ordinary Kriging (OK) uses only distances and assumes a constant mean, while universal kriging allows for a spatial trend. Co-kriging generalises this to include secondary variables such as elevation or roughness, enabling the use of terrain information to improve extrapolation quality [68].

Thin-plate splines (TPS) offer a smooth alternative by fitting a continuous surface that exactly matches the known correction values and minimises the integrated second derivative (i.e., bending energy) [69, 70]. The result is a differentiable correction map that spreads mast information smoothly without requiring an explicit variogram model. TPS has been applied in wind mapping over complex terrain due to its ability to preserve large-scale gradients while avoiding artificial artefacts between mast locations.

BCDG is an extension, proposed in [71] of the BCG method that can be used for spatial inter/extrapolation from specific points to a gridded field, named after it contributors. The method constructs the field in multiple passes over the data, each pass correcting the working solution at each grid-point based on the residuals at the measurement locations and the distance to them. This iterative approach also constitutes the main difference to IDW. BCDG is the second step used in Gridded MOS [71], following a LR step (MOS) at each measurement location. Gridded MOS is developed and used for short term forecasting applications, but its method can be similarly used to the problem of hindcasting investigated here.

A related family of methods, known as statistical downscaling, is used to learn a mapping between a coarse, low-resolution gridded model output and a finer, high-resolution model prediction. This is not the exact setup used in the previously mentioned methods, but is equally a spatial extrapolation technique, as it is a method of adding spatial information. Statistical downscaling approaches include both classical regression models and machine-learning methods such as convolutional neural networks (CNNs). In paper [13], a CNN is trained to downscale ERA5 input data at 0.25° spatial resolution to near-surface wind speeds at 1 km resolution, using multiple meteorological predictors. The training targets are high-resolution WRF simulations, which represent the reference data during the training period. The CNN captures nonlinear spatial dependencies and produces spatially coherent high-resolution wind speed fields. Statistical downscaling is further discussed in [6], which outlines its flexibility in incorporating diverse predictors and its ability to learn site-specific mappings. Furthermore, it mentions the reduced computational cost compared to running a full NWP model as another reason for its utilization. While its input-output structure differs from traditional extrapolation methods, often mapping from and to gridded predictors rather than from mast values to the full grid, it occupies a similar role in post-processing and site-specific refinement of wind fields.

Vertical Extrapolation

A subdomain of spatial extrapolation is extrapolating in height. The height of interest is the turbine hub height, which can be used to calculate the expected power yield, but measurements might have been taken at a different height. When these heights don't align the wind profile can be used to solve this. The wind profile, or wind shear, gives the rate with which the wind speed changes with height. The reason wind speed increases with height is due to the frictional drag caused by earth's surface, which slows the air down more when the terrain is very rough and less when the terrain is more smooth. The two most common descriptions of the wind shear are the Power Law (or Wind Profile Law) and the Logarithmic Law. The Power Law, given by equation 2.1, is an empirical, straightforward formulation that gives a good approximation.

$$V_2 = V_1 \cdot \left(\frac{H_2}{H_1}\right)^{\alpha} \tag{2.1}$$

V_1 and V_2 give wind speeds at heights H_1 and H_2 respectively. α is the *power law exponent* and a good value has to be found for each new site. The Logarithmic Law, given by equation 2.2, is derived more directly from fluid dynamics and described the wind profile in the atmospheric boundary layer.

$$V_2 = V_1 \cdot \frac{\ln(H_2/z_0)}{\ln(H_1/z_0)} \tag{2.2}$$

V_1, V_2, H_1 and H_2 have the same definition as in equation 2.1. ln is the natural logarithm, and z_0 indicates the height above ground at which the wind speed should theoretically be zero. For an open grassland this would be around 0.03m and for urban areas this could for instance be between 0.5m and 1.0m.

For vertical extrapolation ML has also been applied successfully. [72] used an ANN to extrapolate from the reference height to the target height. Domain knowledge is used to arrive at three smarter input features, which are dimensionless velocity ratio over 20 meter height difference, turbulence intensity, and velocity at extrapolation level from the previous timestep, leading to a 65% decrease in MAE compared to the Log Law and a 50% decrease compared to the Power Law. [73] used LSTM successfully, focusing specifically on complex terrain locations.

2.2.3. Joint Spatio-Temporal Extrapolation

It is also possible to extrapolate the bias corrections over the spatial and temporal dimensions jointly. Very few studies take this approach, to try to estimate the wind field in space and time with one single model. Most research and industry workflows treat the two dimensions separately: first a temporal extension of a single measurement location, then a spatial extrapolation over the entire field. The handful of papers that do attempt some form of joint spatio-temporal extrapolation often don't directly translate to the problem this thesis is aimed at solving, which is long term hindcasting for WRA. However, from these methods inspiration might be drawn for the extrapolation of found biases.

[12] proposes a frequency-domain physics-informed neural network (FD-PINN) that maps the four-component coordinate (x, y, z, t) directly to the wind vector U. Physical laws (the logarithmic wind-profile and the Davenport coherence) are added to the loss so the network respects shear and turbulence physics even when only ten to thirty percent of the measurement points are available. In a synthetic, 400 second, large-eddy data set with 121 probes the model reconstructs the full three-dimensional unsteady field with root-mean-square errors below about five percent. This method is used, however, to generate a wind field from distinct measurements, and not for extrapolating bias of an existing model. Furthermore, the timescale (400 s) over which the temporal extrapolation is performed is of a different magnitude than that of a long-term WRA.

Other research performing spatio-temporal extrapolation is done by [74] and [28], both of which address forecasting instead of hindcasting. Both papers handle the two dimensions in the same order: they tackle the spatial part first and then model the temporal evolution. In Spatio-temporal DeepKriging [74] the authors use a feed-forward network with spatio-temporal basis functions to extrapolate or impute values across the spatial field at each observed time slice. Only after that step do they feed the resulting series into an LSTM/ConvLSTM that forecasts the future time steps. In the Predictive Spatio-Temporal Network [28] a CNN is applied to every "spatial wind-speed matrix" to extract spatial features, and those features are then passed to an LSTM that learns their temporal dependence and produces the prediction, again making the spatial operation the first stage and the temporal operation the second. This does not directly translate to the context this Thesis is looking at, because these methods are used to generate a wind field forecast from past and current measurements, and are not used for correcting an existing model across time and space.

3

Methodology

This chapter will detail the methodology used in exploring potential improvements for the temporal and the spatial extrapolation independently. Section 3.1 covers the temporal extrapolation, while Section 3.2 is concerned with the methodology of the spatial extrapolation.

3.1. Temporal Extrapolation

Temporal extrapolation entails using observations taken during a limited time interval and extending this to new and unseen time-steps. The goal is to predict what the measured wind speed would have been at those time steps more accurately than the LES model. The input to give to such a prediction model originates from the LES model. The terms 'prediction model' and 'correction model' will be used interchangeably throughout this section.

The methodology of the temporal extrapolation part is structured as follows. First, the datasets, baseline models, and experimental setup are defined in Sections 3.1.1, 3.1.2, and 3.1.3, respectively. Following this, Section 3.1.4 describes the methodology for comparing the baselines using mesoscale data against when using LES data, addressing the first Research Gap. Sections 3.1.5 through 3.1.10 describe alterations to be made to the baseline method, which need to be tested to see to what degree they improve or degrade the performance. To learn their effect each alteration is first studied in isolation (Sections 3.1.5 - 3.1.9), using Site 1 data, after which the promising strategies are combined (Section 3.1.10) and evaluated on Site 1, Site 3 and Site 4.

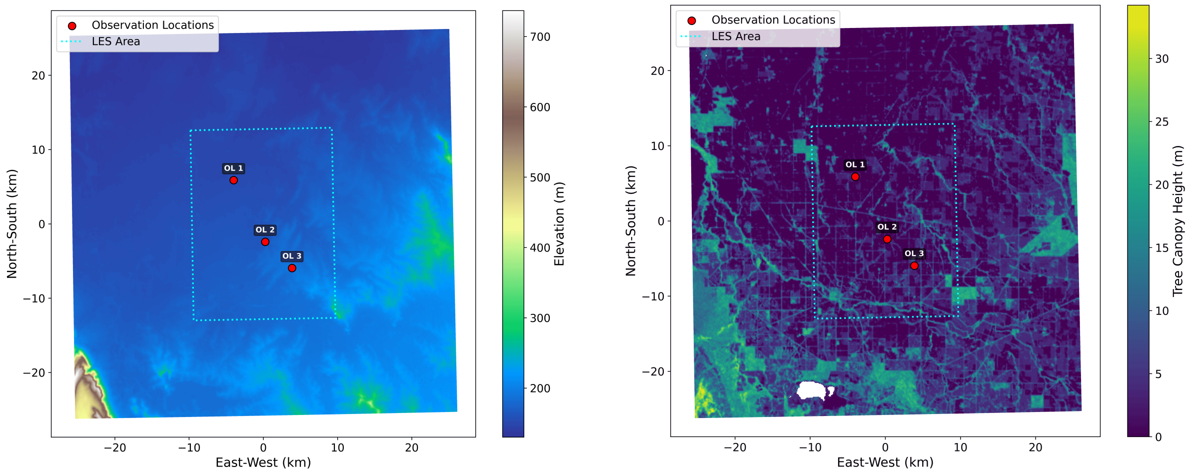
3.1.1. Data

This section details the data used for the temporal extrapolation part of the research. It covers the sites, the LES data, the observational data and the mesoscale data.

Sites

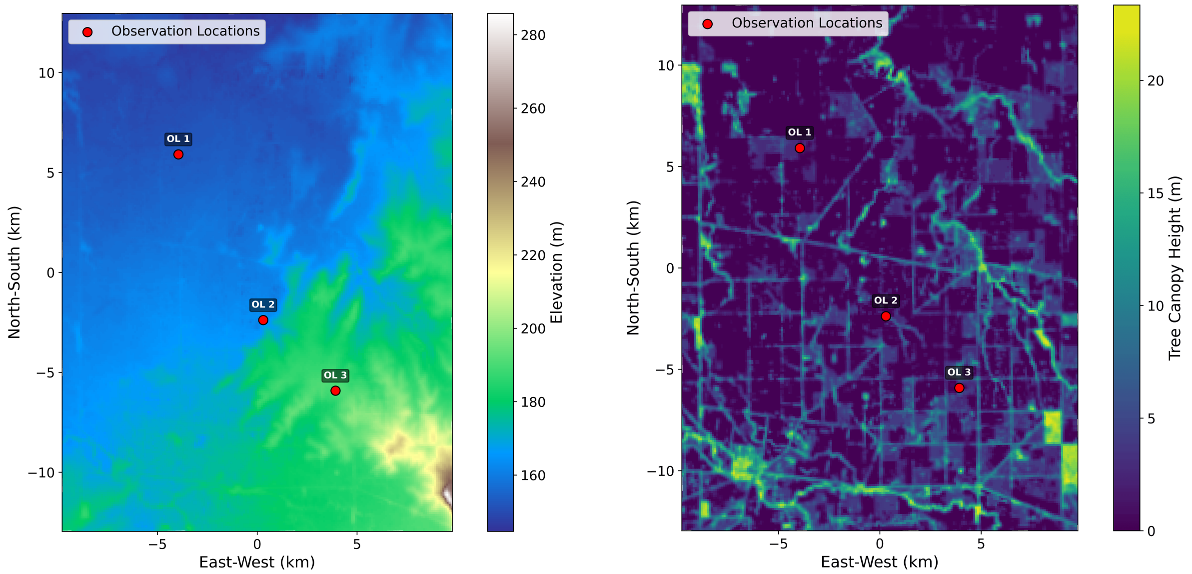
A site is the term used for a potential wind farm locations. Site 1 has a terrain of limited complexity and is used as the primary site on which the methods are developed. The ground height within the site does not vary more than 100 meters. Outside of the site itself, roughly 15 km to the south-west from the bottom left corner of LES area there is a ridge reaching a height of 500 meters above the ground level of the site. Similarly, roughly 15 km to the east of the LES area there is a ridge reaching up to 200 meters above the ground level of the site. Figure 3.1 illustrates the key terrain characteristics, specifically elevation (a, c) and tree canopy height (b, d), for both the site and its wider surroundings. The exact location and the name of the site cannot be disclosed, as requested by the third-party company owning the data.

Two additional sites, which will be referred to as 'Site 3' and 'Site 4', will be used for validation of the final resulting method developed on Site 1. The data from these sites is not used in any of the experiments, such that they are completely unseen until validation. Site 3 has terrain of very low complexity, while the terrain of Site 4 is of moderate complexity. The terrain maps of both sites can be found in Figure 3.2. For these sites it again holds that the exact location, nor the names of the sites, can be disclosed.



(a) Elevation (LES and surrounding area)

(b) Tree canopy height (LES and surrounding area)



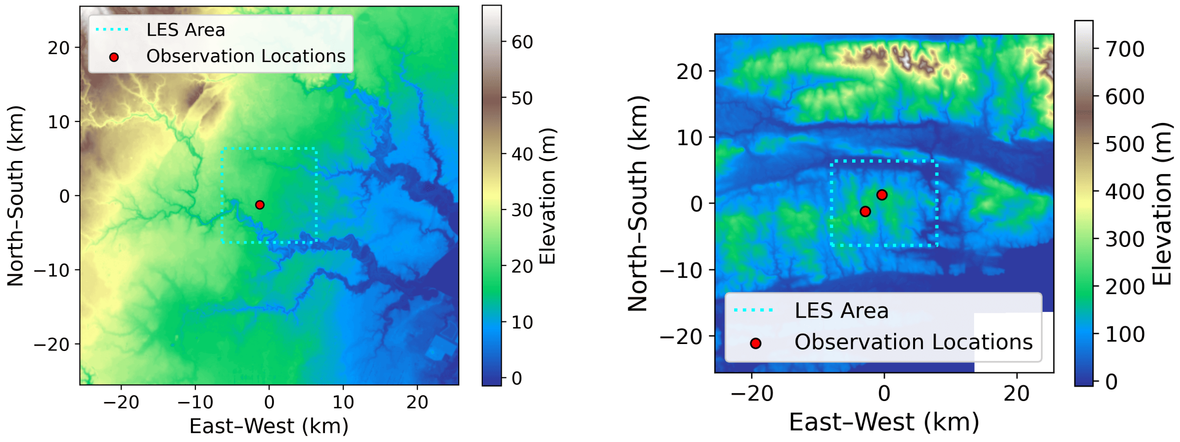
(c) Elevation (LES area)

(d) Tree canopy height (LES area)

Figure 3.1: Terrain characteristics of Site 1 and its surroundings. The top row shows the larger area, while the bottom row provides zoomed-in plots showing only the area for which the LES model is run.

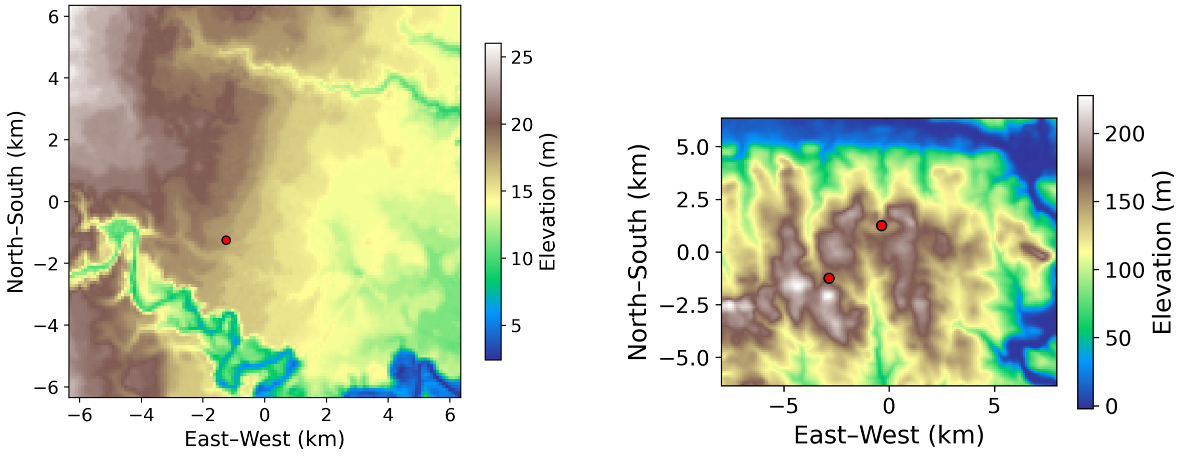
LES Data

The LES data of Site 1 covers an area of 19.2 km wide and 25.6 km tall and has a resolution of 100m, meaning that one cell covers an area of 100 by 100 meters. The LES has been ran per 25 hour block of which the first hour is spin-up and thus disregarded, leaving 24 hours of usable data. The wind fields produced by the LES have a temporal resolution of 1 hour. A wind field contains the wind conditions (wind-speed, wind-direction and resolved-turbulence) as simulated by the LES at all LES grid-cells, meaning at all site locations, for one timestep, in this case one hour. Similarly the terms wind-speed field and wind-direction field refer to the same concept, but limited to only the wind-speed and wind-direction at each location respectively. The datetime associated with each wind field is backwards looking, meaning it describes the aggregate wind conditions in the hour prior to the exact datetime. The LES data covers the period from 14:00 4/10/2022 until 14:00 4/10/2023, totaling 8760 hourly wind fields. The LES produces predictions at multiple heights above ground level, but only the height of 100m was used due to its relevance given common hub heights and it lining up with the observation data height of 101m. The attributes included in the LES data consist of the wind speed, wind direction and the resolved variance in the wind speed. Furthermore the time-of-day and the day-of-year can be added to each datapoint. Figure 3.3 plots five wind fields given by the LES at randomly sampled times



(a) Site 3: Elevation (Large area)

(b) Site 4: Elevation (Large area)



(c) Site 3: Elevation (LES area)

(d) Site 4: Elevation (LES area)

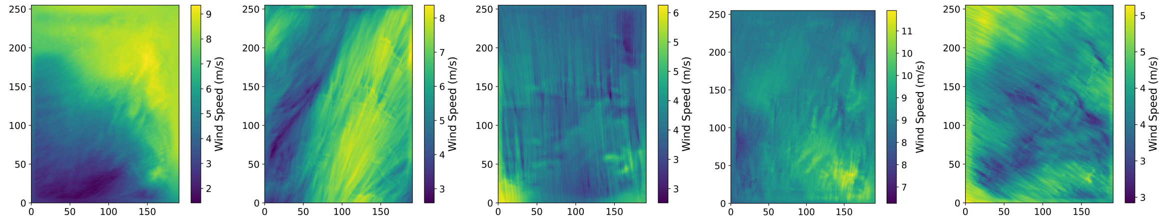
Figure 3.2: Terrain elevation maps for Site 3 and Site 4. The top row displays the larger surrounding region, where the cyan box indicates the domain used for the Large Eddy Simulation (LES). The bottom row shows the specific topography within that LES domain.

to illustrate the data. Each pixel corresponds to one LES cell and thus covers a 100 by 100 meter area, with color indicating the wind speed at that location.

The LES data for Site 3 covers an area of 12.8 km by 12.8 km. The LES data for Site 4 covers an area of 16.0 km wide and 12.8 km tall. Both LES datasets have the same characteristics as the LES dataset for Site 1 regarding both temporal and spatial resolution, features, spin-up period and the height used. Similar to the dataset for Site 1, there is exactly one year of LES data available for both Sites 3 and 4, with the first recorded hours being 07:00 on 01-04-2018 for Site 3, and 02:00 on 19-08-2018 for Site 4.

Observational Data

Measurements are taken at three locations on Site 1 using LiDAR. The three locations will be called Observation Location (OL) 1, 2 and 3, or OL 1, OL 2 and OL 3 for short. The measurements used are taken at a height of 101m. The 1m difference with the simulation height of the LES is assumed to not have any consequences. The temporal resolution of the observations is 10 minutes. The observations are averaged per hour to match the temporal resolution of the LES. The dataset contains many missing values, as is common when using LiDAR, and the decision was made to only include an hour in the dataset if at least four out of six 10-minute measurements are present. After the mentioned data processing their datasets contain 6618, 6237 and 3857 data points respectively. Figure 3.4 shows the number of hourly data points available per OL for each day. Note that for OL 3 no observations are



(a) 26-12-2022 16:00 (b) 01-03-2023 01:00 (c) 25-03-2023 23:00 (d) 05-07-2023 22:00 (e) 11-07-2023 00:00

Figure 3.3: Five example LES wind-speed fields at Site 1 sampled at random times. Each pixel is one LES grid-cell (100 m × 100 m). Color indicates wind speed ($\mathrm{m\,s^{-1}}$). Note that each subfigure uses its own colorbar scale.

available for a large part of the year. Wind speed, standard deviation in wind speed, wind direction and turbulence intensity are measured, of which only wind speed is used as the prediction target.

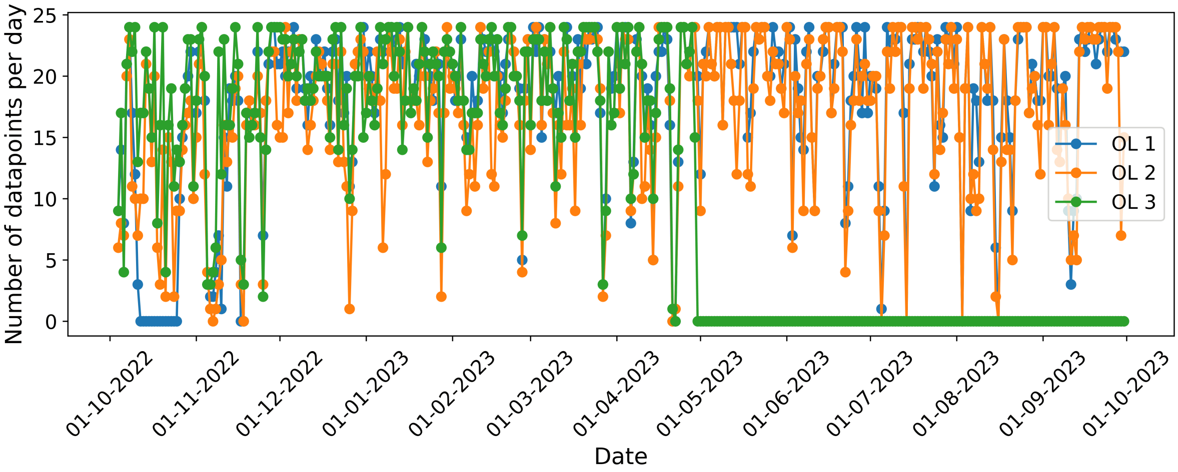


Figure 3.4: The number of available observational datapoints per day for the three Observation Locations (OL) at Site 1.

Site 3 contains one LiDAR observation location, which will be called OL 4, while Site 4 contains 2 LiDAR observation locations, which will be called OL 5 and OL 6. These locations are indicated in Figure 3.2. Similar to the observational data for Site 1 the observations used are taken at a height of 101m, have a temporal resolution of 10min which will be aggregated to hourly, and the data contains many missing values. The datasets for OL 4, 5 and 6 contain 5531, 6493 and 5092 data points respectively.

Mesoscale Data

A mesoscale dataset is included as an alternative long-term reference source for the MCP baseline comparison for all observation locations. The datasets are provided by Whiffle, the same company that supplied the LES data, and consist of time series of wind speed, wind direction, standard deviation in wind speed and turbulence intensity. The data are available at fourteen heights between 40 m and 181 m, including 101 m. The temporal resolution is 10 minutes, and the time series cover the same one-year period as the LES data in all cases. Similar to the LES data, the mesoscale values are aggregated to hourly means.

The dataset does not contain a spatial grid or any fields describing the underlying mesoscale model, and only the point time series at the observation locations are used. Since MCP relies solely on long-term temporal correlations at the observation locations, this level of detail is sufficient.

3.1.2. Baselines

A set of five distinct regression models is employed to establish a comprehensive baseline for performance evaluation. The first of these, representing the current industry standard for prediction, is Linear Regression (LR) based MCP. The long-term reference data is given by the wind speed modeled by the

Table 3.1: Hyperparameter search distributions per model. The Python package SciPy is used for the distribution functions *loguniform()*, *uniform()* and *randint()*.

Model	**Hyperparameter**	**Distribution / Values**
Ridge	α	loguniform$(10^{-3}, 10^{3})$
MLP	hidden_layer_sizes	(128,64,32,16), (128,64,32), (64,32,16), (64,32)
	max_iter	1000
	learning_rate_init	loguniform$(10^{-4}, 1)$
	α	loguniform$(10^{-4}, 10^{2})$
	early_stopping	True
	n_iter_no_change	randint$(5, 50)$
	tol	loguniform$(10^{-5}, 10^{-1})$
SVR	C	loguniform$(10^{-3}, 10^{3})$
	ϵ	loguniform$(10^{-3}, 1)$
	γ	loguniform$(10^{-4}, 10)$
XGBoost	early_stopping_rounds	randint$(5, 50)$
	n_estimators	randint$(100, 2000)$
	max_depth	randint$(1, 10)$
	learning_rate	loguniform$(10^{-4}, 1)$
	min_child_weight	loguniform$(10^{-2}, 10^{1})$
	subsample	uniform$(0.5, 1.0)$
	colsample_bytree	uniform$(0.5, 1.0)$
	γ	loguniform$(10^{-4}, 10^{1})$
	reg_alpha	loguniform$(10^{-2}, 10^{3})$
	reg_lambda	loguniform$(10^{-2}, 10^{2})$

LES, at the grid-cell containing the OL. The target data consists of the wind speed measurements taken at that same OL. LR minimizes the residual sum of squares, and a LR is fitted for each OL individually.

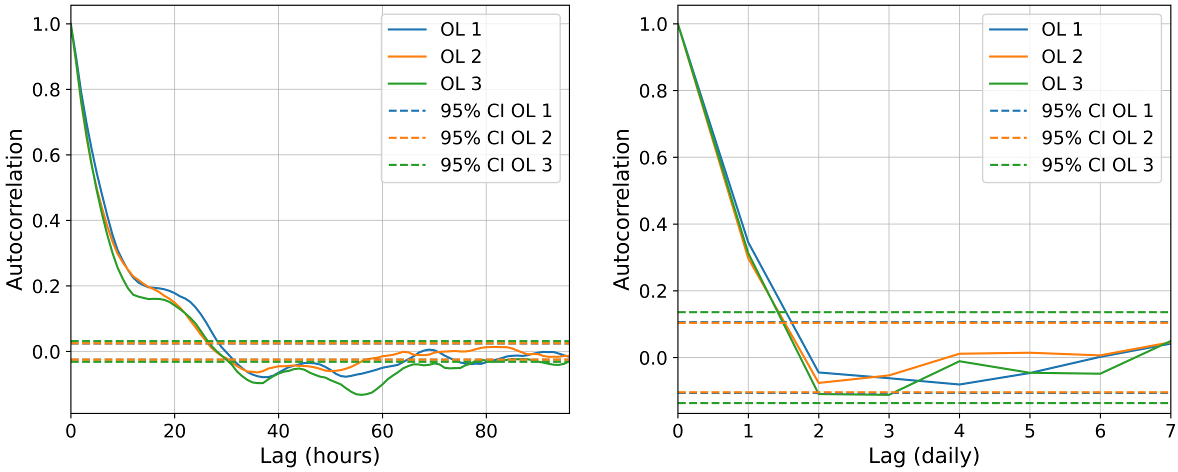
Furthermore, Ridge Regression is used as a baseline that acts in a very similar way as LR, but incorporates L2 regularization to improve over it. While this is not expected to perform different in the baseline scenario where only one value is given as input, regularization becomes more important when the input becomes of higher dimension. Since it is a straight-forward and existing method it is used as a baseline in this research.

Lastly, although ML models are still uncommon in WRA they have been used and shown to give superior performance in multiple studies. Therefore, ML models will be used as another baseline on which to improve. The additional ML models used are Multi-Layer Perceptron (MLP), Support Vector Regressor (SVR) and XGBoost (XGB). The baseline version of these models takes as input only the wind speed predicted by the LES, similar to LR and Ridge. As the input on which to make a prediction grows larger, the ML models are expected to outperform LR more than in the baseline scenario. Furthermore, each ML model has hyperparameters whose correct tuning is vital to the models performance. Section 3.1.3 details the manner in which hyperparameter tuning is performed. The hyperparameters to tune for each model, along with the distribution from which values will be sampled is given in Table 3.1. The Python packages *scikit-learn* and *XGBoost* are used, keeping their default values for all hyperparameters not specified in Table 3.1.

3.1.3. Experimental Setup

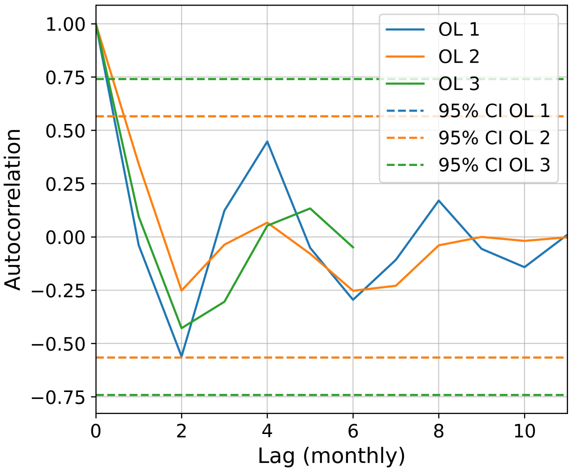
This section describes the setup used when running any of the experiments that are yet to be described. This ensures a constant and unified approach is taken, enabling a fair comparison of the results.

- All input features are standardized by removing the mean and scaling to unit variance.
- Nested cross-validation is used with 6 outer folds and 5 inner folds.
- The dataset is not shuffled before creating the folds. This avoids information leakage from the test set to the training set.
- Hyperparameter search is performed by the inner cross-validation, using random search over parameter distributions.



(a) Hourly

(b) Daily averaged



(c) Monthly averaged

Figure 3.5: Autocorrelation of the observed wind speeds at the three OLs of Site 1, using (a) hourly data, (b) daily averages, and (c) monthly averages. The plots show no seasonal patterns or other strong autocorrelations that could cause significant information leakage when using unshuffled 6-fold cross-validation.

- Hyperparameter search optimizes for RMSE and is done for 100 iterations (hyperparameter samples).
- The outer cross-validation is used for evaluating the performance of the trained model on unseen data.

Using chronological folds instead of shuffled splits prevents information leakage from the test set into the training data which would otherwise likely occur, given that the data is a timeseries. Autocorrelation analysis was performed to verify the absence of significant leakage in chronological folds (Figure 3.5). The results show that autocorrelation falls below the 95% confidence bounds within 28 hours. The rapid decay of autocorrelation ensures that sufficiently large chronological folds are statistically independent of their neighbors, regardless of their position in the time series. Therefore, predicting a middle fold (temporal interpolation) presents the same challenge as predicting the first or last fold (temporal extrapolation), validating the use of unshuffled cross-validation. Moreover, no seasonal patterns were detected that could bias results if a test fold coincided with a particular season.

The model performance is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide a comprehensive assessment of error magnitude and variance explanation, and are commonly seen in related research.

The **MAE** calculates the average absolute difference between predicted (p_i) and observed (o_i) values. It treats all errors equally, providing a robust performance measure, as shown in Eq. (3.1):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |p_i - o_i| \tag{3.1}$$

The **RMSE** measures the square root of the average squared errors. Unlike MAE, the RMSE penalizes large deviations heavily, making it sensitive to outliers (Eq. 3.2):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (p_i - o_i)^2} \tag{3.2}$$

Finally, the $\mathbf{R^2}$ expresses the proportion of variance in the observed data explained by the model (Eq. 3.3). A value approaching 1 indicates the model successfully reproduces observed variability.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (p_i - o_i)^2}{\sum_{i=1}^{n} (o_i - \bar{o})^2} \tag{3.3}$$

where n represents the total number of samples and $\bar{o}$ denotes the mean of the observed values.

3.1.4. Mesoscale vs. LES as MCP Reference Data

The first research gap this thesis aims to address is the evaluation of the existing go-to method, MCP, when using a microscale model, or more specifically an LES model, as the reference data source.

The experiment employs all five baseline correction models (LR, Ridge, MLP, SVR, and XGB) across all six available Observation Locations (OL 1 through OL 6) from Site 1, Site 3, and Site 4. The models are run using wind speed as the sole input feature for both the Mesoscale and LES reference sources. The prediction target remains the measured wind speed at each location.

Training and hyperparameter tuning adhere to the established nested cross-validation procedure outlined in Section 3.1.3. Performance evaluation is based on MAE, RMSE and R^2.

The size of the improvements obtained by the baseline MCP corrections will serve as a reference point for assessing the significance of all subsequent methods. The magnitude of the gains achievable by standard MCP establishes a practical scale against which later enhancements can be meaningfully compared.

3.1.5. Additional Input Features

In addition to wind speed the LES provides other attributes that can be used as input to the prediction model. The hypothesis is that additional attributes will allow the prediction model to better find patterns in the data that differentiate between situations in which the predicted wind speed is an overestimation or underestimation and by how much. The attributes available are:

- resolved variance in wind speed (m^2/s^2)
- wind direction (deg.)
- time-of-day
- day-of-year

Some attributes are circular, meaning that their smallest and largest value are in reality right next to each other. This is the case for wind direction, time-of-day and day-of-year. To maintain their circularity they can be represented by two values which are given by Equations 3.4 and 3.5.

$$f(x) = sin(2\pi\frac{x}{x_{max}}) \tag{3.4}$$

$$g(x) = cos(2\pi\frac{x}{x_{max}}) \tag{3.5}$$

x_{max} is the value at which one rotation of the loop the variable represents is completed, which is 360, 24 and 365 for wind direction, time-of-day and day-of-year respectively.

The *resolved variance in wind speed* is the measure of turbulence that the LES model directly simulates by solving eddies that are bigger than its grid size. This is separate from the smaller-scale eddies, of which the effect is instead estimated using a simplified mathematical model.

A prediction model is trained for every possible set of additional input variables. The baseline to compare with is the empty set, leaving the wind speed as the only input variable.

3.1.6. Additional Area

In addition to the LES cell containing the observation location, surrounding cells might contain useful information. The baseline is to only use data from the LES cell containing the observation location. When using a weather model with a coarser resolution a single cell covers an area of multiple km^2. Using data from surrounding cells in that case is less likely to contain useful information, due to the distance being too large. This could explain why this is not seen in any literature. Since the LES model has a resolution of 100m, it is possible that surrounding cells contain important information for correctly predicting the wind speed at the original cell. Furthermore, due to the high resolution of the LES there is a higher chance that a single cell captures only a local phenomena, missing information about the more global weather situation taking place.

To test this hypothesis data extracted from a range of different area sizes are used as input for the prediction model. The baseline is an area of 1 cell by 1 cell (100m by 100m). This is increased by 1 surrounding layer of cells at a time, keeping the original cell at the center of the area. This results in an area of 3 by 3 cells, 5 by 5 cells, etc., until a region of 11 by 11 cells (1100m by 1100m) is reached. Beyond this size areas of 15 by 15, 21 by 21 and 31 by 31 cells are tested. The largest area tested thus corresponds to a border of 1.5km on each side of the original cell. Figure 3.6 visualizes these input configurations. The baseline single-cell input is shown in Figure 3.6a. The progressively larger spatial contexts are illustrated in Figure 3.6b (11×11), Figure 3.6c (21×21), and Figure 3.6d (31×31). Finally, Figure 3.6e displays the full LES domain to provide a sense of scale for these extractions relative to the total simulated area.

To investigate the limit of useful spatial information beyond the 31×31 threshold, an extended analysis is conducted using area sizes increasing in increments of 10 cells, up to a maximum of 111×111. This largest area effectively reaches the boundaries of the LES domain. Due to the quadratic growth of the input dimensionality, which renders the training of complex non-linear models computationally prohibitive, this specific experiment is restricted to the Ridge Regression baseline.

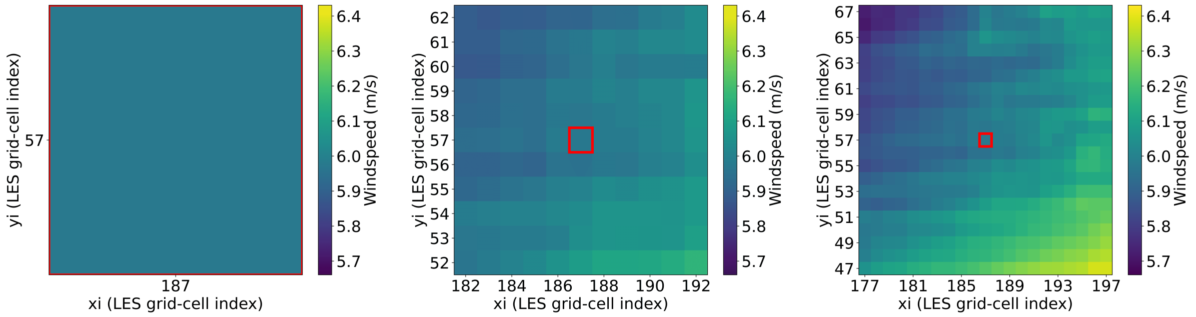
3.1.7. Additional Time Steps

As successive wind speeds are highly correlated, the wind speeds modeled by the LES at the time steps before and after the target time step might prove useful to include in the input to the prediction model. Intuitively the correlation between wind speeds at successive time steps makes sense: if it is very windy right now, it is likely to also be windy in 1 hour. The correlation of wind speeds in time can also be rigidly analyzed and confirmed using autocorrelation analysis. Figure 3.5 shows the autocorrelation found between the wind speed at arbitrary time t and the wind speed at any amount of time into the future. This reveals there is indeed a significant correlation for up to 28 hours into the future.

The baseline to compare to is using data only from the time step of the LES model for which we are trying to predict the observed wind speed. This can be compared to using data from all time steps within a certain timewindow. The timewindow will be of odd length, centered on the target time step and spanning an equal interval into the past and future. The smallest timewindow is of size 3, including timesteps t_{-1}, t_0 and t_1, where t_0 is the time step for which wind speed needs to be predicted and t_{-1} and t_1 are the time steps immediately prior and following t_0. The timewindow size will be increased repeatedly by including 1 more sequential time step in the past and in the future and for each size the performance will be recorded. This process will be stopped at the point where it becomes clear from the data that further increases to the timewindow size will not yield further improvement.

Due to the nature of WRA, which always uses data from the past, the model simulations of both previous and future time steps are readily available. This is in contrast with forecasting, where future information is never available. Since this research only focuses on WRA this is not a limitation.

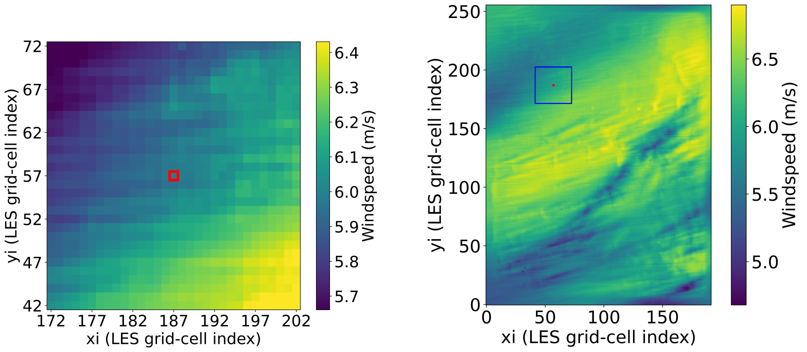
The use of a timewindow comes at the cost of not being able to correct a number of time steps at the beginning and end of the modeled period, with the number depending on the size of the timewindow.



(a) 1 × 1 (100m × 100m)

(b) 11 × 11 (1.1km × 1.1km)

(c) 21 × 21 (2.1km × 2.1km)



(d) 31 × 31 (3.1km × 3.1km)

(e) Full LES domain

Figure 3.6: LES wind speed fields at 14:00 07-10-2022 centered on Observation Location (OL) 1. Figure 3.6 visualizes these input configurations. The baseline single-cell input is shown in (a). The progressively larger spatial contexts are illustrated in (b), (c), and (d). Finally, (e) displays the full LES domain to provide a sense of scale. The cell containing OL 1 is indicated with a red rectangle. In the full LES plot, the blue rectangle marks the largest 31×31 area. Note that the colorbar scale is consistent among the first 4 plots but different for the plot of the full LES.

This could be accommodated for by either running the LES for some additional hours at the start and end of each run, or using a second model that does not use a timewindow for these time steps. Since rerunning the LES for this research is not an option and the number of lost datapoints is a very small fraction of the total data, the data points at the beginning and end will simply be left out during the experiment.

3.1.8. Joint Multi-Location Predicting

Instead of training a separate prediction model for each individual Observation Location (OL), a single prediction model can be trained that is able to make predictions for all OLs. For each OL local phenomena and patterns are at play which differ from location to location. However it can also be hypothesized that patterns exist that are the same across multiple locations. This seems especially likely when the locations are all from within the same site and thus not completely random nor dissimilar and when the terrain of a site is simple.

Since learning is limited by the amount of training data available, training one model on data from multiple OLs of the same site could improve performance. Each data point is enriched with a one-hot encoding, informing the model about the site from which the data originates and for which a prediction is required. Ideally the model then learns global patterns from the entire dataset, while learning local patterns separately for each OL. The local patterns are not expected to be learned better than when using individual models, but global patterns now can be learned more accurately because more data is available.

The new composite dataset will contain 16,712 data points, as given by the sum of the sizes of the individual observation datasets. The one-hot encoding will add a number of dimensions equal to the number of used OLs, which is three for Site 1.

The performance will be assessed for each OL individually. In that manner it can also be compared to the baseline performance of individual models.

3.1.9. Autoencoded Wind-speed Fields

A convolutional autoencoder (CAE) will be used to encode the LES wind-speed fields, which will be provided as additional input features to the prediction model. The CAE consists of an encoder and a decoder. The encoder takes as input the wind-speed field, meaning the wind speeds predicted by the LES across the site, and outputs a 1 dimensional vector. This vector is the *encoding* of the wind-speed field in the *latent space*, and the length of this vector defines the dimensionality of the latent space. The decoder takes the encoded wind-speed field as input and outputs a 2d vector of the same size as the original wind-speed field. The goal of the CAE is to reconstruct the input wind-speed field at the end of the decoder.

Upon initialization the CAE will produce random noise, but since it is in essence simply a specific type of neural network it can be trained using backpropagation in the same manner. During this training the target is defined as being equal to the input. CAEs contain a information bottleneck, which is the dimensionality of the latent space. This is the point in the network where the dimensionality is smallest. It is easy to see that any information present at the end of decoding must be have passed through the bottleneck, and is thus contained in the encoding. The encoding of a wind-speed field is added to the input of the prediction model, providing the prediction model with essentially the entire wind-speed field but represented in a drastically smaller number of dimensions.

Enriching the input to the prediction model with the encoding generated by the CAE adds site-wide information the model can use to improve predictions. With input variables only taken from the closest LES cell, which is the baseline, or even from surrounding cells, as proposed in Section 3.1.6, all information is extremely local. This site-wide information can inform the prediction model about the larger context, enabling it to better distinguish between situations and find systematic errors in the LES model.

For training the CAE all 8760 wind-speed fields available for Site 1 are used. No distinction is made between train and test data, nor is the availability of observation data at each wind-speed fields timestep considered. Including LES wind-speed fields which will subsequently be contained in the test set is not a problem in this case, as the CAE can be seen as a (complex) feature transform, and is not part of the prediction model. The prediction model itself does not have access to the encoded wind-speed fields of the test set during training. During WRA, the usecase this research is concerned with, all LES wind-speed fields are similarly available from the start due to the fact that WRA looks only at historic data, as is visualized in Figure 2.1.

Multiple CAEs with differing latent space dimensions will be trained. Finding a good latent space dimension might prove crucial to the effectiveness of this approach. It is important that the latent space is large enough to be able to accurately represent the wind-speed field in it, whilst being small enough to not overload the prediction model with too many additional features. Models will be trained with latent space dimensionalities 8, 16, 32, 64 and 128.

The CAE is implemented using the Python package *PyTorch*, using *nn.Conv2d*, *nn.Linear*, *nn.ConvTranspose2d*, *nn.ReLU* and *nn.Flatten* to build the model. The structure of the encoder consists of five convolutional layers and two linear layers, with a ReLU activation between each of the layers. The size of the layers are chosen to work with the site-specific size, which for Site 1 is 192 by 256. The decoder consists of the same layers and layer sizes as the encoder, but in reversed order and with ConvTranspose2d replacing the normal Conv2d layers used in the encoder. A schematic representation of the CAE model and its layers for Site 1 can be found in Figure 3.7. Each CAE is trained for 200 epochs, with each epoch feeding the network all wind-speed fields one time. This duration was determined empirically, as preliminary experiments demonstrated that the reconstruction loss consistently stabilized well before this point, ensuring that models of all latent dimensions fully converged. Wind speeds are normalized before being passed on to the encoder by subtracting the dataset minimum and dividing by the range, so all values lie between 0 and 1. The reverse mapping is applied to the outputs of the decoder. The CAE is optimized using the Adam optimizer with a learning rate of 0.001. All Convolutional layers use a stride of 2 and a padding of 1.

For Site 3 and Site 4, the input dimensions differ from Site 1, resulting in different intermediate tensor sizes after each convolutional layer and the *Flatten* operation. However, the channel sizes, kernel sizes, strides, and padding are kept identical, as is the linear layer configuration, ensuring that the overall CAE functioning remains consistent across all sites.

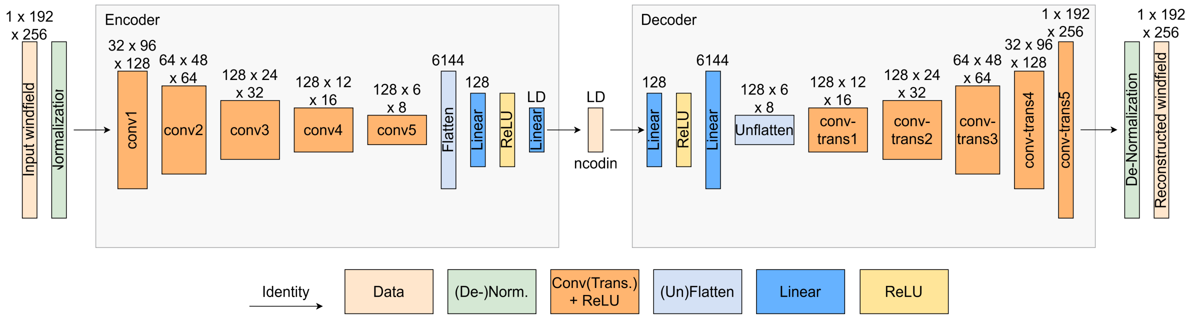


Figure 3.7: Architecture of the convolutional autoencoder (CAE). The encoder compresses the LES wind field input into a latent representation, and the decoder reconstructs it back to the original resolution. The dimensionality of the latent space is variable and indicated by LD. Note that the tensor sizes displayed correspond to the Site 1 wind field dimensions.

3.1.10. Combining Strategies

If multiple strategies are found to be effective on their own, they are subsequently combined to determine whether a composite approach further improves performance. The results of the individual experiments are analyzed to find the hyperparameter that yields the most improvement. These hyperparameters are i) the set of additional input features, ii) the area size, iii) the time window length, iv) the dimensionality of the latent space encoding, and v) whether to use individual models for each OL or a single model trained for all OLs of a single site. The results from the individual experiments inform the design of the combined strategy.

When combining the multiple strategies performance could worsen due to the quickly growing dimensionality of the input to the prediction model. The dimensionality of the input is given by the equation

$$d_{\text{input}} = t \cdot \left(f \cdot a^2 + l\right) + u \tag{3.6}$$

where f is the number of features used from every LES cell, a denotes the size of the area used, t is the length of the time window, l is the latent space dimensionality and u is 0 if the model is trained for a single location or the number of locations the model is trained to predict jointly otherwise. Note that the wind direction, time of day and day of year are made circular using Equations 3.5 and 3.4 and thus produce two features. Machine learning models are subject to the curse of dimensionality, meaning that overly high dimensionality often harms performance due to sparsity of data in the feature space. The optimal values found during the individual experiments could therefore be different than the optimal values for the combined setup. If for any of the individual parts multiple hyperparameters perform similarly, the hyperparameter resulting in the lowest input dimensionality is expected be best in the combined strategy, as this combats the curse of dimensionality. This assumption will inform the design of the combined strategy.

Note that combining a time window with any other strategy implies that the second strategy is applied to every timestep within that window. For example, when using both a time window and a CAE-generated encoding of the wind-speed field, the input to the prediction model will contain the encodings for all wind-speed fields across the entire time window.

Since the combined strategy is designed using the results of the individual experiments on Site 1, its performance must be validated on sites that were not used in its design. For this purpose, Site 3 (OL 4) and Site 4 (OL 5 and OL 6) are used as independent test locations.

3.2. Spatial Extrapolation

Spatial extrapolation in this context refers to the transfer of information from a limited set of observation locations to the entire site. At the observation locations a ratio is found with which the LES model needs to be multiplied to perfectly match the observed wind speed. This ratio will be referred to as the correction factor. The goal is to find a method of extrapolation that most accurately predicts the correction factors for the entire site, using the correction factors found at the observation locations.

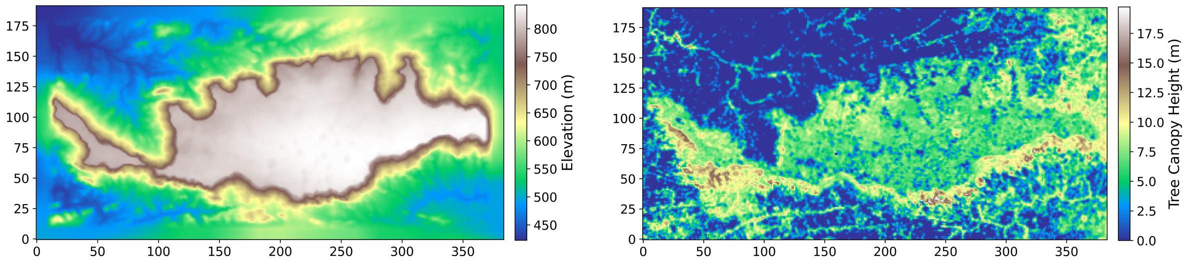
The methodology of the spatial extrapolation part is structured as follows. First, the datasets, baseline method, and experimental setup are defined in Sections 3.2.1, 3.2.2, and 3.2.3, respectively. Following this, Section 3.2.4 details the methodology for incorporating the LES wind field structure into the extrapolation process, introducing Wind-Speed Enhanced IDW (WS-IDW) and Wind-Direction Enhanced IDW (WD-IDW) to address the third Research Gap. Finally, Section 3.2.5 describes the approach for integrating terrain information, specifically elevation and tree canopy height, to further refine the spatial weighting.

3.2.1. Data

This section details the data used for the spatial extrapolation part of the research. It covers the site, the LES data and the observational data.

Site

Two sites will be used for the spatial extrapolation experiments: Site 1 and Site 2. Site 1 is described in Section 3.1.1. Site 2 has terrain of moderate complexity, with the area of interest elevated several hundred meters higher than its surrounding. The elevated area itself is however rather flat. The elevation map can be found in Figure 3.8a. The Tree Canopy Height (TCH) map can be found in Figure 3.8b. Site 2 is selected for the large number of met masts that have been placed at the site to collect wind data. The exact location, nor the name of the site cannot be disclosed, as requested by the third-party company owning the data.



(a) Elevation (m)

(b) Tree Canopy Height (m)

Figure 3.8: Terrain feature maps of the LES simulation domain at Site 2. (a) shows the elevation, which is defined as the ground height above sea level, and (b) shows the tree canopy height. Both terrain features are given in meters.

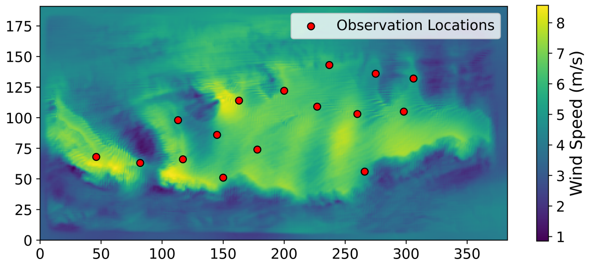


Figure 3.9: Wind-speed field at Site 2 at 09:00 on 08/03/2014 including the observation locations indicated in red. The axes give the LES grid-cell indices, with each grid-cell covering a 100m × 100m area.

LES Data

The LES model covers an area of 38.4 by 19.2 km. The spatial resolution of the LES is 100 by 100 m and the temporal resolution is hourly. Similar to the LES data detailed in Section 3.1.1 the datetimes are backward looking, the height above ground level used is 100m, and the attributes available include wind speed, wind direction and the resolved variance in the wind speed. The LES is run for 10 representative days. These runs start at 02:00, at which point the next hour is the spin-up period the LES needs for accuracy, after which the period from 03:00 until 03:00 the next day is modeled. The 10 days of which the LES data consists are 08/03/2014, 01/06/2014, 08/07/2014, 18/07/2014, 04/10/2014, 12/11/2014, 18/11/2014, 22/11/2014, 12/01/2015 and 29/01/2015. The LES dataset thus consists of 240 wind fields. An example of a LES wind-speed field, together with the observation locations, can be found in Figure 3.9.

Observational Data

There are 16 met. masts that have simultaneously collected data at Site 2. The placement of these masts across the site can be found in Figure 3.9. The temporal resolution of the measurement data is 10 min, which is averaged to hourly in order to match the temporal resolution of the LES data. In contrast to the observational data for the Temporal Extrapolation there are no missing values in the data. Observations are available for all 16 masts from 01/02/2014 at 03:00 until 19/02/2014 at 00:00. The limiting factor is thus the availability of LES wind fields, limiting the dataset to 240 datapoints. To match the height above ground level of the LES data, the observations taken at 101m are used. Of the multiple variables recorded by the met. masts only the wind speed will be used as target for the spatial extrapolation.

3.2.2. Baseline

The industry standard used for the spatial extrapolation procedure described is Inverse Distance Weighting (IDW). IDW calculates the correction factor for every location LES grid-cell individually as the weighted sum of the correction factors found at the observation locations. The weights given to each correction factor is inversely proportional to the square of the distance between the target location and the observation location. Equation 3.7 gives the exact formula used to calculate this.

$$\hat{Z}(\mathbf{x}_0) = \frac{\displaystyle\sum_{i=1}^{N} \frac{Z(\mathbf{x}_i)}{d(\mathbf{x}_0, \mathbf{x}_i)^p}}{\displaystyle\sum_{i=1}^{N} \frac{1}{d(\mathbf{x}_0, \mathbf{x}_i)^p}}, \tag{3.7}$$

In Equation 3.7, $\hat{Z}(\mathbf{x}_0)$ is the interpolated value at the target location $\mathbf{x}_0$, $Z(\mathbf{x}_i)$ is the value found at location $\mathbf{x}_i$, $d(\mathbf{x}_0, \mathbf{x}_i)$ is the distance between the target and observation locations, $p > 0$ is the power parameter controlling distance decay, and N is the number of available observations.

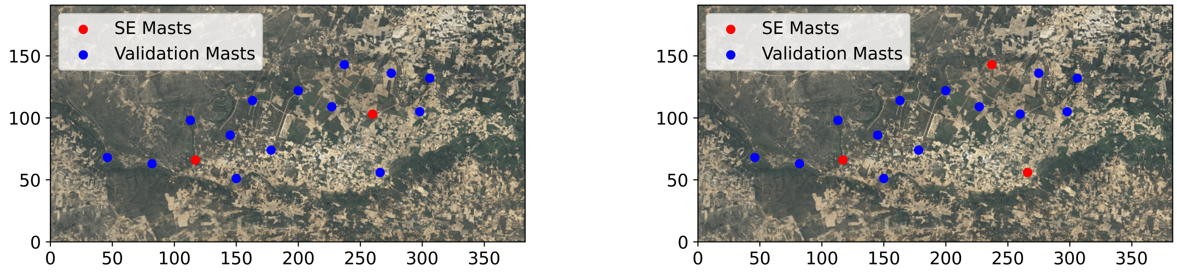
The intuition behind IDW is straightforward and easy to grasp. If a wind speed is measured at one location that wind speed is likely to be a very good estimate for a location very closeby, and a less good estimate for a location further away. When calculating the correction factor for an unobserved location (x, y) and there are multiple observations, all revealing a distinct correction factor at their respective locations, more weight should be given to a correction factor found close to (x, y) than a correction factor found further away.

The power parameter p is the only hyperparameter that needs to be set for IDW. It controls the rate at which the influence of an observation decays with distance, allowing the interpolation to vary from a global average ($p = 0$) to a highly localized estimate (high p). Prominent industry tools, such as OpenWind, use $p = 2$ [66], and this research will follow that standard.

3.2.3. Experimental Setup

This section outlines the experimental setup used across different experiments, which ensures a fair comparison can be made. As mentioned the dataset of Site 2 consists of 240 LES wind fields and observations from 16 meteorological masts places across the site. The availability of this many masts is extraordinary, with a site usually only containing 1, 2 or 3 masts. Because the goal is find the best method that can be used on real sites in real world scenarios, this situation is simulated by only using observations from 2 or 3 masts for the spatial extrapolation. This leaves the observations from the remaining 14 or 13 masts for validation. The masts used for Spatial Extrapolation will be referred to as SE masts, while masts used for validation will be referred to as validation masts. Using 1 mast is not considered because in that scenario the spatial extrapolation is straightforward, and consists of a uniform scaling using the found correction factor.

When masts are placed for a WRA project, their locations are chosen in a way that maximizes the information gained from their measurements. It would for instance not make sense to place two masts near each other while leaving a large part of the site unobserved. Because of this, it is valid to select the SE masts in a way that is suspected to maximize information. The SE masts selected for the experiments using 2 and 3 SE masts are indicated in Figures 3.10a and 3.10b respectively. These sets are selected based on the intuition that they should be spread out, covering the site as good as possible, whilst remaining central in their respective section of the site they are covering, as to maximize overall information and not choose the furthest, most extreme locations. These selections were verified by a domain expert at Whiffle to confirm they represent realistic measurement campaigns.



(a) 2 SE Masts; 14 Validation Masts

(b) 3 SE Masts; 13 Validation Masts

Figure 3.10: The plots show the placement of the chosen Spatial Extrapolation (SE) masts (indicated in red) on the site map for experiments using (a) two, and (b) three SE masts. The validation masts are shown in blue. The sets of SE masts were chosen in a way that mimics what could have been the real set if only n masts were placed.

Although the observations from the validation masts could be used to tune the hyperparameters in this research, this would not be possible in a real world scenario where all masts are used for extrapolation and no masts are withheld for validation. For the baseline, IDW, this problem is circumvented by using the standard $p = 2$ for IDW. For hyperparameters introduced by the additional strategies presented in subsequent sections it is important to realize this distinction between the experimental setup and the real world. To get an unbiased performance evaluation for any additional strategies that uses hyperparameters, the hyperparameter tuning either shouldn't rely on the validation masts, or the subsequent performance validation should be performed on a different site. Using distinct time steps or time periods is not sufficient in this case, because the spatial characteristics of a site remain the same over time.

Note that given the hyperparameter $p = 2$ the baseline method IDW doesn't actually learn anything, it just operates deterministically according to its specification. Therefore there is no need for splitting the data into a training and test set, and instead we can directly evaluate the performance of the method on all data points.

For methods introduced in Sections 3.2.4 and 3.2.5 hyperparameters have to be learned. As these hyperparameters relate to the spatial structure of a site the hyperparameter tuning and the method validation have to be performed on separate sites to be able to give an unbiased performance score. For the scenario in which 2 SE masts are used the hyperparameter tuning can be performed using the data from Site 2, after which the validation can be done using Site 1, where two masts at a time will be used as SE masts while the third mast will be used for validation. Since there are 2955 timesteps at which observations are available for all three observation locations of Site 1, and for each timestep each of the observations can be used for validation while the others are used for extrapolation, the validation set consists of 8865 datapoints.

For the scenario in which three SE masts are used such unbiased validation on Site 1 is not possible, as a fourth mast would be necessary for validation.

The performance of the spatial extrapolation techniques detailed in this section is quantified using the MAE (Eq. 3.1) and the RMSE (Eq. 3.2).

3.2.4. Using the LES Wind Field Structure

The first concept for improving spatial extrapolation is to use the wind field structure at the specific timestep for which the correction factors are to be extrapolated. The wind field contains the wind-speed field and the wind-direction field, which are the wind speeds and wind directions as modeled by the LES at all locations at a single timestep.

Wind Speed

Visual inspection of the wind-speed fields reveals distinct regions of high and low wind speeds, and it can be hypothesized that all locations within such a region need a similar correction. In other words, a correction factor found at an observation location is most informative for other locations that share a similar wind speed.

Figure 3.11 shows a wind-speed field from Site 1 where such regions can clearly be seen (indicated by a dashed ellipse), and where it is plausible that relying only on distance weighting in the spatial extrapolation could be suboptimal. The top-left observation location (red dot) is in a high wind speed region and would, according to the hypothesis, be most informative for all other locations in the region. However, there are many locations for which this does not hold when using only geographical distance, one of which is indicated with an orange dot. The orange dot is geographically closest to the middle observation location, rather than the top left one, and thus the middle observation location will be given more weight than the top left one in traditional IDW. Here an example from Site 1 is used rather than from Site 2, for the sole reason that the wind-speed fields for Site 1 contain a simpler structure and thus illustrate the intuition better. There is no reason to assume the same wouldn't apply to Site 2.

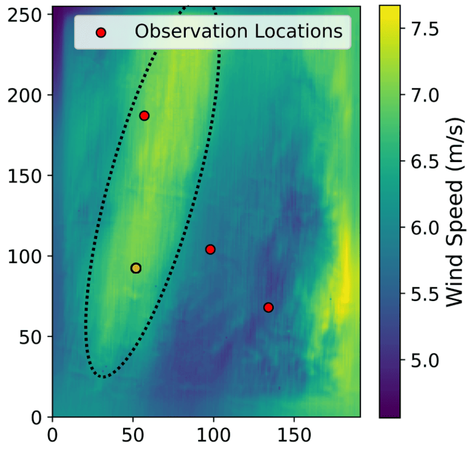


Figure 3.11: Example wind-speed field illustrating the importance of wind-speed similarity in spatial extrapolation. A high wind speed region is indicated by the dotted black line and it encapsulates the top-left observation location, which ideally is then given increased weight in correcting the entire region. The orange dot indicates a location within this region that is geographically closest to the middle observation location, and thus will give the largest weight to the correction factor found there if the traditional IDW is used for SE.

More abstractly, locations where the LES models similar wind speeds should be corrected in similar ways. Consequently, when determining the correction factor for an unobserved location (x, y), correction factors from observation locations with more similar LES wind speeds should be weighted more heavily than those with less similar wind speeds.

While wind speed similarity is hypothesized to be a useful feature, it is unlikely that it should entirely replace the dependence on geographical proximity. The geographical distance between points must still be incorporated into the spatial extrapolation process. Therefore, a balance must be found between geographical distance and wind speed difference.

The wind-speed field structure can be incorporated into spatial extrapolation by redefining the distance between points to include the difference in wind speed, and subsequently using IDW with this new definition of distance for the spatial extrapolation. In its standard implementation, IDW defines a location by its two-dimensional geographical coordinates (x, y). Now the difference in wind speed can additionally be calculated for any two locations, and added to the geographical distance.

Furthermore, the relative importance of the wind speed difference versus the geographical distance can be managed by using a tunable weight α. The wind speed component will be multiplied by this weight, while the geographical distance will be multiplied by $(1-\alpha)$. This weight, α, is a hyperparameter of the proposed Wind Speed-enhanced IDW (WS-IDW) method.

The definition of the correction factor for unseen location (x, y), denoted as $\hat{C}_{(x,y)}$, is given by Equation 3.8, which is analogous to the formulation given in Equation 3.7 but refers to a different distance definition. Equation 3.9 details the altered distance definition.

$$\hat{C}_{(x,y)} = \frac{\sum_{i=1}^{N} \frac{C_i}{d_{i,(x,y)}}}{\sum_{i=1}^{N} \frac{1}{d_{i,(x,y)}}} \tag{3.8}$$

$$d_{i,(x,y)} = (1-\alpha) \cdot \sqrt{(x-x_i)^2 + (y-y_i)^2} + \alpha \cdot \left| v_{(x,y)} - v_i \right| \tag{3.9}$$

where:

$\hat{C}_{(x,y)}$	The estimated correction factor for the unobserved location (x, y).
C_i	The correction factor calculated at the i-th observation location (the ratio of measured wind speed to LES-modeled wind speed at observation location i).
N	The total number of observation locations used for spatial extrapolation.
$d_{i,(x,y)}$	The Wind Speed-enhanced distance between the unobserved location (x, y) and the i-th observation location.
(x, y)	The geographical coordinates of the unobserved location where the correction factor is being estimated.
(x_i, y_i)	The geographical coordinates of the i-th observation location.
$v_{(x,y)}$	The LES-modeled wind speed at the unobserved location.
v_i	The LES-modeled wind speed at the i-th observation location.
α	A constant scalar (hyperparameter) used to weight the influence of the wind speed difference relative to the geographical distance.

It is important to note that the units used for the distance and the wind speed do not influence the performance of this method. Neither does scaling or normalization. These operations will influence the value of α that should be used, but given that these are only scalar operations they are undone by correctly setting α, which is also basically a scalar.

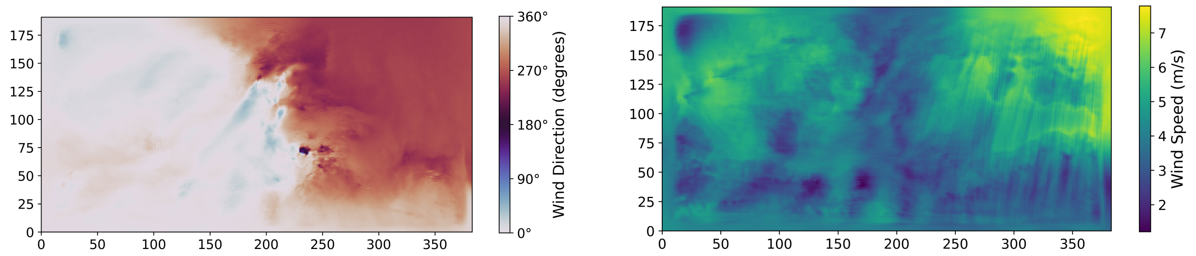
Furthermore it should be noted that although the weights for the geographical distance ($(1-\alpha)$) and the wind speed distance (α) by definition add up to 1, they do not represent percentages and should rather be thought of as a relative scalar. For example, using $\alpha = 0.5$ does not mean 50% weight is given to each, but rather that no relative scaling is performed and the original aspect between the two parts is maintained. Using $\alpha = 0.9$ would subsequently mean that the influence of the second part is scaled up by a factor 9 ($0.9/(1-0.9) = 9$).

Lastly, also note that this method assumes that the LES wind-speed field structure provides a reasonably accurate representation of the true flow field. This assumption is not guaranteed to hold, as the very goal of this research is to correct imperfections in the LES predictions. Nevertheless, the method benefits from its continuous weighting scheme rather than discrete regional boundaries. As a result, inaccuracies in the LES structure only gradually affect the extrapolation, causing weighting errors to scale proportionally with the deviation of the LES wind-speed field from reality.

Wind Direction

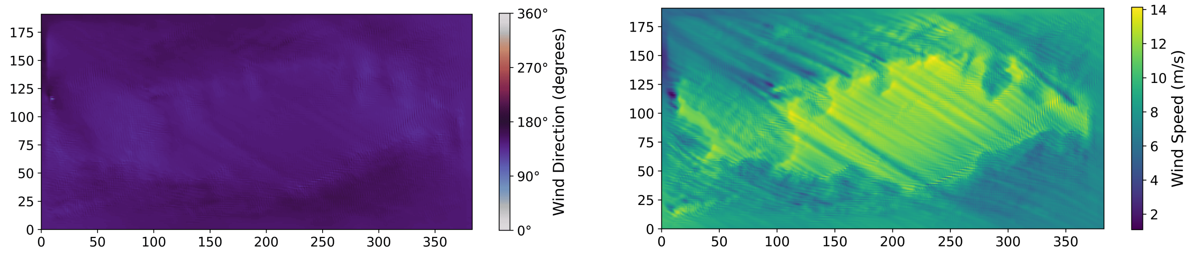
When correcting an unseen location based on correction factors found at observation locations, giving more weight to the correction factors at observation locations where a similar wind direction is modeled by the LES could prove useful. The importance of including wind direction in the input of the temporal correction model of the LES is hypothesized in Section 3.1.5 and shown in Section 4.1.2. It is therefore fair to assume that the required correction is dependent on the wind direction, and should thus be informed by it. Furthermore, in the IDW implementation for wind distributions, rather than for timeseries, of the industry software *OpenWind* [66] an additional factor is included in IDW that depends on the wind direction, indicating it’s usefulness in spatially extrapolating wind distributions. It is plausible that this transfers to timeseries prediction.

Visually inspecting the wind direction fields it is found that when the wind speed is sufficiently large there is very little variation in the wind direction across the site. For time steps where the wind speed is lower however, the variation in wind direction across the site is large, and the structures in its field can be used to guide the spatial extrapolation. An example of both scenarios at Site 2 can be seen in Figure 3.12.



(a) Wind-direction field (low wind speed)

(b) Wind-speed field (low wind speed)



(c) Wind-direction field (high wind speed)

(d) Wind-speed field (high wind speed)

Figure 3.12: Examples of LES wind-direction and wind-speed fields at Site 2 for two timesteps with contrasting wind conditions. The top row shows a timestep with low wind speeds where the wind direction varies substantially across the site, while the bottom row shows a high-wind case where the direction field is nearly uniform. These examples illustrate how wind direction structures may inform spatial extrapolation.

The correction factor at location (x, y), $\hat{C}_{(x,y)}$, is then still given by Equation 3.8, but the distance $d_{i,(x,y)}$ is now calculated using Equation 3.10. The resulting Wind Direction-enhanced IDW will be referred to as WD-IDW.

$$d_{i,(x,y)} = (1-\alpha)\sqrt{(x-x_i)^2 + (y-y_i)^2} + \alpha \cdot min(|dir_{(x,y)} - dir_i|, |dir_i - dir_{(x,y)}|) \tag{3.10}$$

where dir_i gives the wind direction modeled by the LES at the observation location i, and $dir_{(x,y)}$ gives the wind direction at the location (x, y). The wind directions are given in degrees, and because wind direction is cyclical the minimum of the two orderings is taken to find the smallest distance between two wind directions. α is then used to weight the influence of the wind direction (dis)similarity against the geographic distance. All other variables are the same as in Equation 3.9.

Note that the remarks regarding the units and scaling of the variables, the meaning of α, and the reliance on the correctness of the structure of the data produced by the LES are all equally valid for wind direction, as they are for wind speed.

3.2.5. Using Terrain Information

Similar to how the distance used in IDW can be extended with the wind speeds or the wind direction at each location, as outlined in the previous section, the distance can also be extended with terrain features. This will embed terrain information in the spatial extrapolation process. The terrain features investigated are elevation, defined as the ground height above sea level, and tree canopy height. The feature maps of these two features can be seen in Figure 3.8. The hypothesis is that locations with similar terrain require similar correction. In both cases the correction factor at (x, y) is given by Equation 3.8, but the distance $d_{i,(x,y)}$ is redefined as

$$d_{i,(x,y)} = (1-\alpha)\sqrt{(x-x_i)^2 + (y-y_i)^2} + \alpha \cdot (t_{(x,y)} - t_i) \tag{3.11}$$

where $t_{(x,y)}$ gives the value of the chosen terrain feature, either elevation or tree canopy height, at location (x, y) and t_i gives the value of the chosen terrain feature at observation location i. α is then used to weight the influence of the terrain (dis)similarity against the geographic distance. All other variables are the same as in Equation 3.9.

4

Results

This chapter presents the quantitative results obtained from the experiments designed in Chapter 3. The analysis is structured into two primary components: temporal extrapolation and spatial extrapolation. Section 4.1 focuses on the temporal dimension, beginning with a validation of LES-based MCP against a mesoscale baseline, followed by an evaluation of individual enhancements such as multi-step time windows and autoencoded wind fields, and culminating in the assessment of a combined correction strategy. Section 4.2 addresses spatial extrapolation, comparing the performance of the proposed flow-aware and terrain-aware IDW variants against the industry-standard traditional IDW method.

4.1. Temporal Extrapolation

This section evaluates a range of strategies for temporal extrapolation, the task of correcting the prediction at a given observation location (OL) at an unseen time. First, the results of the baseline methods are presented and compared when using mesoscale data and when using LES data as the reference source. Thereafter performance of the five baseline estimators from Section 3.1.2 is reported individually for the OLs of Site 1, followed by experiments on five design choices evaluated on these same OLs: adding input features, using data from additional LES area, using a time window, training a joint multi-location model, and using a convolutional autoencoder (CAE) encoding of the wind-speed field. Finally, the most effective elements are combined into two combined strategies, and results are presented as absolute scores and percentage improvements for MAE, RMSE, and R^2. The design of the composite strategy is guided by the input dimensionality in Equation 3.6.

The evaluation of strategies in isolation is done for the 3 OLs of Site 1. The evaluation of the combined strategies, as well as the initial comparison to using mesoscale NWP data, is done for all OLs of Site 1, Site 3 and Site 4.

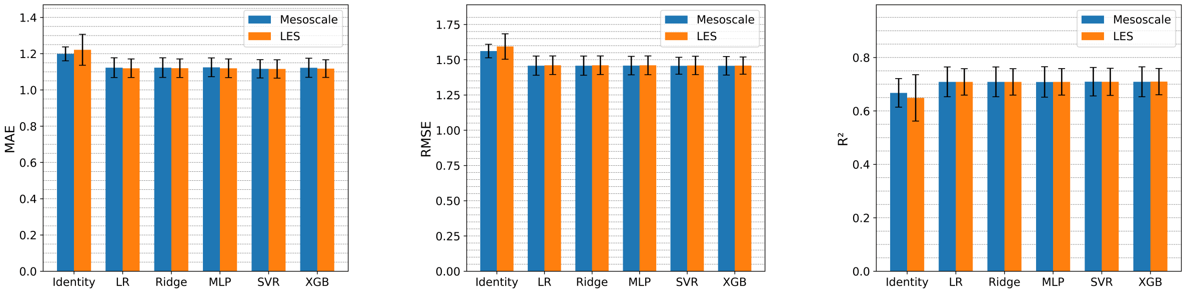
4.1.1. Mesoscale vs. LES as MCP Reference Data

The performance of the baseline MCP models was evaluated using both mesoscale and LES data as the long-term reference source. The results, aggregated over the six Observation Locations, are presented in Figure 4.1.

The uncorrected baselines, indicated with *identity*, reveal a clear difference between the two reference sources. The Identity case based on the LES time series exhibits higher MAE and RMSE and a lower coefficient of determination than the identity based on mesoscale data. This indicates that the raw mesoscale time series aligns better with the measurements than the raw LES series before any correction is applied.

Applying the MCP models yields consistent improvements across all estimators for both reference sources. Linear and nonlinear models reduce MAE and RMSE and increase R^2 to a similar degree. After correction, the performance obtained using mesoscale and LES as the reference is essentially the same. The two sets of corrected results are nearly indistinguishable across all three metrics, and their error bars overlap fully for every estimator.

Aggregated over all estimators, LES-based corrections improve MAE, RMSE and R^2 by 8.42%, 8.39% and 9.27% respectively, while mesoscale-based corrections improve these metrics by 6.43%,



(a) Mean Absolute Error (MAE) (b) Root Mean Square Error (RMSE) (c) Coefficient of Determination (R^2)

Figure 4.1: **Mesoscale versus LES as reference data source for baseline MCP models.** Performance of the baseline MCP models when using mesoscale data (blue) or LES data (orange) as the reference time series. Bars show the mean MAE, RMSE and R^2 aggregated over six observation locations from three sites; error bars indicate the corresponding standard deviation. In addition to the five correction models, the identity mapping is included to indicate the accuracy of the mesoscale and LES time series before any correction.

6.66% and 6.20% respectively. Aggregating further over metrics as well gives an average improvement of 8.69% for LES and 6.43% for mesoscale.

The five estimators behave consistently in this comparison. Ridge, Linear Regression, MLP, SVR and XGB achieve comparable performance, with only minor variations between them. This uniformity appears for both reference sources and across all metrics, which indicates that the MCP task has low sensitivity to the choice of model.

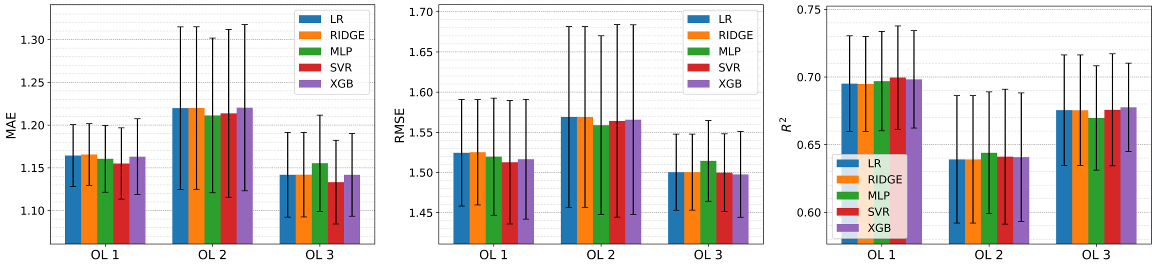
4.1.2. Baselines

The performance of the baseline methods, as described in Section 3.1.2, can be found in Figure 4.2 and Table 4.1.

The performance of the five baseline estimators (LR, Ridge, MLP, SVR, XGB) across the three observation locations at Site 1 is presented in Figure 4.2 and Table 4.1. In this baseline scenario, where the models rely solely on the LES wind speed at the observation location as input, two main trends are observed.

First, performance varies noticeably between observation locations. OL 1 yields the highest R^2 values approaching 0.70, while OL 3 obtains the lowest MAE and RMSE scores. OL 2 shows the lowest performance for all three metrics.

Second, the choice of estimator has a minimal impact on performance in this baseline setup. The complex non-linear models (MLP, SVR, XGB) do not meaningfully outperform the linear baselines (LR, Ridge). This suggests that when inputs are restricted to the local modeled wind speed, the relationship between the LES and the observations is predominantly linear, leaving little room for non-linear models to extract additional gain without further contextual information.

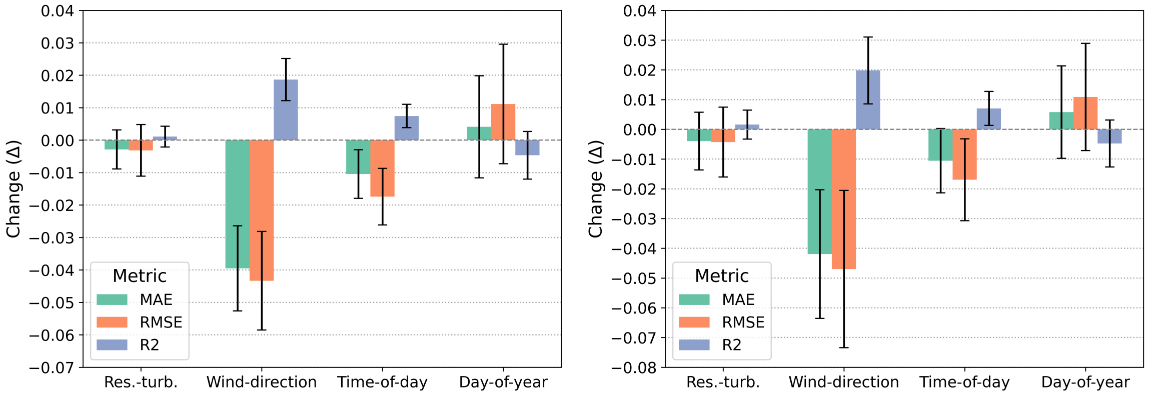


(a) Mean Absolute Error (MAE) (b) Root Mean Square Error (RMSE) (c) Coefficient of Determination (R^2)

Figure 4.2: Performance of the baseline models (LR, Ridge, MLP, SVR, XGB) for Temporal Extrapolation for each of the three observation locations. The displayed mean value and standard deviation (error bars) for each model are aggregated from the scores obtained across the six individual cross-validation folds.

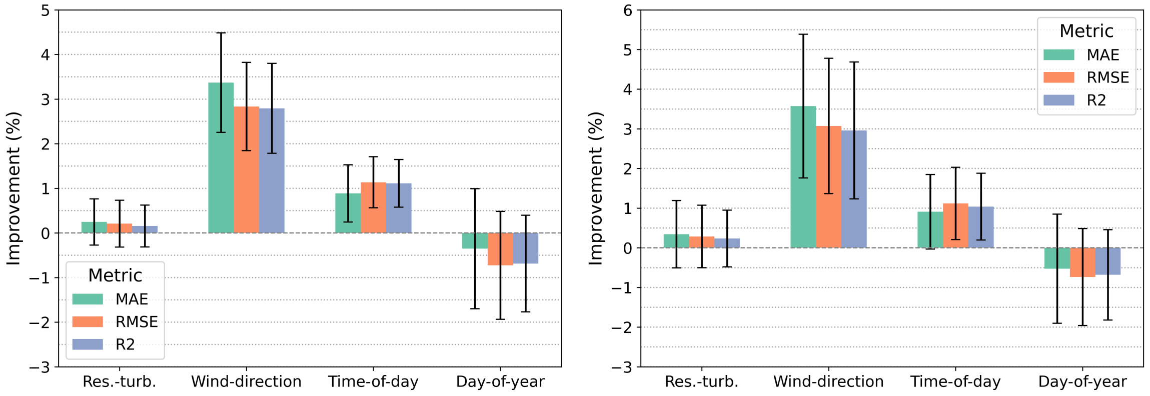
Table 4.1: Temporal Extrapolation Baselines Performance for Observation Locations (OL) 1, 2 and 3

OL	Method	MAE	RMSE	R^2
1	LR	1.1644 ± 0.0361	1.5245 ± 0.0663	0.6951 ± 0.0353
	Ridge	1.1656 ± 0.0360	1.5252 ± 0.0655	0.6948 ± 0.0350
	MLP	1.1606 ± 0.0391	1.5196 ± 0.0729	0.6969 ± 0.0367
	SVR	1.1551 ± 0.0417	1.5127 ± 0.0769	0.6995 ± 0.0382
	XGB	1.1631 ± 0.0444	1.5164 ± 0.0745	0.6983 ± 0.0360
2	LR	1.2199 ± 0.0950	1.5690 ± 0.1125	0.6391 ± 0.0471
	Ridge	1.2199 ± 0.0950	1.5690 ± 0.1124	0.6391 ± 0.0471
	MLP	1.2113 ± 0.0905	1.5588 ± 0.1112	0.6439 ± 0.0450
	SVR	1.2137 ± 0.0981	1.5642 ± 0.1198	0.6411 ± 0.0499
	XGB	1.2203 ± 0.0972	1.5656 ± 0.1181	0.6407 ± 0.0475
3	LR	1.1418 ± 0.0495	1.5003 ± 0.0474	0.6755 ± 0.0408
	Ridge	1.1419 ± 0.0494	1.5004 ± 0.0473	0.6754 ± 0.0408
	MLP	1.1553 ± 0.0563	1.5144 ± 0.0502	0.6697 ± 0.0385
	SVR	1.1332 ± 0.0490	1.4996 ± 0.0485	0.6757 ± 0.0414
	XGB	1.1418 ± 0.0485	1.4975 ± 0.0534	0.6776 ± 0.0326



(a) Fixed-Baseline Analysis (Δ Absolute)

(b) Context-Dependent Analysis (Δ Absolute)



(c) Fixed-Baseline Analysis (Δ Percentage)

(d) Context-Dependent Analysis (Δ Percentage)

Figure 4.3: **Isolated impact of additional features on model performance.** The plots show the mean change (Δ) in performance metrics resulting from adding a specific feature, aggregated across all model and location combinations. The analysis is presented using both absolute change (Δ) and percentage change (Δ%). **Top Row (a, b):** Absolute change in metrics. **Bottom Row (c, d):** Percentage change in metrics. **Fixed-Baseline Analysis (a, c):** Quantifies the isolated, marginal benefit of adding a feature to the baseline set (wind speed only). **Context-Dependent Analysis (b, d):** Quantifies the isolated, marginal benefit of adding a feature to any set not containing it already. In plots (a) and (c), negative Δ in MAE and RMSE, and positive Δ in R^2, indicates performance improvement.

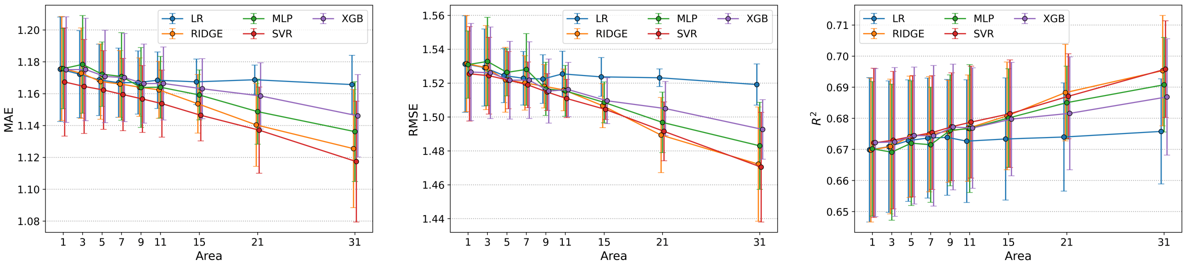
4.1.3. Additional Input Features

Adding additional features to the input given to the prediction models causes a change in performance of those prediction models. The absolute and relative impact each feature has on the three evaluation metrics used (MAE, RMSE and R^2) when added to the baseline set (wind speed only) is visualized in Figures 4.3a and 4.3c respectively. The observed impact to the evaluation metrics is aggregated over the 3 different OLs of Site 1 and the 5 different models by taking the mean and standard deviation. For the RMSE and MAE a negative delta signifies an improvement, while for R^2 an improvement is indicated by a positive delta. Figures 4.3b and 4.3d visualize the absolute and relative impact of adding a specific feature to any possible feature set not already containing it. The observed impact is thus aggregated over the 3 different locations, 5 different models, and 8 feature sets to which the target feature is added. Both Figures show that the addition of resolved-turbulence (in the wind speed), wind-direction and time-of-day increase the performance on all metrics, whilst the addition of day-of-year decreases performance on all metrics. The delta obtained by adding wind-direction is most significant, giving an improvement of 3%, followed by time-of-day, giving an improvement of 1%. The improvement given by addition of resolved-turbulence is small (< 0.5%), with a standard deviation larger than the mean improvement, indicating that although the addition of this feature in general gives an improvement, it also worsens performance for some cases. The complete results for every individual (OL, Model, Input Feature set)-combination can be found in Appendix A.2.

The results shown in Figure 4.3 warrant a detailed examination of the feature set *{wind-speed, resolved-turbulence, wind-direction, time-of-day}*. This specific four-feature combination proved to be the optimal feature set in 15 out of 45 unique experiments (across 5 models, 3 locations, and 3 metrics). For the remaining 30 experiments where it was sub-optimal, the performance gap between this set and the truly optimal set remained insignificant, with the mean difference (Δ) in MAE, RMSE and R^2 being 0.0033, 0.0050 and 0.0020 respectively.

4.1.4. Additional Area

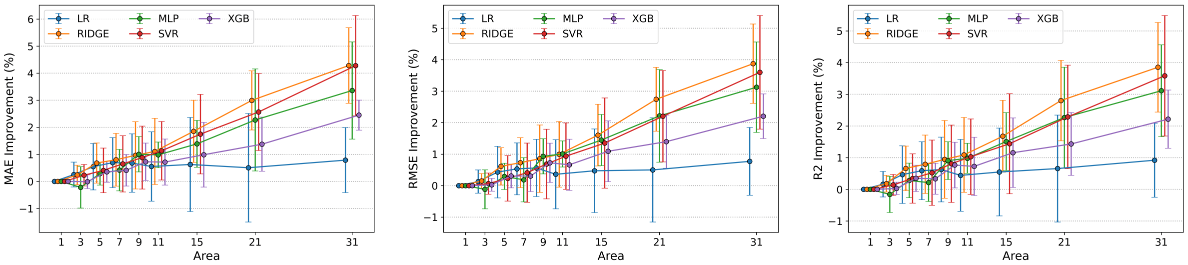
The effects of including a larger LES area in the model input on the absolute metric score, and relative improvement with regard to the baseline, are shown in Figures 4.4 and 4.5 respectively. When aggregating results over the three OLs, a clear and seemingly linear improvement can be observed as the input area increases. This trend holds consistently across all model types (except for LR) and all metrics. By further averaging the relative improvements across MAE, RMSE, and R^2 at area size 31, a single representative improvement per model type is obtained. This gives, from smallest to largest average improvement, LR: 0.8%, XGB: 2.3%, MLP: 3.2%, SVR: 3.8% and Ridge: 4.0% improvement over the baseline.



(a) MAE (b) RMSE (c) R^2

Figure 4.4: Performance of the baseline models when including spatial information from increasingly larger areas around the observation locations. Results are aggregated over the three observation locations, with the marker indicating the average and the error bars showing the standard deviation. For each model-type performance is shown in terms of MAE, RMSE, and R^2.

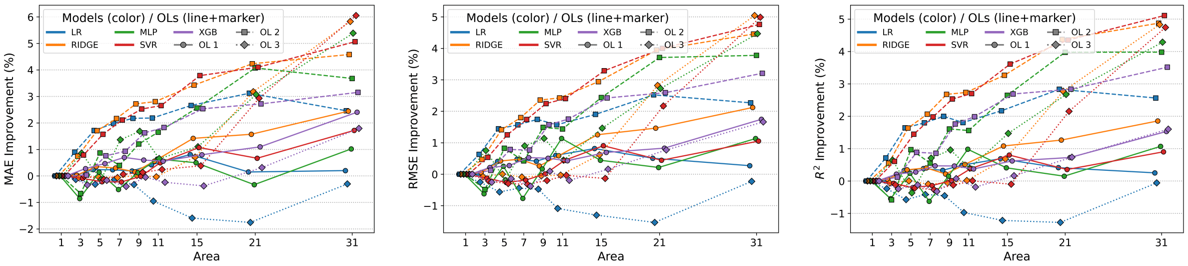
However, when examining the results per observation location individually, the improvement pattern shows itself to be less uniform. The improvement obtained for each OL individually is shown in Figure 4.6. The counterpart of this Figure showing absolute metric values can be found in Appendix A.3. For OL 1 the gains are noticeably smaller than those seen in the aggregated results. For OL 2 the improvement increases steeply at smaller area sizes but begins to flatten for the largest areas. In contrast, OL 3 shows little improvement up to an area size of 15×15, after which the performance improves sharply for areas 21×21 and 31×31. These differences indicate that the aggregated trend



(a) MAE Improvement (b) RMSE Improvement (c) R^2 Improvement

Figure 4.5: Relative performance improvement of each model and area size compared to the baseline (area = 1). Improvements are expressed as a percentage change in mean metric values, disregarding the standard deviation. Results are aggregated over the three observation locations, with error bars representing the standard deviation of the found improvement percentages.

masks variability between locations, as each observation location responds differently to increasing area sizes.



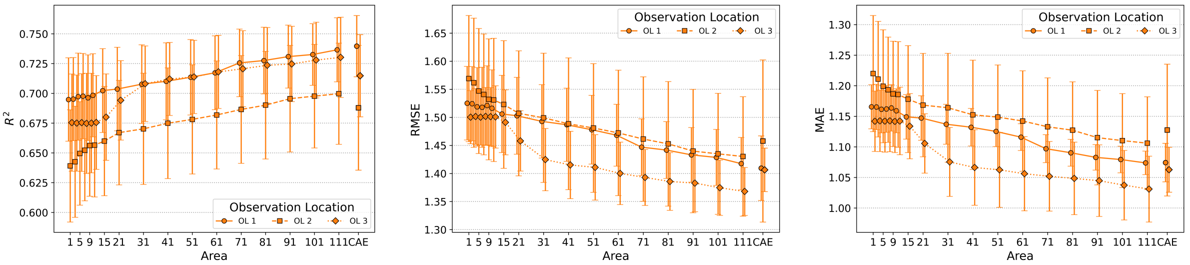
(a) MAE Improvement (b) RMSE Improvement (c) R^2 Improvement

Figure 4.6: Relative performance improvement of each model and observation location as a result of using a larger area compared to the baseline (area = 1). Improvements are expressed as a percentage change in mean metric values, disregarding the standard deviation.

The observed improvements show that using information from surrounding LES cells enhances performance compared to using only the single cell containing the OL. This demonstrates that neighboring cells contain relevant information not fully represented by the central cell alone. However, looking at Figure 4.6 it can be seen that the improvement obtained varies significantly between locations and models.

It should be noted that the input dimensionality increases quadratically with the area size. For an area of 31×31, the number of input variables grows by a factor of $31^2 = 961$ compared to the baseline. This substantially increases training times. The XGB model is most affected, with training times rising from roughly 5 minutes for the baseline (1×1) setup to around 8 hours for the 31 × 31 area size. Although training time is not a performance criterion, since in a real world WRA project a model only needs to be trained once, this computational cost is important to acknowledge and the quadratic relation does pose a practical limit. This is also why simply giving the entire LES domain is not feasible for most models.

Figure 4.7 shows the impact that adding increasingly large area sizes to the prediction input has on performance. This analysis aims to determine what size increase continues to add useful information and to potentially identify the range at which additional spatial context becomes uninformative. Because extremely large areas cause the input space to grow quadratically, rendering complex baseline models computationally infeasible, only Ridge Regression was used to explore this dynamic up to the theoretical limit of the domain. It can be seen that performance keeps increasing up to the very largest possible square area around the Observation Locations (111 × 111 cells) before reaching the border of the LES domain, without reaching a distinct plateau. A consistent, nearly linear improvement is observed across all area sizes, extending up to the maximum tested size of 111. Also included for comparison is the performance obtained by Ridge when, instead of the direct additional area, it is provided with a wind-speed field encoding of size 128.



(a) R^2

(b) RMSE

(c) MAE

Figure 4.7: Performance of Ridge regression for increasing LES area size up to 111. Shown are R^2, RMSE, and MAE across the three observation locations. The rightmost point (CAE) shows the performance when the additional area information is provided via a CAE wind-speed-field encoding with latent dimension 128 instead of direct grid-cell inputs.

4.1.5. Additional Time Steps

The use of additional time steps is found to significantly improve the performance of all models for all observation locations tested, as indicated by all of the used evaluation metrics. Figure 4.8 shows the performance in absolute terms for all each of the five models for all three locations of Site 1 when an increasing number of time steps is included in the input. The figure shows the mean and standard deviation of the MAE, RMSE and R^2 from cross-validation.

From this figure it becomes clear that as the time window increases in size from 1 (the baseline) to a time window of size 11, which includes the 5 preceding time steps, the target time step, and the 5 following time steps, results in a significant performance increase on all metrics, model and OLs.

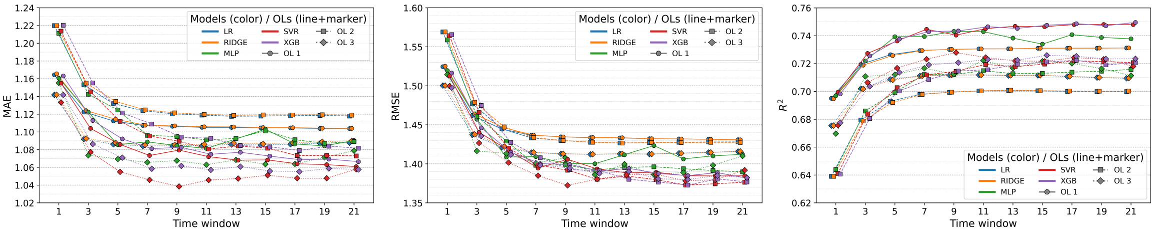
Continuing to use larger time windows does not yield a further increase for both LR and for Ridge, both of which plateau after this point in all cases.

The use of a time window larger than 11 causes the performance to drop slightly for MLP at OL 1 and OL 3, compared to a time window of size 11. For OL 2 the same deterioration of MLP is not seen, but neither is a further improvement.

For both SVR and XGB the best performance is most commonly observed for a time window of size 19 or 21. Although the improvement compared to a time window of size 11 is small, it can be observed across most of the results. The exception being the performance of SVR at OL 3, which is best at a time window of 9.

The improvements are quantified in terms of percentages in Figure 4.9. Improvements as large as 12% are found for OL 2 across all metrics for both SVR and XGB models. At OL 1 and OL 3 improvements up to 9.0% and 8.5% are found respectively. Similar to Figure 4.8, Figure 4.9 shows that the a significant improvement is observed across the board for up to a time window of size 11, after which further improvements are absent or diminishing.

Appendix A.4 contains a complete account of the individual results of all experiments.

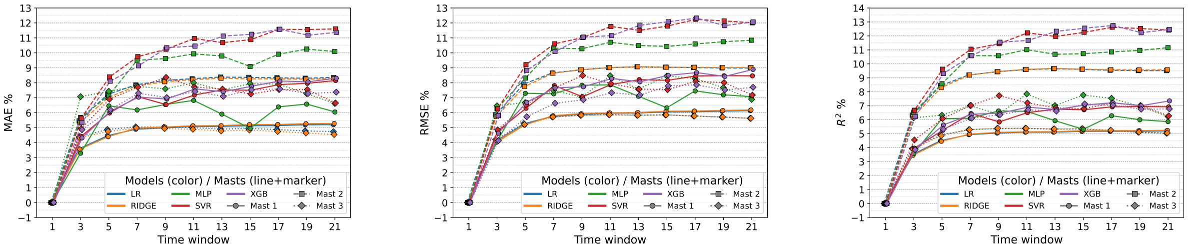


(a) Mean Absolute Error (MAE)

(b) Root Mean Square Error (RMSE)

(c) R^2

Figure 4.8: Performance of the baseline models adapted to include input features from an increasing time window. The figure is organized by performance metric (MAE, RMSE, and R^2), with each plot showing the results for all three observation locations (OLs). The x-axis 'Time window' represents the number of (hourly) time steps used in the model input, centered on the prediction time step. The markers represent the mean metric value, and the error bars show the standard deviation from cross-validation.

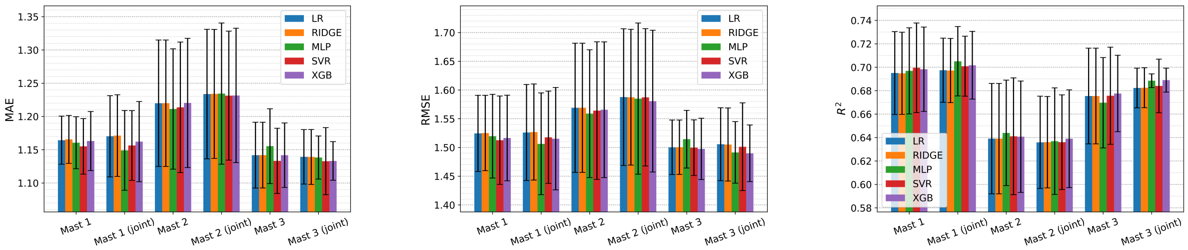


(a) MAE Improvement (%)

(b) RMSE Improvement (%)

(c) R^2 Improvement (%)

Figure 4.9: **Percentage Improvement from Increasing Time Window.** The plots show the percentage improvement in the mean of the performance metrics (MAE, RMSE, and R^2) compared to the baseline model (Time Window $= 1$). The x-axis 'Time window' represents the number of (hourly) time steps used in the model input, centered on the prediction time step. The improvement percentage is determined as the absolute difference over the baseline metric score.



(a) Mean Absolute Error (MAE)

(b) Root Mean Square Error (RMSE)

(c) Coefficient of Determination (R^2)

Figure 4.10: **Joint Training vs. Single Mast Performance.** This figure compares the performance metrics (MAE, RMSE, and R^2) for models trained exclusively on data from a single mast (e.g., **Mast 1**) versus models trained on data aggregated from all three locations (e.g., **Mast 1 (joint)**). The "joint" models are single models (one per model type) whose performance is evaluated on the individual mast test sets to allow for direct comparison with the baseline single-mast models.

4.1.6. Joint Multi-Location Prediction

See Figure 4.10. From Figure 4.10 it is clear that the joint model does not on average outperform the baseline. Aggregating over all experiments, on average the joint model performs 0.26% worse than the baseline. Therefore a further detailing of results is omitted.

4.1.7. Autoencoded Wind-speed Field

Figure 4.12 illustrates the absolute performance metrics (MAE, RMSE, and R^2) across the three OLs as a function of the Latent space Dimensionality (LD). To quantify the relative gains, Figure 4.13 presents the percentage improvement over the baseline for each metric. The baseline model, corresponding to a latent dimension of None, represents the performance without the CAE-encoded wind-speed field included in the model input. The ability of the CAE to compress and reconstruct the LES wind-speed fields was quantified using the Mean Squared Error (MSE) loss. The reconstruction losses for the tested LDs are: LD=8 (0.0998), LD=16 (0.0684), LD=32 (0.0557), LD=64 (0.0531), and LD=128 (0.0737). The difference in MSE between models is not clear when visually comparing the reconstructed wind-speed fields, indicating that the difference is small. An example of an original LES field and the reconstructions after encoding and decoding by the CAE models is given in Figure 4.11.

The inclusion of the CAE-encoded wind-speed field leads to a significant improvement in prediction accuracy across all models and OLs of Site 1 when compared to the baseline. For MLP, SVR and XGB the largest improvement is seen when going from the baseline (LD = None), to a latent dimension of size 8, for all three performance metrics. Further increasing the latent dimension has mixed effects, with in some cases a small further improvements, while in others the same level of performance is maintained or performance even deteriorates. LR and Ridge show a different pattern, improving more steadily as the latent dimensionality grows. This pattern for LR and Ridge is universal across the OLs and performance metrics, which are in all cases best when LD = 128.

The largest improvements, of up to 11% for MAE and no lower than 8.5% for the other metrics, are seen for LD=128, using XGB on OL 1 and OL 2.

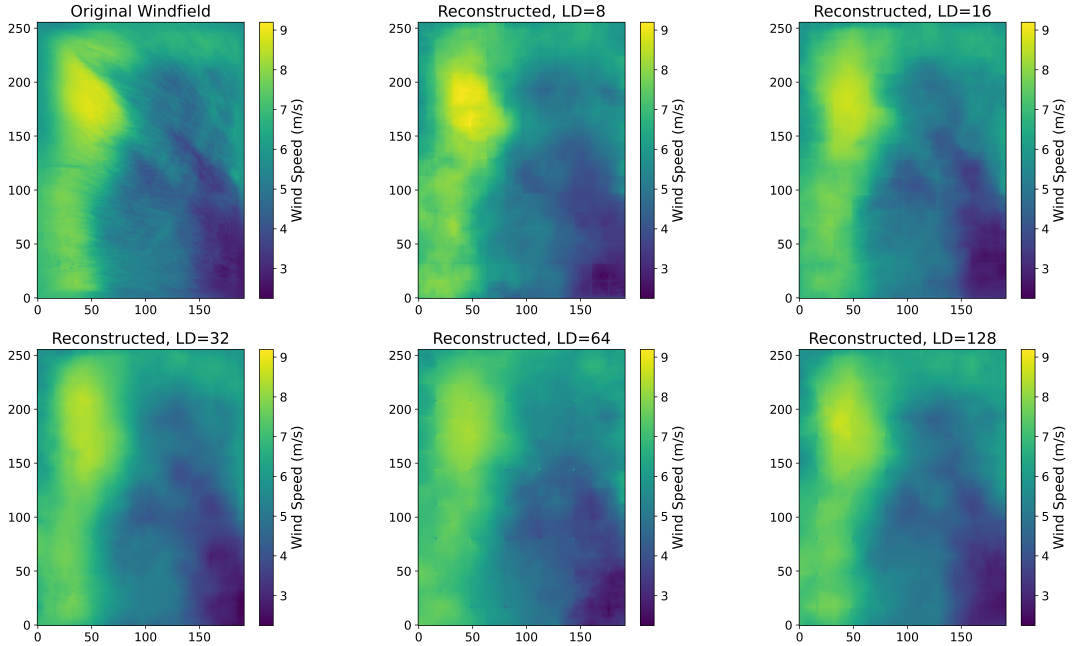
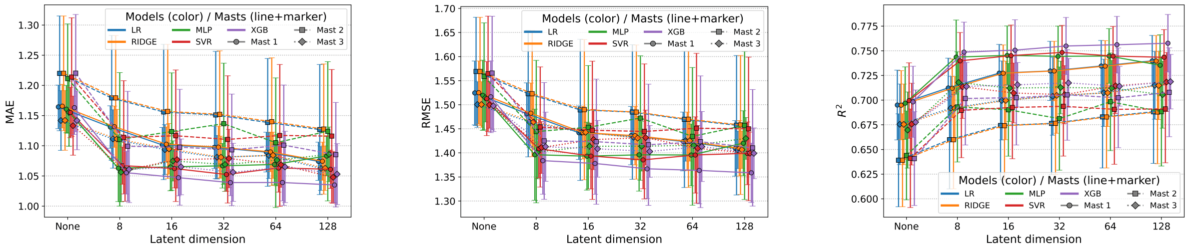
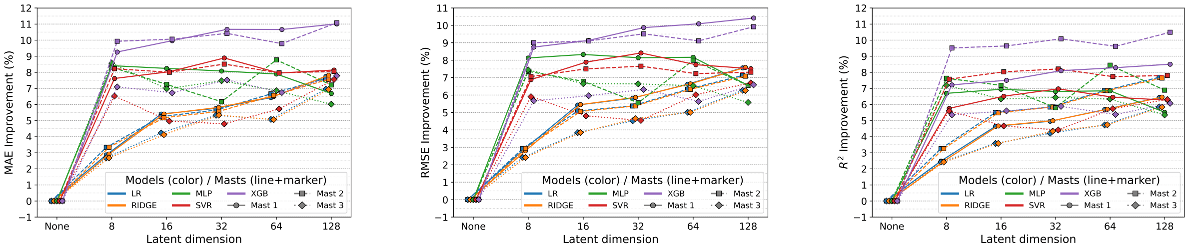


Figure 4.11: Reconstruction of an LES windfield from Site 1 using Convolutional AutoEncoders (CAEs) with varying latent dimension sizes (LD). The top-left panel shows the original LES windfield, while the remaining panels display the reconstructed windfields after encoding and decoding through CAEs with LD = 8, 16, 32, 64, and 128. The color scale indicates the modeled wind speed in m/s and is consistent across all windfields.



(a) MAE (b) RMSE (c) R^2

Figure 4.12: Performance of the baseline models adapted to include the windfield encoding as given by CAEs with various latent dimensions. The Figure is organized with Observation Locations (OL 1, OL 2, OL 3) in rows and performance metrics (MAE, RMSE, R^2) in columns. The x-axis 'Latent dimension' represents the dimensionality of the latent space of the CAE used to encode the input windfield(s). The markers represent the mean metric value, and the error bars show the standard deviation from cross-validation.

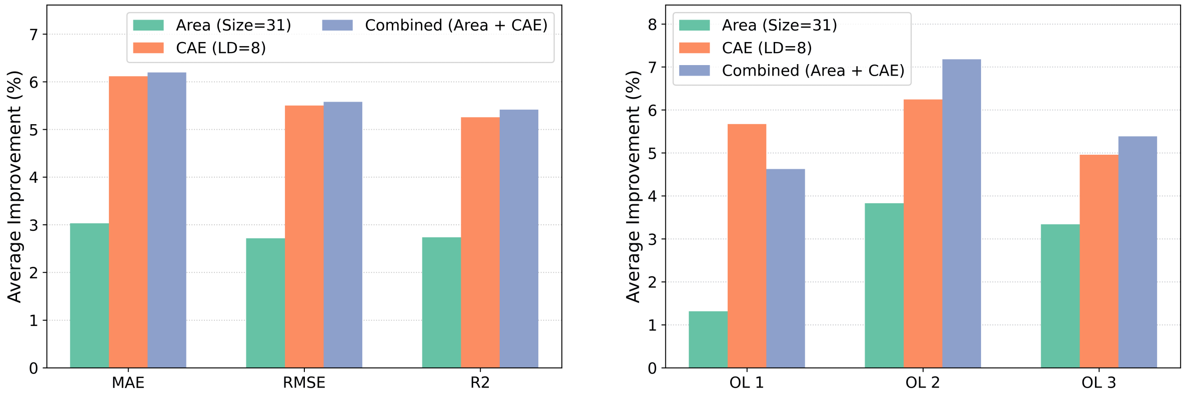


(a) MAE Improvement (%) (b) RMSE Improvement (%) (c) R^2 Improvement (%)

Figure 4.13: Percentage Improvement from CAE Latent Dimension. The plots show the percentage improvement in the mean of the performance metrics (MAE, RMSE, and R^2) compared to the baseline model (Latent Dimension = $\mathbf{None}$, i.e., no CAE input). The x-axis 'Latent dimension' represents the number of dimensions in the compressed windfield representation (the CAE bottleneck). The improvement percentage is determined as the absolute difference over the baseline metric score.

Relation Between Additional Area and CAE Strategies

To investigate how the spatial information captured by the Additional Area and CAE strategies relate, an additional comparison was performed. Three model configurations were evaluated: one using only the largest additional area (31×31), one using only the CAE with latent dimension 8, and one combining both approaches. The resulting average improvements are shown in Figure 4.14, with Subfigure 4.14a showing aggregated results per metric and Subfigure 4.14b showing aggregated results per observation location.



(a) Average improvement (%) for each Metric

(b) Average improvement (%) for each Observation Location.

Figure 4.14: Comparison of average improvement percentages for different setups (Area, CAE, Combined). Subfigure (a) shows aggregated results per metric, while Subfigure (b) shows aggregated results per observation location.

Aggregating results over model types and locations reveals the average performance of each method per metric (Figure 4.14a). Across all metrics, the CAE alone yields roughly twice the improvement of the explicit 31×31 area input. Combining both approaches results in only a marginal gain compared to using the CAE alone, with an average increase of just 0.09%.

However, examining the results by observation location (aggregated over models and metrics) reveals distinct spatial behaviors. For OL 2 and OL 3, the combined approach leads to a modest but measurable improvement over the CAE alone, with gains of 0.94% and 0.43% respectively. This aligns with the finding that the stand-alone Area strategy performs significantly better at these locations than at OL 1. Consequently, at OL 1, where the explicit Area input adds little value, the combined approach fails to outperform the CAE-only configuration.

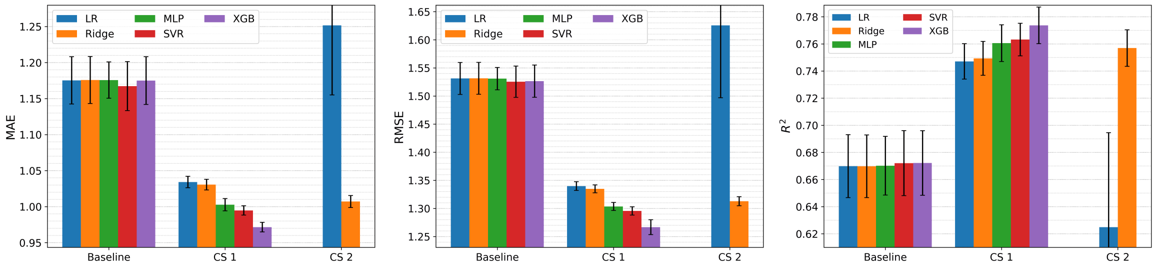
4.1.8. Combining Strategies

Based on the results found for the individual experiments a model is designed that combines the strategies that are found to work. Since the combination of multiple strategies can quickly inflate the number of input features, strategy hyperparameters are chosen in a manner that tries to maximize improvement, whilst minimizing the input feature dimensionality.

Combination Strategies

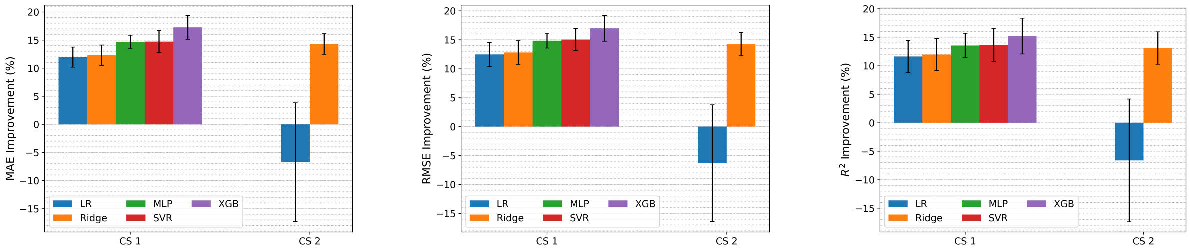
Section 4.1.2 shows that adding the features resolved-turbulence, wind-direction and time-of-day increase performance, so these features are included in the model input. Although Section 4.1.3 shows that the use of a sufficiently large additional area (31 x 31) does significantly improve performance, the results in Section 4.1.7 show that when CAE encoding is used this practically subsumes this improvement. Therefore no additional area is directly included. Section 4.1.5 shows that a time window of size 11 gives optimal or near-optimal performance in all cases, so a time window of size 11 is used. Furthermore, a latent dimensionality of 8 is chosen for the CAE, as Section 4.1.7 shows that this already provides the majority of the performance increase for MLP, SVR and XGB. Lastly, the strategy to train a single model to predict for multiple observation locations is not used, as this was shown to not increase performance in Section 4.1.6. This combination of strategies is referred to as Combination Strategy 1 (CS 1).

Because LR and Ridge are shown to gain most from the use of a CAE if the LD is set to 128, a second Combination Strategy (CS 2) using LD=128 will be tested for these linear models.



(a) MAE (b) RMSE (c) R^2

Figure 4.15: **Absolute performance of Combined Strategy (CS) models, alongside the baseline performance.** The plots show the metric scores (MAE, RMSE, and R^2) achieved by CS 1 and CS 2 models aggregated over the three OLs. The combined strategies use the feature-set {wind-speed, resolved-turbulence, wind-direction, time-of-day}, a time window of size 11, and a CAE-encoded windfield with latent dimension (LD) of 8 and 128 for CS 1 and CS 2 respectively. Lower scores are better for MAE/RMSE, and higher scores are better for R^2.



(a) MAE Improvement (%) (b) RMSE Improvement (%) (c) R^2 Improvement (%)

Figure 4.16: **Percentage Improvement from Combined Strategies (CS) 1 and 2 over the baseline models.** The combination strategy models use the feature-set wind-speed, resolved-turbulence, wind-direction, time-of-day, a time window of size 11, and a CAE-encoded windfield with latent dimension (LD) of 8 and 128 for CS 1 and CS 2 respectively. The results are aggregated over the three OLs, and the mean and standard deviation of this aggregation is given.

The performance achieved by these combination strategies on Site 1 for the different models is presented in absolute terms in Figure 4.15 and compared with the performance of the baseline models in Figure 4.16. It can be clearly seen that XGB benefits most from the applied strategies, resulting in an average improvement of 16.5%. Aggregating improvements across all model types, locations and metrics an average improvement of 13.9% is found for CS 1. The average improvement percentage and its standard deviation for each model and each metric, aggregated over the three observation locations, can be found in Table 4.2. For Ridge it can be observed that performance is slightly higher for CS 2 than for CS 1, benefiting from the additional information it can extract from the larger LD used in CAE. For LR the opposite is seen, as it clearly suffers from the much larger input dimensionality, degrading its performance to below the level of the baseline.

To compare the improvement obtained by CS 1 to the improvements obtained by the individual strategies in isolation they are listed side by side in Table 4.3. This table shows the improvement resulting from using the additional feature set wind-direction, resolved-turbulence, time-of-day, using a time window of size 11, using a CAE with latent dimensionality of size 8, and the improvement from using the combination of all three strategies. Furthermore, it also indicates what the improvement would be if the individual improvements were to be completely disjunct. This is calculated as

$$impr_{disjunct} = 1 - (1 - impr_{feats}) * (1 - impr_{tw}) * (1 - impr_{cae}) \tag{4.1}$$

As seen in Table 4.3 the improvement obtained by CS 1 is 3% lower than if the individual improvements relied on completely disjunct information. CS 1 achieves a large and significant improvement.

The performance of CS 1 is also evaluated on Site 3 and Site 4. Figure 4.17 shows the performances obtained by the baseline implementation and by the use of CS 1 on observation location 4 (OL 4), which is in Site 3. An improvement of between 9% and 14% is found across model types and metrics. SVR

Table 4.2: Performance Improvement over the Baseline of Models Using Combined Strategies. The mean improvement in percentages over the baseline ± the standard deviation from aggregating across OLs.

Strategy	Estimator	MAE (%)	RMSE (%)	R^2 (%)
CS 1	LR	11.960 ± 1.788	12.464 ± 2.075	11.618 ± 2.799
	RIDGE	12.303 ± 1.804	12.803 ± 2.040	11.967 ± 2.797
	MLP	14.692 ± 1.162	14.837 ± 1.274	13.548 ± 2.134
	SVR	14.726 ± 1.958	15.038 ± 1.913	13.654 ± 2.904
	XGB	$\mathbf{17.263 \pm 2.127}$	$\mathbf{16.980 \pm 2.231}$	$\mathbf{15.199 \pm 3.135}$
CS 2	LR	-6.736 ± 10.577	-6.326 ± 10.091	-6.607 ± 10.771
	RIDGE	$\mathbf{14.299 \pm 1.835}$	$\mathbf{14.237 \pm 1.995}$	$\mathbf{13.101 \pm 2.838}$

Table 4.3: Percentage improvement over the baseline for different correction strategies. Results are aggregated over model types and locations.

Strategy	MAE (%)	RMSE (%)	R^2 (%)
Additional Features	5.03	4.51	4.35
Time Window (size = 11)	7.51	8.21	7.79
CAE (LD = 8)	6.12	5.50	5.26
If Completely Disjunct	*17.54*	*17.17*	*16.44*
CS 1	14.19	14.42	13.20

and XGB are seen to improve most from the use of CS1, both achieving improvements between 12% and 14% across metrics. The average improvement of XGB aggregated across metrics is 13.0%. Aggregating over all model types and metrics the average improvement is 11.4%.

Figure 4.18 shows the performances obtained on OL 5 and OL 6 of Site 4, by both the baseline implementations and the CS1 models. Note that the improvements obtained for these locations are lower than those seen for the OLs at Site 1 and Site 3. LR benefits least from CS1 compared to its baseline, with improvements across metrics and OLs varying between 4% and 7%. XGB again achieves the highest performance increase, ranging from 6.7% to 12.5%, averaging at 9.8%, as well as the highest performance in absolute terms. The other models land somewhere between LR and XGB in terms of performance, showing a consistent ranking: LR < Ridge < MLP < SVR < XGB. Aggregating over model types, metrics and OLs the average improvement is 7.2%.

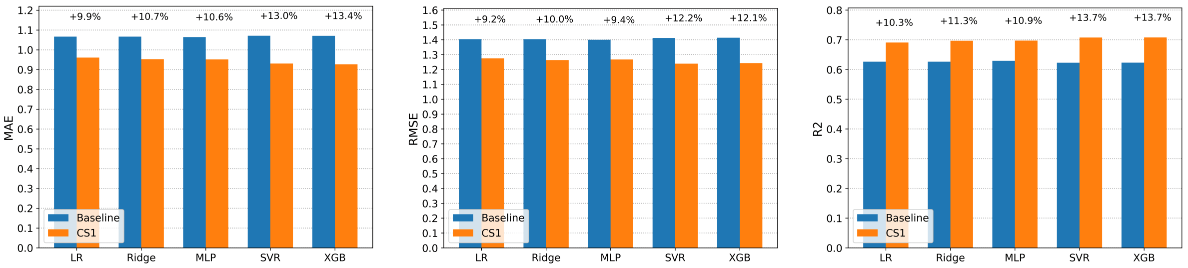
All obtained performances are given in Appendix A.7.

Out-of-Distribution Behavior

It is noted that while the non-linear models (SVR, XGB, MLP) demonstrate superior performance within the range of the training data, their behavior is unbounded regarding out-of-distribution (OOD) inputs. Empirical experimentation indicate that when presented with wind speeds significantly higher than those in the training set, SVR predictions tend to converge towards the mean, while XGB predictions plateau at the maximum wind speed encountered during training. The behavior of MLP was not found to adhere to such a predictable behavior pattern. However, a simple solution is available, and will be detailed in Section 5.1.7.

4.2. Spatial Extrapolation

In this section, the results of the experiments concerning spatial extrapolation are presented. First, Section 4.2.1 establishes the performance of the industry-standard baseline, Inverse Distance Weighting (IDW). Following this, Section 4.2.2 evaluates the proposed enhancements that incorporate the LES wind field structure, specifically utilizing wind speed and wind direction similarity. Finally, Section 4.2.3 assesses the impact of integrating terrain features, namely elevation and tree canopy height, into the extrapolation weighting.

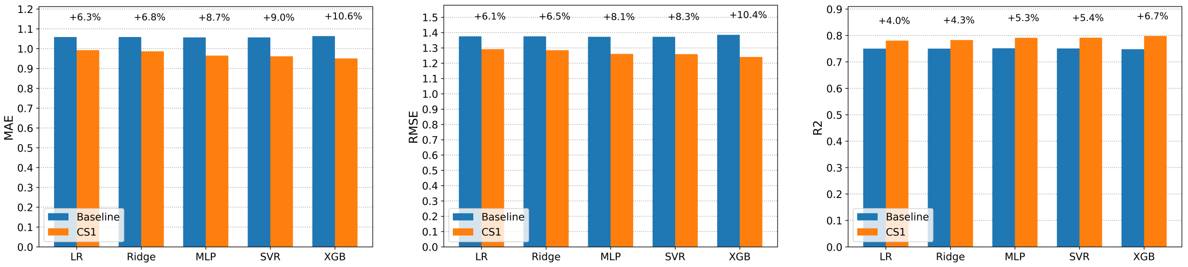


(a) MAE

(b) RMSE

(c) R^2

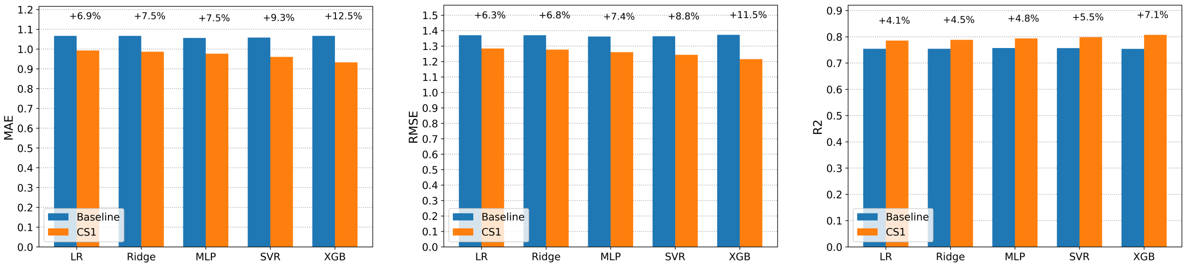
Figure 4.17: Comparison between baseline and combined strategy 1 (CS1) across all estimators for Site 3. The percentage labels above the bars indicate the relative improvement of CS1 compared to the baseline.



(a) OL 5 – MAE

(b) OL 5 – RMSE

(c) OL 5 – R^2



(d) OL 6 – MAE

(e) OL 6 – RMSE

(f) OL 6 – R^2

Figure 4.18: Performance comparison of Baseline and CS1 for Site 4 at OL 5 and OL 6. The percentage labels above the bars indicate the relative improvement of CS1 compared to the baseline.

4.2.1. Baseline

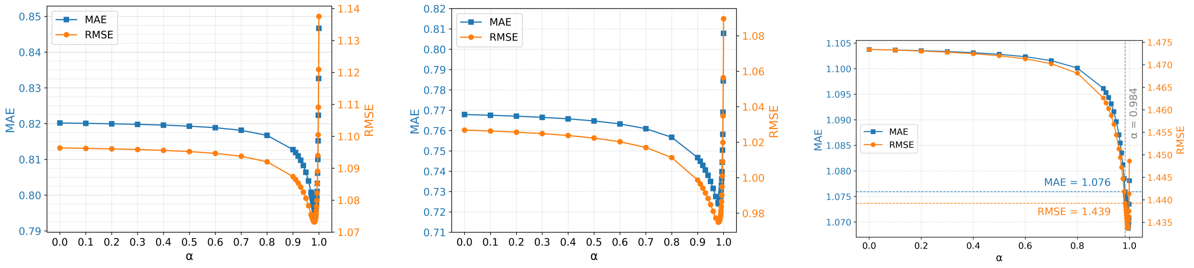
The baseline used to compare the additional methods against is IDW. The performance of IDW on Site 1 and Site 2 can be found in Table 4.4. Note that for Site 1 only the 2 SE masts scenario is given because the 3 SE mast scenario cannot be evaluated. For evaluation at least one mast needs to be withheld, and for Site 1 only 3 OLs are recorded.

Table 4.4: Spatial Extrapolation Baselines Performance for Sites 1 and 2, given by IDW.

Site	SE Masts	MAE	RMSE
1	2	1.1038	1.4734
2	2	0.8202	1.0963
	3	0.7679	1.0268

4.2.2. Using the LES Wind Field Structure

The first step in using the LES wind field structure in spatial extrapolation is to perform hyperparameter tuning, to learn what weight to give to the distance in wind speed or wind direction, compared to the geographical distance. Figures 4.19 and 4.20 shows the MAE and RMSE achieved when using different values for the scaling factor α for both wind speed and wind direction, for the scenario of using 2 SE masts as well as 3 SE masts.



(a) Site 2, 2 SE Masts (b) Site 2, 3 SE masts (c) Site 1, 2 SE masts

Figure 4.19: Performance of WS-IDW, given in MAE and RMSE, for various values of α. For Site 2 both the 2 SE mast scenario (a) and 3 SE mast scenario (b) are shown. For Site 1 only the 2 SE mast scenario is given (c). In (c) the α resulting from hyperparameter tuning on Site 2 is shown, as well as the MAE and RMSE at that α.

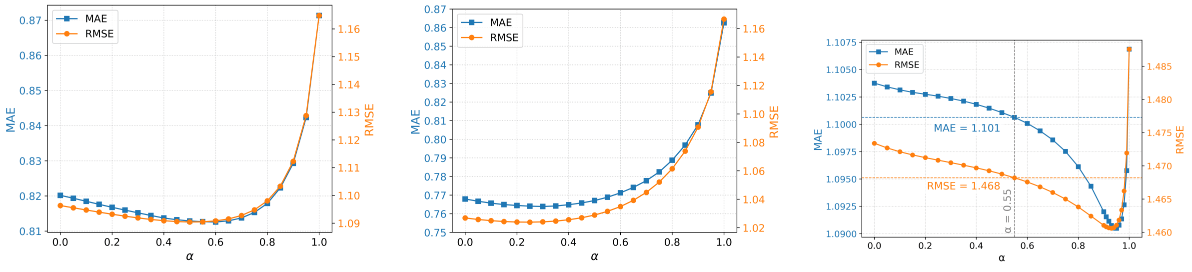
Table 4.5: Summary of WS-IDW Performance Metrics Across Sites and α Values.

Site	SE Masts	α	Comment	MAE	RMSE	MAE Imp. (%)	RMSE Imp. (%)
1	2	0.984	Tuned on Site 2	1.0759	1.4392	2.52	2.32
		0.996	Lowest MAE	1.0688	1.4342	**3.17**	**2.66**
		0.994	Lowest RMSE	1.0692	1.4337	3.13	2.70
		1.000	Wind Speed Only	1.0781	1.4486	2.33	1.68
2	2	0.986	Optimal MAE	0.7957	1.0738	2.99	2.06
		0.982	Optimal RMSE	0.7964	1.0732	2.90	2.11
		0.984	Chosen	0.7959	1.0733	2.96	2.10
	3	0.983	Optimal MAE	0.7240	0.9753	5.72	5.02
		0.980	Optimal RMSE	0.7244	0.9750	5.67	5.05
		0.982	Chosen	0.7240	0.9751	5.71	5.04

Wind Speed

Some MAE and RMSE values, and their respective improvements over the baseline, will be highlighted in following text. These results can also be found in Table 4.5.

Figure 4.19a shows that for incorporating wind speed with 2 SE masts $\alpha = 0.986$ is found to be best when optimizing MAE, while $\alpha = 0.982$ is found to be best when optimizing RMSE. The difference between the MAE and RMSE at both values for α is insignificant. At $\alpha = 0.986$ the MAE has dropped



(a) Site 2, 2 SE masts

(b) Site 2, 3 SE masts

(c) Site 1, 2 SE masts

Figure 4.20: Performance of WD-IDW, given in MAE and RMSE, for various values of α. For Site 2 both the 2 SE mast scenario (a) and 3 SE mast scenario (b) are shown. For Site 1 only the 2 SE mast scenario is given (c). In (c) the α resulting from hyperparameter tuning on Site 2 is shown, as well as the MAE and RMSE at that α.

Table 4.6: Summary of WD-IDW Performance Metrics Across Sites and α Values.

Site	SE Masts	α	Comment	MAE	RMSE	MAE Imp. (%)	RMSE Imp. (%)
1	2	0.950	Optimal MAE	1.0905	1.4611	**1.20**	0.84
		0.930	Optimal RMSE	1.0907	1.4606	1.18	**0.87**
		0.550	Tuned on Site 2	1.1006	1.4682	0.28	0.35
2	2	0.600	Optimal MAE	0.8126	1.0908	**0.92**	0.51
		0.500	Optimal RMSE	0.8129	1.0904	0.89	**0.54**
		0.550	Chosen	0.8127	1.0904	0.91	0.54
	3	0.300	Optimal MAE / Chosen	0.7639	1.0241	**0.53**	0.27
		0.250	Optimal RMSE	0.7640	1.0239	0.51	**0.29**

from 0.8202 to 0.7957 and the RMSE has dropped from 1.0963 to 1.0738. At $\alpha = 0.982$ the MAE is 0.7964 and the RMSE is 1.0731. A middle ground is found by selecting $\alpha = 0.984$. This gives a MAE of 0.7959 and a RMSE of 1.0733, resulting in an improvement of 2.96% for MAE and 2.10% for RMSE.

For the case of incorporating wind speed with 3 SE masts (Figure 4.19b), the optimal value of α is found to be 0.983 when optimizing MAE and 0.980 when optimizing RMSE. The performance difference between these two optima is negligible. Compared to the baseline (normal IDW), which gives a MAE of 0.7679 and a RMSE of 1.0268, the inclusion of the wind speed term improves performance. At $\alpha = 0.983$, the MAE decreases to 0.7240 and the RMSE to 0.9753. At $\alpha = 0.980$, the MAE and RMSE are 0.7244 and 0.9750, respectively. $\alpha = 0.982$ is selected as the value for the hyperparameter, which shows an improvement of 5.71% for MAE and 5.04% for RMSE.

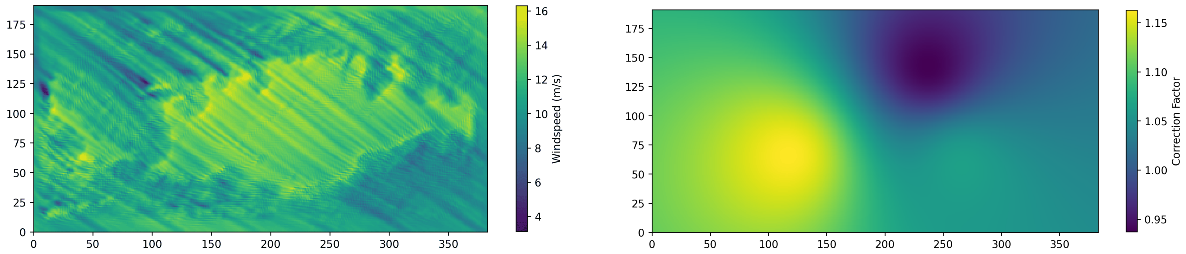
Note that the optimal values for α for the 2 SE mast scenario and the 3 SE mast scenario are very similar. The found values for α are 0.984 for the 2 SE mast scenario and 0.982 for the 3 mast scenario. From the plots in Figure 4.19 it can be seen that these values can be used interchangeably at Site 2 without a notable impact on the resulting MAE and RMSE. This indicates that a value for α found for either scenario generalizes to the other scenario, at least for the same site.

The found value for α, 0.984 for the 2 SE mast case, is subsequently used to perform wind-speed enhanced IDW (WS-IDW) on Site 1 to obtain an unbiased performance evaluation. It is found on Site 1 that using WS-IDW with $\alpha = 0.984$ in the 2 SE masts scenario, gives an improvement of 2.52% for MAE and 2.32%. This is similar to the improvement seen during the hyperparameter tuning on Site 2.

To further investigate generalizability of α across sites, as well as the potential of this method, different values for α are tried for Site 1. This results in the curves that can be seen in Figure 4.19c. These curves show that the lowest MAE would be obtained by $\alpha = 0.996$, going as low as 1.0688, and the lowest RMSE would be obtained by $\alpha = 0.994$, dropping to 1.4337. These values signify a 3.17% improvement in MAE and a 2.70% improvement in RMSE. The improvement percentage for each metric drops only by 0.04% (to 3.13% and 2.66%) when using the α optimal for the other metric. Comparing these improvements with the improvements obtained when using $\alpha = 0.984$ it can be observed that a large part of the maximum improvement is still captured by the $\alpha = 0.984$ found during hyperparameter tuning on the other site.

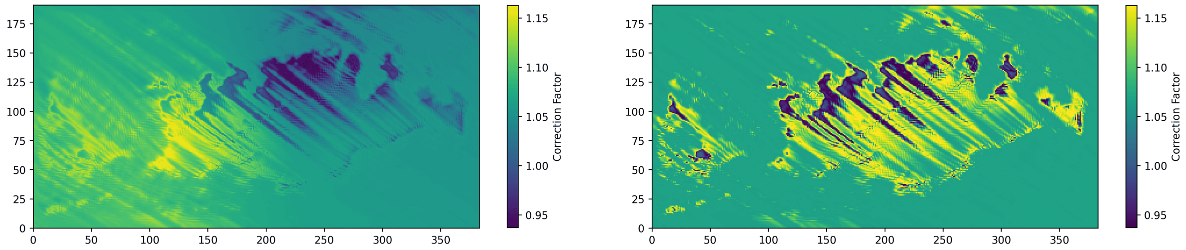
Note also that for Site 1 the MAE and RMSE scores are lower at $\alpha = 1$ than at $\alpha = 0$. Looking back at Equation 3.9 it can be found that when $\alpha = 0$ the difference in wind speed is given no weight and normal IDW is used. Similarly, when $\alpha = 1$ the geographical distance is given no weight and only the difference in wind speed is used. This means that for Site 1, spatial extrapolation only based on the LES wind-speed field is better than IDW based on geographical distance, improving the MAE by 2.33% and RMSE by 1.68%.

To illustrate the working of WS-IDW Figure 4.21 shows the LES wind-speed field at a selected timestep together with the corresponding correction maps obtained using IDW and WS-IDW for three values of α. The IDW map (or WS-IDW with $\alpha = 0$) produces a smooth correction surface determined solely by the geographical distances to the observation locations. When the wind-speed term is included at the optimal found setting ($\alpha = 0.982$), the correction map is seen to follow spatial patterns present in the LES field at that timestep. For $\alpha = 1$, where only the wind-speed term is used, the correction map reflects the instantaneous spatial structure of the LES field even more strongly. These visualizations illustrate how varying the wind-speed weight modifies the spatial structure of the resulting correction map.



(a) Original LES

(b) IDW ($\alpha = 0$)



(c) WS-IDW ($\alpha = 0.982$)

(d) WS-IDW ($\alpha = 1$)

Figure 4.21: Wind-speed field and correction maps for IDW and WS-IDW under varying wind-speed weight α.

Wind Direction

The results for the Wind Direction-enhanced IDW (WD-IDW), where wind direction is weighted by α, are summarized in Table 4.6. The performances found during hyperparameter tuning on Site 2 are also visualized in Figure 4.20 for both the 2 SE mast and the 3 SE mast case.

For incorporating wind direction with 2 SE masts on Site 2, $\alpha = 0.600$ is found to be optimal when minimizing MAE, yielding an improvement of 0.92%. $\alpha = 0.500$ is found to be optimal when minimizing RMSE, resulting in an improvement of 0.54%. A middle ground value of $\alpha = 0.550$ is selected. At this chosen value, the MAE is 0.8127 and the RMSE is 1.0904, which is a 0.91% improvement for MAE and a 0.54% improvement for RMSE over the baseline (MAE: 0.8202, RMSE: 1.0963).

For the case of incorporating wind direction with 3 SE masts on Site 2, the optimal value is found to be $\alpha = 0.300$ when optimizing MAE, and $\alpha = 0.250$ when optimizing RMSE. $\alpha = 0.300$ is selected as the hyperparameter value for this scenario, which yields an MAE of 0.7639 and an RMSE of 1.0241. This shows an improvement of 0.53% for MAE and 0.27% for RMSE over the baseline.

The chosen α value of 0.550 from the Site 2 (2 SE masts) tuning is subsequently used to perform WD-IDW on Site 1 for an unbiased performance evaluation. On Site 1, using WD-IDW with $\alpha = 0.550$ in the 2 SE masts scenario, the improvement is only 0.28% for MAE and 0.35% for RMSE. This improvement is much smaller than the improvement obtained by WS-IDW.

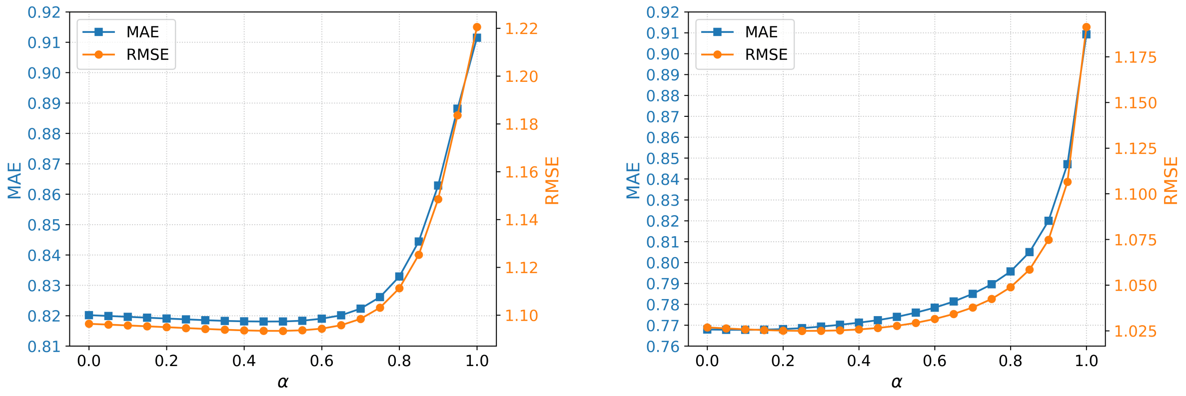
Further investigation of the performance of WD-IDW on Site 1 reveals that the largest possible improvements to MAE and RMSE are 1.20% and 0.87%, and are obtained when using $\alpha = 0.95$ and

$\alpha = 0.93$ respectively. This shows that even when using the optimal value for α the improvement is small.

Note that the optimal values for α differs significantly from Site 1 ($\alpha = 0.93/0.95$) to Site 2 ($\alpha = 0.55$). The same holds for the 2 SE mast ($\alpha = 0.55$) and 3 SE mast scenario ($\alpha = 0.30$) on Site 2. The generalizability of α in WD-IDW thus seems small.

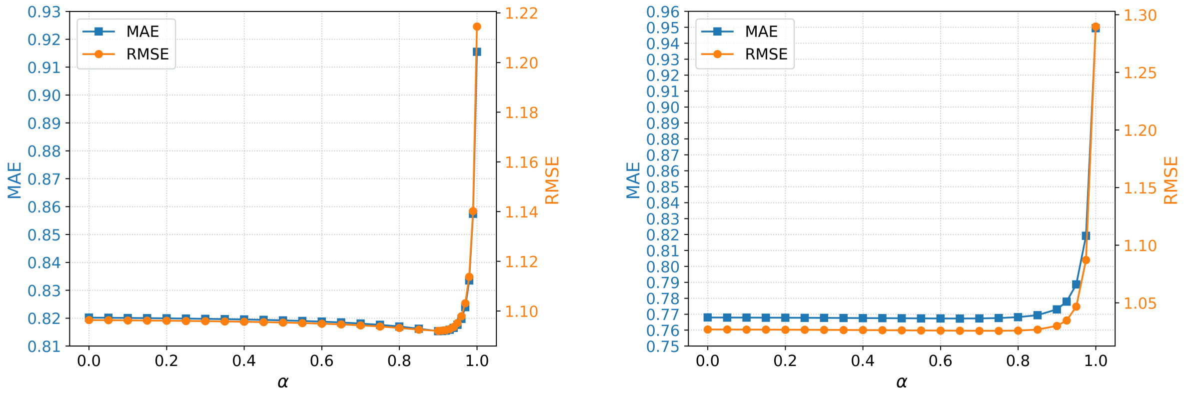
4.2.3. Using Terrain Information

Two terrain features, elevation and tree canopy height, are used to enhance SE and the impact they have on performance is shown in this section.



(a) Elevation, 2 SE masts

(b) Elevation, 3 SE masts



(c) TCH, 2 SE masts

(d) TCH, 3 SE masts

Figure 4.22: Performance of Elevation-enhanced and Tree Canopy Height (TCH)-enhanced IDW, given in MAE and RMSE, for various values of α on Site 2. Subfigures (a) and (b) show the Elevation-enhanced IDW results for the 2 and 3 SE-mast scenarios, while (c) and (d) show the corresponding TCH-enhanced IDW results.

Elevation

The weight given to elevation difference compared to the weight given to the geographical distance, is controlled by α. The performance of the Elevation-enhanced IDW (E-IDW) for different values of α is given in Figures 4.22a and 4.22b for the 2 and 3 SE mast scenarios respectively.

For the 2 SE mast scenario at Site 2 the largest improvements found for MAE and RMSE respectively, are 0.26% at $\alpha = 0.45$ and 0.27% at $\alpha = 0.50$. For the 3 SE mast scenario the largest improvements are found are 0.02% at $\alpha = 0.1$ for MAE, and 0.19% at $\alpha = 0.25$ for RMSE. It is apparent that the inclusion of elevation does not produce any noteworthy improvements. Therefore, further investigation on Site 1 is not performed.

Tree Canopy Height

The performance of Tree Canopy Height-enhanced IDW (TCH-IDW) when using various values for α is shown in Figures 4.22c and 4.22d. These Figures respectively show the 2 and 3 SE mast scenarios for Site 2.

For the 2 SE mast case a small improvement can be seen in both metrics, both reaching their lowest values at $\alpha = 0.9$. This amounts to a 0.59% improvement for MAE and a 0.39% improvement for RMSE. For the 3 SE mast case it can be seen that no noticeable improvements are made. These results indicate that the incorporation of TCH does not hold any promise, and therefore further investigation on Site 1 is not performed.

5

Discussion

This chapter interprets the experimental results presented in Chapter 4, analyzing the effectiveness of the proposed methods. The discussion is organized into two main parts, mirroring the research structure. Section 5.1 evaluates the temporal extrapolation strategies, while Section 5.2 addresses spatial extrapolation, discussing the potential and limitations of incorporating flow and terrain structure into IDW. Throughout the chapter, the physical mechanisms driving the observed improvements are explored, alongside an assessment of the generalizability and operational feasibility of the proposed methods.

5.1. Temporal Extrapolation

This section discusses the findings regarding the temporal correction of LES wind speed predictions at observation locations. The analysis begins by establishing the validity of using microscale LES data as a long-term reference source in MCP, comparing it against the use of mesoscale data. Subsequently, the impact of specific methodological enhancements is examined in isolation, determining how the inclusion of additional input features, spatial and temporal context alters prediction accuracy. The section concludes with an analysis of the combined correction strategy, evaluating the compound effects of these components and the distinct behaviors of linear versus non-linear estimators.

5.1.1. Mesoscale vs. LES as MCP Reference Data

The results reveal that the uncorrected LES time series performs worse than the mesoscale reference on all three metrics. Even though the microscale simulation operates at a higher spatial resolution and therefore exhibits a richer variability, this does not guarantee improved accuracy at a single point location. Visual inspection of the wind-speed fields (Figure 3.3) shows coherent bands of elevated wind speed. Such structures presumably exist in reality as well, yet it is plausible that the LES does not position them with sufficient accuracy to reproduce mast measurements at an exact location. This misalignment leads to a poorer Identity performance and warrants some caution when interpreting the raw LES output at a single point as a long-term reference.

After applying MCP, the performance obtained using mesoscale and LES reference data becomes virtually indistinguishable. The initial discrepancy between the two sources is fully removed by the correction. This demonstrates that the systematic offset in the LES time series is easily corrected and that, once adjusted, LES reaches the same accuracy as mesoscale data. This is a valuable outcome for the present work, because it shows that MCP applied to LES is a valid approach for long-term correction in wind resource assessment. Any improvement obtained by enhancing the prediction model in later chapters can therefore be interpreted as an improvement over a mesoscale-based MCP workflow as well.

The five MCP estimators yield almost identical post-correction performance. When using only the predicted wind speed as input, the best attainable correction appears to be a simple linear transformation. All models, whether linear or nonlinear, converge to this same mapping. This also explains why the mesoscale and LES results coincide after correction, since both are likely aligned through nearly the same linear bias-removal step.

The performance gains obtained by the baseline corrections are small. Across both reference sources, the improvement remains below ten percent for all metrics, and often closer to five. These values provide a clear baseline for evaluating the contributions developed later in this thesis. Any method that achieves another improvement of similar magnitude would already double the effectiveness of MCP relative to these standard approaches. This puts the size of the challenge into perspective: the baseline MCP correction removes only a modest portion of the error, and the remaining discrepancy is substantial.

5.1.2. Additional Input Features

The results in Section 4.1.2 show that both the wind-direction and the time-of-day are valuable additions to the input feature set, while the addition of resolved-turbulence gives only a very small improvement, and day-of-year actively worsens the prediction.

The improvement achieved by including wind-direction can be explained by the fact that with this addition the prediction model is able to better distinguish between situations. When the input feature set consists of only the wind-speed, for any wind speed ws the model will always predict the same corrected wind speed ws_c even though the situations might be completely different and require distinct corrections. The addition of wind-direction enables the model to distinguish between situations better, e.g. between the situation where wind speed is ws and wind is coming from the west and the situation where wind speed is ws and wind is coming from the east, and return distinct corrected wind speeds ws_{c1} and ws_{c2}.

The inclusion of time-of-day in the feature set allows the correction model to account for incorrect diurnal patterns. By including the time-of-day the true diurnal pattern can be learned from the training data as well as at what times the LES model systematically diverts from it. An intuitive example of this might be that the LES is found to systematically models wind speeds that are too low at night.

Similar to how the time-of-day allows for learning the diurnal pattern, the day-of-year was expected to allow for the model to learn the seasonal pattern. As indicated by the results, this is not what actually happens, and there are two likely causes for this. Firstly, and most simply, this could be because there is no seasonal pattern to learn for this specific site. This suspicion is supported by the autocorrelation analysis, especially on monthly averages, which can be found in Figure 3.5 c. Secondly, the day-of-year values encountered in the test set is, by design, out-of-distribution with regards to the training set. As outlined in Section 3.1.1 and 3.1.3 the data spans exactly one year, and is kept in chronological order when splitting it into folds for cross-validation. If the prediction models give the day-of-year any weight in the prediction, the out-of-distribution values of the test set could worsen the performance.

A limitation of this research is that only a limited amount of features are explored. The LES model has a large variety of features that can be recorded during the simulation run, such as temperature and air-pressure, which could further improve the performance. These features have to be recorded during the LES run, otherwise they are lost and can't be recovered later. Since the dataset made available by Whiffle for this project only included these features, only these features were considered.

5.1.3. Additional Area

The overall improvement observed with increasing area size confirms that local spatial context benefits the temporal correction model. This indicates that nearby LES grid cells contain useful flow information and that the LES grid resolution of 100 m is fine enough for local structures to be relevant. However, it is found that the improvement obtained varies notably between locations and models.

The variability between observation locations shows that the improvement obtained from larger input areas is not universal but depends on both location and model type. The differences between models can partly be explained by their ability to handle the quadratic increase in input dimensionality with area size, as well as by how effectively they can learn spatial relationships between neighboring cells. LR performs worst at the largest area size, which is expected given its sensitivity to high dimensionality. XGB also shows limited improvement, likely because its tree-based structure with restricted depth makes it difficult to capture continuous spatial patterns.

Regarding the differences between observation locations, the largest improvements are found for OL 2 and OL 3, while the improvement for OL 1 is smaller. As seen in Figure 3.1c, the local terrain at OL 2 and OL 3 is more complex than at OL 1, suggesting that incorporating LES data from surrounding cells is especially beneficial in regions with more complex local flow and terrain features.

The results from the extended area analysis on Site 1 using Ridge (Figure 4.7) indicate that, at least for this site, the inclusion of a larger area translates directly to more useful information. The performance trend does not plateau even at the largest possible size of 111×111, suggesting that even larger areas, extending beyond the current LES domain, could hold further information and improve performance. This persistent linear improvement implies that the relevant error structures in the LES are not confined to the immediate microscale vicinity but are linked to larger, site-wide flow patterns that the linear model can effectively leverage despite the high dimensionality. The observation that the explicit area input outperforms the CAE encoding (at LD 128) in this specific experiment does not necessarily imply that the CAE captures less information. Rather, it may be explained by Ridge Regression's inability to fully exploit the complex, non-linear information compressed into the latent space. This hypothesis is supported by the fact that other models, such as XGB, were observed to gain significantly more performance from the CAE encoding than Ridge.

While model performance improves with larger input areas, the quadratic growth in input dimensionality quickly becomes computationally expensive, making the use of the entire LES field infeasible for some models. Despite this limitation, the findings confirm that including local spatial information in the model input improves performance, supporting the idea that LES correction benefits from spatial context.

5.1.4. Additional Time Steps

A clear improvement is observed by adding additional time steps to the input of the prediction model, as can be seen in Section 4.1.5. A gap can be seen between the improvement percentages of the linear models (LR and Ridge) and the non-linear models (MLP, SVR and XGB) for all masts and metrics. This gap strongly indicates that there is a non-linear component that the linear models are incapable of learning, while the non-linear models can leverage this to improve performance.

Comparing the non-linear models, MLP is seen to be slightly outperformed by SVR and XGB in most of the cases. Furthermore, the performance of MLP is seen to be less predictable, as indicated by the mean performance increasing and decreasing, or 'jumping around', as time window increases steadily. A possible explanation for this is the random initialization of MLP, which can cause a 'lucky' or 'unlucky' initialization, leading to better or worse performance. SVR on the other hand is deterministic, given the hyperparameters. XGB is non-deterministic, with the randomization coming only from the randomly sub-sampled features and data rows for each tree. This randomization seems to be less impactful than the randomization for MLP.

Analysis of model performance across time windows ranging from 1 to 21 reveals a clear performance plateau beginning around a time window size of 11 (representing 5 preceding and 5 succeeding time steps). This plateau suggests that information further away than 5 hours is generally irrelevant for further prediction improvement. This finding is notable because the preceding autocorrelation analysis indicated a statistically significant correlation (outside the 95% confidence interval) extending up to a 28 hour distance. The performance plateau, therefore, stands in unexpected contrast to the statistical correlation indicated by the time series analysis.

The plateau suggests that the mechanism driving performance gain is not only dependent on temporal autocorrelation. A potential explanation is that the added time steps primarily enhance the model's ability to distinguish different dynamic wind scenarios (e.g., rapidly increasing wind speed, steady peak flow, decreasing wind speed, etc.). The improvement is thus derived from the capacity to find and utilize the trend of the wind speed, enabling the model to apply slightly different corrections for different states. Following this hypothesis, the plateau at an 11-step window size represents the temporal distance beyond which previous or future time steps contain little additional information relevant to defining the immediate, controlling trend at the target time.

A direct limitation of using a time window that uses x preceding and x subsequent time steps is that this model can't be used to correct the LES model for the first and last x time steps of a run. There are two easy solutions for this. First, extending the period for which the LES is run x hours into the past and future, or second, training a second model which doesn't use a time window and using that to correct the first and last x time steps.

5.1.5. Joint Multi-Location Prediction

As mentioned in Section 4.1.6, the joint approach, predicting corrected wind speeds for multiple locations with a single model, overall does not improve over the baseline. Some metrics for some models at some locations slightly improved, while other metrics for other models at other masts slightly worsened.

Overall the performance of both strategies is comparable. This indicates one of two things.

Firstly, it could indicate that the challenge of correcting the LES is completely distinct at each location, and therefore nothing is to be gained from also seeing the training data of other locations. This seems extremely unlikely, and can be debunked by using a model that is trained on one mast location to correct the LES at another mast location, which is found to improve the performance compared to not correcting the LES model at that location. This proves that the corrections required at each location are not completely distinct.

Secondly, the lack of improvement could suggest that the amount of data available for each location is sufficient, and is not a limiting factor. If the amount of data is not a limiting factor, the availability of more data doesn't solve any problem, and thus wouldn't improve the model. This explanation is particularly plausible given that the joint model relied solely on LES wind speed as input. With such a low-dimensional feature space, the correction function is likely simple enough that the single-location datasets are already sufficient for convergence, rendering the additional data from other locations redundant. This strategy could prove more useful when larger, more complex input is used.

5.1.6. Autoencoded Wind-speed Field

Using the encoding of the entire wind-speed field, as given by an CAE, as part of the input to the prediction model gives a significant improvement in all cases. This strongly suggests that the correction for the LES requires information concerning site-wide flow conditions and cannot be optimized using only values at the OL itself. To reiterate, a wind-speed field consists of the wind-speed as given by the LES model at every location.

The effectiveness of the CAE encoding relies on the assumption that similar wind-speed fields should be corrected similarly, and that the encoding provides the prediction model with an condensed vector that can be used to determine similarity. These assumptions seems to hold, as indicated by the increase in performance. In theory, the encodings of similar wind-speed fields will be close together in the latent space.

Comparing the different models a notable difference can be observed between the linear models and the non-linear models. The non-linear models already obtain the largest part of the improvement for a latent dimension as small as 8, and a larger latent dimension only allows for small improvements, while the linear models see a more gradual improvement as the latent dimension increases. This can be explained by the fact that as the latent dimension becomes smaller, the CAE has to learn an increasingly non-linear function to be able to represent the full wind-speed field in the latent space. This is supported by the reconstruction losses found for the different CAEs, which are all similar, indicating that with a smaller LD the model doesn't encode less information, it just encodes it more complexly. The non-linear models are better equipped to exploit this highly compressed, non-linear representation. The linear models, conversely, continue to improve more gradually as the latent dimension increases, because the higher dimensionality reduces the non-linear complexity and provides representations that are more linearly-separable.

Further comparing the non-linear models it can be seen that XGB is best able to leverage this additional input resulting in the largest improvements, especially for large LDs. This indicates that XGB is best able to deal with the high non-linearity and large input feature sets. Furthermore, it can again be observed that MLP is slightly more unpredictable, as indicated by its average performance fluctuating as LD increases, due to its random optimization.

Relation Between Additional Area and CAE Strategies

The comparison between the Additional Area, CAE, and combined strategies provides further insight into how these methods capture spatial information. On average, the CAE with LD = 8 achieves roughly twice the improvement of the explicit 31×31 area input, while the combined approach yields only a marginal additional gain. This would lead to the assumption that the CAE largely subsumes the information provided by the area expansion, making explicit inclusion of neighboring LES cells unnecessary in most cases. However, when examining the results per observation location, a more nuanced picture emerges. For OL 2 and OL 3, where the local terrain is more complex, the combined approach slightly outperforms the CAE alone (by 0.94% and 0.43%, respectively), suggesting that additional local context remains informative. For OL 1, where the local terrain is relatively simple, adding the local area increases dimensionality without contributing new information, leading to a small deterioration in performance. Overall, the CAE alone suffices for capturing the dominant spatial structure, while explicit local area information only adds value at locations with pronounced terrain complexity.

The CAE results, especially when compared to the 31×31 area experiment, indicate that the correction model benefits from information extending beyond the local neighborhood. This reinforces the idea that the relevant spatial context for LES correction spans the entire site, which is consistent with the results found in the Additional Area experiment using area sizes up to 111 × 111 (Figure 4.7). Similarly, it reiterates that the site-wide wind-field structure is key to improving performance. Consequently, because it is found that it is not only the local context that improves performance, a similar approach may also enhance coarser model corrections, such as mesoscale models.

5.1.7. Combining Strategies

Combination Strategy (CS) 1 integrates resolved turbulence, wind direction, and time of day with a time window of 11 and a CAE with latent dimensionality 8. CS 2 uses a latent dimensionality of 128 and is evaluated only for linear models to test whether they can exploit the larger encoding.

Across metrics and locations on Site 1 the gains are clear and consistent. CS 1 improves performance by about 14% on average, with XGB around 16.5% and other estimators close behind (Figures 4.15, 4.16; Tables 4.2, 4.3). The improvement is not driven by a single OL of Site 1, as indicated by the limited spread of the error bars. These gains signal meaningful reductions in MAE and RMSE and a higher R^2 which together signify a significant improvement in the capacity to correct the LES.

Subsequent verification on the unseen Sites 3 and 4 reveals more nuanced performance. On Site 3, the average aggregated improvement is 11.4%, slightly lower than the 13.9% observed on Site 1. XGBoost consistently outpaces the average, achieving a 13.0% improvement on Site 3. For Site 4, the gains are significantly smaller, with the average improvement dropping to 7.2%, roughly half of that seen on Site 1, though XGBoost again demonstrates robustness by securing a 9.8% gain. Crucially, these figures must be interpreted within the broader context of WRA correction standards. As established in Section 4.1.1, baseline MCP methods were found to yield performance gains between 6.4% (mesoscale) and 8.7% (LES). Consequently, even the lower performance observed on Site 4 represents a substantial advance, effectively doubling the error reduction capacity compared to standard MCP workflows.

Lower performance increase for Site 4

The question becomes why the improvements obtained on Site 4 are significantly lower then on Site 1 and Site 3. Looking at the terrain maps of these three site it can be seen that the terrain of Site 1 and Site 3 are very simple, while the terrain of Site 4 is significantly more complex (Figures 3.1, 3.2). Firstly, since CS1 is constructed based on what was shown to be effective on Site 1, a simple terrain site, it is not surprising that it works best for sites with simple terrain. Doing all individual experiments on Site 4, and using those results to inform a combined strategy might have resulted in a different composition more effective for more complex sites and less effective for simple sites. Secondly, the use of a CAE encoding the wind-speed field of the entire site might be less informative for OLs with complex terrain. If the terrain is simple the wind speeds are more homogeneous and correlated, and the range of that correlation is larger before disappearing. Subsequently, the site-wide wind-speed field, or its encoding, are more informative in simple terrain than in complex terrain. This could explain the reduced improvement for Site 4. This is basically a more specific extension of the first point made. A third hypothesis is that the correction task is simply more difficult for Sites with more difficult terrain. The task of correcting the LES can also be stated as the task of predicting the residuals. Given that the LES model is a simplification of the real world, as any model is, this might fit a simple site well, leaving only some predictable bias. On the other hand, for a complex site the difference between the complex and chaotic reality and the simplified model results in chaotic and difficult to predict residuals. This would result in the observed lower improvement for more complex sites, given the same tools (data and method).

Lower performance for Site 3 compared to Site 1

Now a hypothesis will also be offered to (partially) explain the difference in performance increase between Site 1 and Site 3, which are of comparable terrain difficulty. An important reason to note is CS1 is designed using experiments on Site 1, using the same dataset for the isolated experiments, as for the evaluation of the combined strategy, and thus the found performance isn't completely unbiased. This is the foremost reason, making a slight drop as seen for Site 3 expected and acceptable. However, a second possible cause exists. The size of the sites differs and thus the amount of information the CAE can capture in the encoding differs as well. Site 1 has a size of 19.2 by 25.6 km, while Site 3 has a size of 12.8 by 12.8 km. This means Site 3 is only one-third the size of Site 1. Having three times the spatial context encoded by the CAE and included in the input the performance increase might be larger and more similar to the results seen for Site 1. The same argument goes for Site 4, with a size of 12.8 by 16.0 km, thus covering an area roughly 40% of Site 1.

Model dependence
The size of the improvement is model dependent. From the results of Site 1 it is observed that non-linear learners benefit more than the linear models from CS 1, which fits the expectation that these models can better use the temporal context and the highly non-linear CAE encoding. This same pattern is observed in the results of Site 3 and Site 4, supporting this interpretation. Ridge gains further under CS 2 which suggests that it is able to exploit the larger CAE encoding, while controlling the increasing input dimensionality with strong regularization. LR degrades under CS 2 which conversely points to its sensitivity to high dimensionality and its inability to deal with it. These results, obtained on Site 1, underline that the best latent dimensionality is estimator specific, depending on a models ability to regularize and extract non-linear information.

Although the size of the improvement obtained differs from model to model, in all cases an improvement was observed, with the exception of CS2 using LR. As noted in the Section 2.2.1 different studies report different ML models as the top performer for different sites, indicating that this varies from site to site. The present result is therefore valuable, since the combined strategy increases performance across the board, and can thus be applied to whatever ML model is found to be best for a specific site.

Notably, in this research XGB is found to be the top performer for all sites. This is suspectedly because for each model hyperparameter tuning is done with hyperparameter domains chosen after manual experimentation, and the hyperparameter domains chosen are not equally good for each model. This is because there is no direct way to assess the quality of the hyperparameter domains chosen, nor to compare them across model types. In this research this may have favored XGB, while in another paper this might result in another model ending up as the top performer.

Accreditation of the composite parts
The three ingredients of CS 1 interact but are nonetheless complementary. The realized CS 1 improvement on Site 1 is only a few percentage points below the “disjunct” upper bound as given in Table 4.3, which implies limited overlap. Without further experiments that run two-out-of-three combinations it is not possible to localize the remaining overlap with precision. The most plausible interpretation is that, since the time window and the CAE both encode aspects of spatio-temporal structure, this is where the overlap is most prominent. Alternatively the wind-direction might be extracted from the wind-speed field by the CAE, as the overall structure of the wind-speed field strongly indicates this, causing the redundancy of explicitly adding the wind-direction feature. As Figure 4.3 shows that the improvement resulting from the inclusion of wind-direction is roughly 3%, this would fit quantitatively.

Operational Stability and Out-of-Distribution Behavior
A critical limitation observed in the non-linear models is their inability to generalize beyond the training domain. When exposed to out-of-distribution wind speeds, these models exhibit physically implausible behaviors. SVR tends to converge toward the mean and XGB plateaus at the training maximum. The behavior for MLP is ill-defined OOD.

To mitigate this risk without sacrificing the accuracy gains provided by non-linearity, a residual-based modeling architecture is proposed for future implementation. In this setup, a simple Linear Regression model (based solely on reference wind speed) would serve as the baseline predictor. The non-linear models would subsequently be trained to predict only the residuals (the deviation between the measurement and the linear prediction).

This approach ensures that for extreme wind speeds or OOD conditions, the combined system effectively falls back on the robust, physically plausible extrapolation of the linear model. While this architectural change is not expected to negatively impact the performance gains observed in this study, its validation remains a subject for future research.

In a WRA context the improvements observed for this combination strategy matter. The objective is to correct LES fields at sparse and incomplete mast records rather than to predict from scratch. The combined strategy reduces typical error levels across sites and metrics which improves the reliability of long-term corrections and narrows the uncertainty that feeds into energy yield, availability assessments and finance-grade reporting. To give context to the significance of these gains, although empirically, the head of R&D at Whiffle stated at the start of this thesis that improvements on the scale of 10-20% are reason to further investigate and possibly change the methods they use.

Based on these findings CS 1 is recommended as the default temporal correction method for nonlinear learners and LR. Ridge can profit from CS 2.

5.2. Spatial Extrapolation

This section discuses the results, which can be found in Section 4.2, of the spatial extrapolation methods introduced in Section 3.2.

A universal limitation of the spatial extrapolation experiments lies in the method of performance validation. The objective of the spatial extrapolation step is to obtain corrected wind speeds for all locations within the LES domain, yet performance can only be evaluated where ground-truth measurements exist. Moreover, the locations used for extrapolation cannot simultaneously serve for validation. As a result, evaluation is restricted to a limited number of observation locations. For Site 1, this means that in each experiment only one location is available for validation, leading to three validation instances in total when the experiments are repeated three times and a different location is left out each time. For Site 2, validation is based on 14 and 13 masts in the 2 and 3 SE mast scenarios respectively. Consequently, performance assessment relies on a small subset of all spatial points within the domain.

This restriction also raises questions about representativeness. The observation masts are not randomly distributed but are typically placed at locations of high wind activity. The performance metrics derived from these masts may therefore not fully capture the behavior across the entire domain. This limitation is universal across all spatial extrapolation methods, including the baseline IDW method. Given the lack of alternative ground truth, it is assumed that the results obtained from the available masts form a reasonable proxy for the true performance over the full domain and that observed improvements in these validation locations correspond to genuine improvements in the overall correction quality.

A further limitation of the experiments performed is that the sample size used is small. The methods are evaluated on only 2 sites. It can therefore not be stated with certainty that the results on other sites would be similar. Especially sites with different characteristics could yield different results and effectiveness of methods should be further verified on those types of sites, e.g. off-shore or mountainous sites.

5.2.1. Using the LES Wind Field Structure

The LES wind field enhanced IDW methods WS-IDW and WD-IDW investigates whether incorporating the LES wind field, in terms of wind speed (WS) and wind direction (WD), can improve the accuracy of spatial extrapolation beyond the standard distance-based IDW. The results of these methods are given in Section 4.2.2 and will be discussed here.

Wind Speed-enhanced IDW

WS-IDW demonstrates a clear improvement over standard IDW, with improvements of around 2-3% for the 2 SE mast case on both sites and 5-6% for the 3 SE mast case on Site 2. While the 2-3% improvement is modest, the 5-6% improvement is more significant and suggests that incorporating LES wind speed similarity adds meaningful structure to the spatial extrapolation.

The optimal α values for WS-IDW are high (> 0.98), indicating that the wind speed differences between two locations, which are often only a few meters per second, need to be scaled significantly relative to geographic distance (hundreds to thousands of meters) to noticeably influence the extrapolation. Additionally, on Site 1, the performance at $\alpha = 1$ (pure wind speed weighting) surpasses that at $\alpha = 0$ (pure geographic distance), strongly indicating that the LES wind speed field provides strong spatial information. This finding supports the hypothesis underpinning this method, which is that locations with similar wind speeds require similar corrections.

Spatial Flow Structure

Visual inspection of the correction maps provides additional insight into why WS-IDW yields improved performance. The maps show that, as intended, once the wind-speed weighting is introduced, the spatial structure of the correction begins to follow the flow structure present in the LES field at that timestep. One consequence is that the correction map often inherits streak-like patterns that closely resemble those in the LES field, as can be seen in Figure 4.21. A closer look shows that these streaks in the correction map tend to counteract the streaks present in the original LES, effectively reducing their amplitude.

LES is known to reproduce realistic flow structures, but the appearance of these streaks in the correction map, combined with the observed improvement over standard IDW, supports the suspicion that LES may not always localize these structures precisely at the correct spatial positions. Their partial removal therefore appears to contribute to the improvement achieved by WS-IDW. A similar suspicion arose in the temporal extrapolation experiments, where it was found that the mesoscale-based MCP adhered more closely to the observations than the LES-based MCP, suggesting that fine-scale features in LES can be slightly misplaced in both space and time. When such hyper-local structures are displaced, a purely distance-based method such as IDW cannot adjust for these errors. WS-IDW, however, uses local wind-speed similarity to guide the weighting process, and these streaks are precisely regions regions where coherent wind-speed patterns make the weighting particularly influential.

It is theoretically possible for WS-IDW to have the opposite effect by amplifying the streaks rather than reducing them; for example by further accelerating already high wind-speed regions or further slowing down already low wind-speed regions. However, manual inspection of many wind-speed fields and their corresponding correction maps shows that such amplification is uncommon in practice. This can be explained by the fact that modeled high wind speeds at the observation locations are more often overpredicted than underpredicted, which results in smaller correction factors, while modeled low wind speeds tend to be underpredicted and therefore receive larger correction factors. Note that these two things are relative to each other: a speed up of 1.1 of high wind speed can be a low correction factor if low wind speeds are corrected with a factor of 1.3.

Generalizability

The consistency of the optimal α values across the 2 and 3 SE mast configurations (0.982 vs. 0.984) for Site 2 suggests that α generalizes well across different scenarios on the same site. Furthermore, the successful transfer of the Site 2-tuned $\alpha = 0.984$ to Site 1, which captured a large part of the potential improvement, demonstrates the method's reasonable generalizability across different sites. This indicates that WS-IDW is not overly sensitive to local site-specific conditions. However, these results only test generalizability in each direction (cross-site and across SE mast scenarios) once, and thus are far from conclusive.

Continuing from the indication that α generalizes in both directions for WS-IDW, this method can be useful in both the 2 SE mast and the 3 SE mast scenario for a new unseen real world site. In the 2 SE mast scenario the value for α is used which is learned on Site 2, meaning $\alpha = 0.984$. For the 3 SE mast scenario two approaches are possible. First, similar to the 2 SE mast case, the value learned on Site 2 can be used, which is $\alpha = 0.982$ for the 3 SE mast case. Secondly, since there are 3 masts (or OL) two can be used for SE while the third is used for validation. With this setup α can be tuned specifically for the new site. The first approach relies on generalizability across sites, while the second approach relies on generalizability across number of SE masts. The results show the second case to work better, so that is the recommended method.

A larger improvement is found for the 3 SE mast scenario than for the 2 SE mast case. It is expected that the improvement obtained by WS-IDW becomes larger as more SE masts are used, because with more masts more regions in the LES wind-speed field will be covered, and a larger variety in observed wind speeds can be expected. Since during SE the weight given to each mast is normalized (see Equation 3.8) so that they sum to 1, they need to be thought of as relative to each other. Therefore, when the LES wind speeds at all masts are very similar, the difference in wind speed to any point differs very little between masts, and thus their weights in IDW are virtually unaffected. From this it follows that variety in LES wind speeds across masts is necessary for WS-IDW to be able to exploit this and differ from normal IDW. This leads to the expectation that the improvements will be more pronounced when more masts are available. It is therefore expected that if it were possible to validate the 3 SE mast case on Site 1 an improvement of ca. 5-6% would be found, similar to the improvement found for Site 2.

Despite the improvements, the method's performance remains modest. The obtainment of the improvement is also subject to finding a good value for α, and setting this hyperparameter incorrectly has the potential to actually deteriorate the performance for some sites. Additionally, the method depends on the accuracy of spatial structure of the LES wind-speed field, and inaccuracies in the LES could lead to biased extrapolations.

Wind Direction-enhanced IDW

The Wind Direction-enhanced IDW (WD-IDW) provides only minor improvements over the standard IDW method, typically below one percent. While some improvement can be seen across all tested experiments, the results indicate that wind direction carries limited useful spatial information for reducing extrapolation errors.

As seen in Figure 3.12 the wind direction is largely coherent across the site during high-wind conditions, leaving little spatial variation for the method to exploit. On the other hand, at lower wind speeds, variation in wind direction increases, enabling WD-IDW to exploit this variance. The modest gains therefore likely arise from isolated, low-wind instances where local wind direction patterns deviate meaningfully between observation locations. These low-wind instances contribute less to the overall error because the values, and their errors, are smaller in absolute terms. Furthermore, the low-wind instances are of lower importance during WRA because with low wind speeds there is little wind energy to be harvested, limiting the usefulness of WD-IDW for wind resource assessment.

The optimal α values differ considerably between sites and the 2 and 3 SE-mast scenarios, indicating poor generalizability. This suggests that the relationship between spatial wind direction patterns and the required correction is both site-specific and dependent on the number of SE masts. The lack of a consistent α value makes it difficult to select an appropriate weighting for unseen sites, reducing the method's practical applicability. Overall, WD-IDW can slightly refine spatial interpolation but its effect is minor and inconsistent, making it unsuitable as a stand-alone enhancement.

5.2.2. Using Terrain Features

Both Elevation-enhanced IDW (E-IDW) and Tree Canopy Height-enhanced IDW (TCH-IDW) fail to produce meaningful improvements over the standard distance-based IDW, with observed differences being practically insignificant. The results suggest that the terrain features elevation and tree canopy height have a weak or negligible relationship with the spatial variation of the LES errors. It is likely that the influence of terrain is already captured correctly by the LES model, leaving little additional structure for the correction method to exploit.

The flat performance for TCH-IDW across a large range of the α values can be explained by the much smaller spatial scale of tree canopy height variations, which are typically only a few to several tens of meters, compared to the distances between masts, making its contribution small for many distances unless α approaches 1. For E-IDW the elevation difference can become much larger, causing the performance to degrade already for smaller alpha values.

A further explanation for the weak results may lie in the homogeneity of the terrain features at the mast locations. Since extrapolation and validation can only be performed at mast sites, the evaluation effectively captures performance only for these limited and non-random points. The masts are generally positioned in similar high-elevation areas and likely in open terrain without canopy cover to facilitate installation. This lack of variation among observation locations reduces the relevance of both terrain-based features, thereby limiting the potential of these methods to enhance the spatial extrapolation process.

6

Conclusions

This thesis investigated how data-driven methods can improve post-processing of microscale LES in Wind Resource Assessment (WRA) in both the temporal and the spatial direction. First the existing standard MCP models for temporal extrapolation were validated for microscale LES model data by comparing it to the use of mesoscale data. The baseline results show that although raw LES time series initially perform worse than the mesoscale reference, MCP corrections remove this difference entirely and both reference sources yield the same post-correction accuracy. This validates the use of LES as a long-term reference in MCP-style temporal extrapolation. Building on this baseline, the combined temporal strategy introduced in this thesis, which integrates additional LES features, a time window and a compact CAE encoding of the wind-speed field, improves performance on all three sites. The magnitude of the improvement varies with site complexity, averaging 7.2% on the more complex terrain of Site 4 and 11.4% on the simpler terrain of Site 3. These sites were unseen during the design of the combined strategy and thus provide an unbiased evaluation. For spatial extrapolation a flow-aware IDW method that leverages LES wind-speed similarity improves over distance-only IDW by up to 5.7% in MAE and RMSE. Taken together these experiments show that using the spatio-temporal structure present in LES wind fields leads to systematic improvements in both temporal and spatial correction tasks.

Temporal Extrapolation

Before extending the existing methods, the suitability of LES as a reference source for MCP-style temporal extrapolation was assessed. Standard MCP models were applied using both mesoscale and LES time series as the long-term reference. The uncorrected LES data showed a higher error and lower correlation with the observations than the mesoscale reference. After MCP correction this difference disappeared completely, and both reference sources yielded indistinguishable performance across all models and metrics. This confirms that baseline MCP can fully correct the discrepancy between the LES and mesoscale performance relative to the observations, and that LES as a reference source is on par with mesoscale models for the temporal extrapolation task considered in this thesis. These results also establish the scale of improvement that baseline MCP can deliver, with average gains of 6.4% for the mesoscale data and 8.7% for the LES data.

Building on this validation, the combined strategy for temporal extrapolation integrates the additional LES features resolved-turbulence, wind-direction, and time-of-day, with a time window of size 11 and a CAE wind-speed field encoding (latent dimension 8) improves performance by 14% on Site 1, 11.4% on Site 3, and 7.2% on the more complex Site 4. Gains hold for all baseline model types (LR, Ridge, MLP, SVR, XGB), each Observation Location, and all metrics, indicating robustness. The combined strategy also outperforms each component in isolation, confirming the strategies are complementary. The time window captures short-term flow dynamics, while the CAE supplies site-wide spatial context. Together they provide essential spatio-temporal flow information. Improvements are estimator-specific: non-linear models benefit most, with XGB showing the largest average gain. A second combined strategy with a larger latent dimension (128) for the wind-speed field encoding further improves Ridge Regression yet degrades Linear Regression, indicating that the optimal latent dimensionality is estimator-dependent and should match the model's regularization strength and capacity to exploit non-linear structure.

Using the feature day-of-year, adding data from the surrounding LES grid-cells in addition to the CAE encoding, and adopting a joint multi-mast approach did not improve performance and thus were excluded from the combined strategy.

Spatial Extrapolation

The distance-only IDW was replaced with an enhanced flow-aware version, called WS-IDW, that combines geographical distance and LES-based wind-speed similarity, weighted by a single hyperparameter α. This yields consistent, but modest gains in both MAE and RMSE. The largest improvement potential, 5.7%, appears at Site 2 in the 3-mast SE case. In 2-mast SE at Site 1 and Site 2 we see 2-3% error reduction when compared to IDW. It is found that the hyperparameter α generalizes well across sites and across different numbers of observation locations used, which is crucial for real-world usability of this method. The optimal values for α are found to be in the range ≈0.980-0.996. WS-IDW obtains larger improvement increases when more OLs are available and when OLs sample different flow conditions, because the similarity term can re-weight contributions based on flow differences. The informativeness of the LES wind-speed field is supported by the results found at Site 1, where wind-speed-only spatial extrapolation actually beats distance-only. It is recommended to adopt WS-IDW with $\alpha \approx 0.98 - 0.996$ as a robust default and adjust only if validation shows systematic bias, in which case a lower alpha gives a more conservative spatial extrapolation.

Similar to the implementation of WS-IDW the addition of wind-direction and the terrain features elevation and tree canopy height to IDW have also been explored. These additions were found to hold less potential than the wind speed.

The following are concrete recommendations for practice. Use the combined temporal strategy with the stated additional features, a time window of size 11, and a CAE encoding with LD = 8, provided that sufficient observation data are available for training. For Ridge regression, use LD = 128. On complex sites, a larger LD than 8 may be needed to properly encode spatial context. In this case identify the smallest LD that achieves stable encoding through validation and use that. Perform hyperparameter tuning as done in this thesis, using the same search space, data splits, and early stopping to obtain comparable results, in which case XGB is expected to give the best performance. For spatial mapping, adopt WS-IDW when three observation locations are available. Tune α on the 2-OL subset to obtain a good value, then apply it to the 3-OL case, where improvements are typically larger. In 2-OL settings, adopt WS-IDW only if a 2-3% error reduction is meaningful and set α toward the lower end of the 0.98-0.996 range to avoid performance degradation. Because WS-IDW was tested only on sites of low to moderate complexity, assess terrain similarity and validate before deployment on markedly different sites.

Limitations

The generalization of the results to other sites is verified for the temporal extrapolation, as the methods are tested on three distinct sites, but could be further investigated and made more robust by the addition of more sites. The claim to generalization of the spatial extrapolation is also limited as only two sites have been used. Furthermore, observation locations were not randomly selected and likely favor windier or more accessible locations, so found performance may be biased. Another limitation of this thesis is that the experiments only cover one height, which leaves vertical generalization untested. Lastly, the solution developed for temporal extrapolation has only been tested on sites with simple to moderate terrain, and performance on markedly different terrain types, such as complex mountainous areas or flat offshore conditions, remains unverified.

Future Work

Future work should test reliability and generalizability by repeating the experiments across a broader range of sites, terrain complexities and climate types. The current evaluation covers simple to moderate terrain, yet markedly different environments such as complex mountainous regions, offshore settings or sites with seasonal flow regimes may require different modelling choices. The temporal dimension should also be extended by validating the methods on multi-year LES datasets, ensuring that the performance holds under larger separations between training and validation periods and under interannual variability. Vertical generalization is another open direction, since the present work evaluates only the 100 m level; extending the correction methods across multiple heights would allow assessment of vertical wind shear and height-dependent flow behavior.

The targets should extend beyond wind speed to wind direction and other relevant flow variables so the same correction framework can be assessed on a wider set of atmospheric quantities. An end-to-end evaluation of the full correction pipeline is also needed, in which the temporal correction is applied first and subsequently combined with the spatial extrapolation, followed by a comparison to the MCP and IDW baselines.

At the strategy level the feature space can be expanded with additional LES-derived predictors that add information without excessive dimensionality. Auto-encoding also deserves deeper study. Variational CAEs may yield encodings that downstream models use more effectively, and an encoder-only approach trained on the correction targets, with the encoding added to the inputs of an MLP, would give a compact CNN-MLP pipeline capable of learning the useful features directly. The encoder inputs can also be extended to include wind direction or other relevant fields. Finally, one-at-a-time ablations that remove a single component will quantify each contribution within the combined strategy and clarify how the components interact.

A broader challenge is to design models that handle out-of-distribution conditions more robustly. Starting the non-linear models from the linear relation identified by the LR baseline and letting them learn only the residuals may also improve robustness to out-of-distribution flow conditions, although this requires systematic evaluation. A related future direction is to repeat the design process of the combined strategy on representative sites of different terrain classes. The current strategy was derived from a simple-terrain site and performs best on similar terrain. Repeating the ablation and design steps on moderate, complex or offshore sites would likely yield terrain-specific combined strategies that better reflect flow behavior in those environments.

Finally, the use of time-windowed inputs for the temporal extrapolation can similarly be applied when using mesoscale data, reanalysis data or even meteorological measurements used as a reference source in MCP. While the inclusion of temporal context was demonstrated effective for LES data in this thesis, its application to these more common data sources remains unexplored. Future research should therefore evaluate whether enriching standard MCP workflows with multi-step time windows can unlock similar performance gains.

Bibliography

[1] Rebecca J. Barthelmie and Sara C. Pryor. “Climate Change Mitigation Potential of Wind Energy”. In: *Climate* 9.9 (Aug. 2021). DOI: `10.3390/cli9090136`.

[2] J. C. Y. Lee and M. J. Fields. “An overview of wind-energy-production prediction bias, losses, and uncertainties”. In: *Wind Energy Science* 6.2 (2021), pp. 311–365. DOI: `10.5194/wes-6-311-2021`. URL: `https://wes.copernicus.org/articles/6/311/2021/`.

[3] K.S.R. Murthy and O.P. Rahi. “A comprehensive review of wind resource assessment”. In: *Renewable and Sustainable Energy Reviews* 72 (2017), pp. 1320–1342. ISSN: 1364-0321. DOI: `https://doi.org/10.1016/j.rser.2016.10.038`. URL: `https://www.sciencedirect.com/science/article/pii/S1364032116306918`.

[4] G. Gualtieri. “Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review”. In: *Renewable and Sustainable Energy Reviews* 167 (2022), p. 112741. ISSN: 1364-0321. DOI: `https://doi.org/10.1016/j.rser.2022.112741`. URL: `https://www.sciencedirect.com/science/article/pii/S1364032122006293`.

[5] Evert Wiegant and Remco Verzijlbergh. *GRASP model description & validation report*. Whiffle Weather Finecasting Ltd. 2019. URL: `https://www.whiffle.nl`.

[6] Freddy Houndekindo and Taha B.M.J. Ouarda. “Machine learning and statistical approaches for wind speed estimation at partially sampled and unsampled locations; review and open questions”. In: *Energy Conversion and Management* 327 (2025), p. 119555. ISSN: 0196-8904. DOI: `https://doi.org/10.1016/j.enconman.2025.119555`. URL: `https://www.sciencedirect.com/science/article/pii/S0196890425000780`.

[7] Sandra Schwegmann et al. “Enabling Virtual Met Masts for wind energy applications through machine learning-methods”. In: *Energy and AI* 11 (2023), p. 100209.

[8] Ørsted. *Hornsea 2 Offshore Wind Farm*. Accessed: 2025-06-02. 2022. URL: `https://orsted.co.uk/energy-solutions/offshore-wind/our-wind-farms/hornsea2`.

[9] Yanuar Eka Saputra. *Gansu Wind Farm, China Overview*. Accessed: 2025-06-02. 2023. URL: `https://www.exaputra.com/2023/04/gansu-wind-farm-china-overview.html`.

[10] A. Basse et al. “Seasonal effects in the long-term correction of short-term wind measurements using reanalysis data”. In: *Wind Energy Science* 6.6 (2021), pp. 1473–1490. DOI: `10.5194/wes-6-1473-2021`. URL: `https://wes.copernicus.org/articles/6/1473/2021/`.

[11] A. Rangaraj et al. “Validation and bias correction techniques to improve Numerical Weather Prediction wind speed data”. In: Sept. 2019.

[12] Shaopeng Li et al. “A novel frequency-domain physics-informed neural network for accurate prediction of 3D spatio-temporal wind fields in wind turbine applications”. In: *Applied Energy* 386 (2025), p. 125526. ISSN: 0306-2619. DOI: `https://doi.org/10.1016/j.apenergy.2025.125526`. URL: `https://www.sciencedirect.com/science/article/pii/S0306261925002569`.

[13] Hao Zhou, Qi Luo, and Ling Yuan. “Downscaling and wind resource assessment of climatic wind speed data based on deep learning: a case study of the Tengger Desert wind farm”. In: *Atmosphere* 15.3 (2024), p. 271.

[14] José A. Carta et al. “Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study”. In: *Applied Energy* 158 (2015), pp. 490–507. ISSN: 0306-2619. DOI: `https://doi.org/10.1016/j.apenergy.2015.08.102`. URL: `https://www.sciencedirect.com/science/article/pii/S0306261915010387`.

[15] Ravinesh Deo et al. “Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for windspeed prediction of target site using a limited set of neighboring reference station data”. In: *Renewable Energy* 116 (Sept. 2017). DOI: `10.1016/j.renene.2017.09.078`.

[16] Michael Denis Mifsud, Tonio Sant, and Robert Nicholas Farrugia. “Analysing uncertainties in offshore wind farm power output using measure–correlate–predict methodologies”. In: *Wind Energy Science* 5.2 (2020), pp. 601–621.

[17] Jonas Deksnys. “Wind Resource Variation Between Mesoscale and Downscaled Linear Wind Model – Comparison of Surface Roughness and Proximity to Baltic Coast”. Accessed 2025-06-10. Dissertation. Uppsala University, 2022. URL: `https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-477947`.

[18] Murilo Montanini Breve et al. “Computational performance analysis of pca enhanced anen method”. In: *International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability*. Springer. 2024, pp. 18–33.

[19] S. K. Angenent. “Investigating the Accuracy of the Measure–Correlate–Predict Procedure”. Supervisor: S. J. Watson. Master Thesis. Faculty of Electrical Engineering, Mathematics and Computer Science: Delft University of Technology, 2023.

[20] Santiago Díaz, José A. Carta, and José M. Matías. “Performance assessment of five MCP models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques”. In: *Applied Energy* 209 (2018), pp. 455–477. ISSN: 0306-2619. DOI: `https://doi.org/10.1016/j.apenergy.2017.11.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0306261917315866`.

[21] Fanny Kristianti et al. “Combining weather station data and short-term lidar deployment to estimate wind energy potential with machine learning: a case study from the swiss alps”. In: *Boundary-Layer Meteorology* 188.1 (2023), pp. 185–208.

[22] Daeyoung Kim et al. “A comparison of ground-based LiDAR and met mast wind measurements for wind resource assessment over various terrain conditions”. In: *Journal of Wind Engineering and Industrial Aerodynamics* 158 (Nov. 2016). DOI: `10.1016/j.jweia.2016.09.011`.

[23] Michael D Mifsud, Tonio Sant, and Robert N Farrugia. “A comparison of Measure-Correlate-Predict Methodologies using LiDAR as a candidate site measurement device for the Mediterranean Island of Malta”. In: *Renewable energy* 127 (2018), pp. 947–959.

[24] Yu Li et al. “Numerical weather prediction correction strategy for short-term wind power forecasting based on bidirectional gated recurrent unit and XGBoost”. In: *Frontiers in Energy Research* 9 (2022), p. 836144.

[25] Mirella Lima Saraiva Araujo et al. “Wind power forecasting in a semi-arid region based on machine learning error correction”. In: *Wind* 3.4 (2023), pp. 496–512.

[26] Stefan Jonas et al. “Bias correction of wind power forecasts with SCADA data and continuous learning”. In: *Journal of Physics: Conference Series*. Vol. 2767. 9. IOP Publishing. 2024, p. 092061.

[27] Ravi Pandit et al. “SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends”. In: *Wind Engineering* 47.2 (2023), pp. 422–441.

[28] Qiaomu Zhu et al. “Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction”. In: *IEEE Transactions on Sustainable Energy* 11.1 (2019), pp. 509–523.

[29] Mao Yang, Yunfeng Guo, and Yutong Huang. “Wind power ultra-short-term prediction method based on NWP wind speed correction and double clustering division of transitional weather process”. In: *Energy* 282 (2023), p. 128947.

[30] Gerard Ayuso-Virgili et al. “Measure-correlate-predict methods to improve the assessment of wind and wave energy availability at a semi-exposed coastal area”. In: *Energy* 309 (2024), p. 132904.

[31] Xueliang Zhao et al. “A comprehensive wind speed forecast correction strategy with an artificial intelligence algorithm”. In: *Frontiers in Environmental Science* 10 (2022), p. 1034536.

[32] Keran Chen et al. “A model output deep learning method for grid temperature forecasts in Tianjin area”. In: *Applied Sciences* 10.17 (2020), p. 5808.

[33] Laura Clemente-Harding. *A Beginners Introduction to the Analog Ensemble Technique*. Warner Internship for Scientific Enrichment (WISE) Teaching Material, Version 1.0. Accessed 2025-06-10. 2017.

[34] H. Hersbach et al. "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), pp. 1999–2049. DOI: `https://doi.org/10.1002/qj.3803`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803`. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803`.

[35] R. Gelaro et al. "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)". In: *Journal of Climate* 30.14 (2017), pp. 5419–5454. DOI: `10.1175/JCLI-D-16-0758.1`. URL: `https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0758.1.xml`.

[36] Tiejun Zhang et al. "Bias-correction method for wind-speed forecasting; Bias-correction method for wind-speed forecasting". In: *Meteorologische Zeitschrift* 28.4 (2019), pp. 293–304.

[37] C. Bollmeyer et al. "Towards a high-resolution regional reanalysis for the European CORDEX domain". In: *Quarterly Journal of the Royal Meteorological Society* 141.686 (2015), pp. 1–15. DOI: `https://doi.org/10.1002/qj.2486`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2486`. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2486`.

[38] Bert Blocken. "50 years of Computational Wind Engineering: Past, present and future". In: *Journal of Wind Engineering and Industrial Aerodynamics* 129 (2014), pp. 69–102. ISSN: 0167-6105. DOI: `https://doi.org/10.1016/j.jweia.2014.03.008`. URL: `https://www.sciencedirect.com/science/article/pii/S016761051400052X`.

[39] Matthew Churchfield et al. "A Large-Eddy Simulation of Wind-Plant Aerodynamics". In: Jan. 2012. ISBN: 978-1-60086-936-5. DOI: `10.2514/6.2012-537`.

[40] Björn Maronga et al. "Overview of the PALM model system 6.0". In: (Jan. 2020).

[41] T. et al. Heus. "Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications". In: *Geoscientific Model Development* 3.2 (2010), pp. 415–444.

[42] J. et al. Schalkwijk. "High-performance simulations of turbulent clouds on a desktop PC". In: *Bulletin of the American Meteorological Society* 93.3 (2012), pp. 307–314.

[43] R. J. A. M. Stevens and C. Meneveau. "A concurrent precursor inflow method for Large Eddy Simulations and applications to finite length wind farms". In: *Renewable Energy* 68 (2014), pp. 46–50.

[44] Palmer Cosslett Putnam. *Power from the Wind*. New York: D. Van Nostrand Company, 1948.

[45] Ross B. Corotis, Arden B. Sigl, and Michael P. Cohen. "Variance Analysis of Wind Characteristics for Energy Conversion". In: *Journal of Applied Meteorology and Climatology* 16.11 (1977), pp. 1149–1157. DOI: `10.1175/1520-0450(1977)016<1149:VAOWCF>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/apme/16/11/1520-0450_1977_016_1149_vaowcf_2_0_co_2.xml`.

[46] John L. Walmsley and David L. Bagg and. "A method of correlating wind data between two stations with application to the Alberta oil sands*". In: *Atmosphere-Ocean* 16.4 (1978), pp. 333–347. DOI: `10.1080/07055900.1978.9649041`. eprint: `https://doi.org/10.1080/07055900.1978.9649041`. URL: `https://doi.org/10.1080/07055900.1978.9649041`.

[47] L Landberg and NG Mortensen. "A comparison of physical and statistical methods for estimating the wind resource at a site". In: *15th British Wind Energy Association Conference*. Mechanical Engineering Publications Limited. 1994, pp. 119–125.

[48] José A Carta, Sergio Velázquez, and Pedro Cabrera. "A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site". In: *Renewable and Sustainable Energy Reviews* 27 (2013), pp. 362–400.

[49] Anthony L Rogers, John W Rogers, and James F Manwell. "Comparison of the performance of four measure–correlate–predict algorithms". In: *Journal of wind engineering and industrial aerodynamics* 93.3 (2005), pp. 243–264.

[50] M. L. Thøgersen et al. *An Introduction to the MCP Facilities in WindPRO*. EMD International A/S. Aalborg, Denmark, 2010. URL: `https://help.emd.dk/knowledgebase/content/ReferenceManual/MCP.pdf` (visited on 06/03/2025).

[51] Ciaran King and Brian Hurley. “The SpeedSort, DynaSort and Scatter Wind Correlation Methods”. In: *Wind Engineering* 29.3 (2005), pp. 217–241. DOI: `10.1260/030952405774354868`. eprint: `https://doi.org/10.1260/030952405774354868`. URL: `https://doi.org/10.1260/030952405774354868`.

[52] J. F. D. Addison et al. “A neural network version of the measure correlate predict algorithm for estimating wind energy yield.” In: (Jan. 2000). URL: `https://repository.lincoln.ac.uk/articles/conference_contribution/A_neural_network_version_of_the_measure_correlate_predict_algorithm_for_estimating_wind_energy_yield_/25110713`.

[53] Sergio Velázquez, José A. Carta, and J.M. Matías. “Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site: A case study in the Canary Islands”. In: *Applied Energy* 88.11 (2011), pp. 3869–3881. ISSN: 0306-2619. DOI: `https://doi.org/10.1016/j.apenergy.2011.05.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0306261911002984`.

[54] Santiago Díaz, José A. Carta, and José M. Matías. “Comparison of several measure-correlate-predict models using support vector regression techniques to estimate wind power densities. A case study”. In: *Energy Conversion and Management* 140 (2017), pp. 334–354. ISSN: 0196-8904. DOI: `https://doi.org/10.1016/j.enconman.2017.02.064`. URL: `https://www.sciencedirect.com/science/article/pii/S0196890417301760`.

[55] Serkan Kartal, Sukanta Basu, and Simon J Watson. “A decision tree-based measure-correlate-predict approach for peak wind gust estimation from a global reanalysis dataset”. In: *Wind Energy Science Discussions* 2023 (2023), pp. 1–25.

[56] B. Postema et al. “Estimating long-term annual energy production from shorter-time-series data: methods and verification with a 10-year large-eddy simulation of a large offshore wind farm”. In: *Wind Energy Science* 10.7 (2025), pp. 1471–1484. DOI: `10.5194/wes-10-1471-2025`. URL: `https://wes.copernicus.org/articles/10/1471/2025/`.

[57] Christopher Jung and Dirk Schindler. “Introducing a new approach for wind energy potential assessment under climate change at the wind turbine scale”. In: *Energy conversion and management* 225 (2020), p. 113425.

[58] Conor P Sweeney, Peter Lynch, and Paul Nolan. “Reducing errors of wind speed forecasts by an optimal combination of post-processing methods”. In: *Meteorological Applications* 20.1 (2013), pp. 32–40.

[59] Xindong Peng, Yuzhang Che, and Jun Chang. “A novel approach to improve numerical weather prediction skills by using anomaly integration and historical data”. In: *Journal of Geophysical Research: Atmospheres* 118.16 (2013), pp. 8814–8826.

[60] Haochen Li et al. “A model output machine learning method for grid temperature forecasts in the Beijing area”. In: *Advances in Atmospheric Sciences* 36 (2019), pp. 1156–1170.

[61] Luca Delle Monache et al. “Probabilistic Weather Prediction with an Analog Ensemble”. In: *Monthly Weather Review* 141.10 (2013), pp. 3498–3516. DOI: `10.1175/MWR-D-12-00281.1`. URL: `https://journals.ametsoc.org/view/journals/mwre/141/10/mwr-d-12-00281.1.xml`.

[62] Weiming Hu, Davide Vento, and Shiquan Su. “Parallel Analog Ensemble - The Power of Weather Analogs”. In: May 2020.

[63] L Minah Yang and Ian Grooms. “Machine learning techniques to construct patched analog ensembles for data assimilation”. In: *Journal of Computational Physics* 443 (2021), p. 110532.

[64] Ian Grooms. “Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders”. In: *Quarterly Journal of the Royal Meteorological Society* 147.734 (2021), pp. 139–149.

[65] Wei Zhao et al. “Comparison and correction of IDW based wind speed interpolation methods in urbanized Shenzhen”. In: *Frontiers of Earth Science* 16.3 (2022), pp. 798–808.

[66] UL Solutions. *Adjust to Mast Mean Wind Speeds and Energy Capture*. Accessed: 2025-06-06. 2024. URL: `https://openwind.ul-renewables.com/adjusttomastmeanwindspeedsandenergycapture`
`html`.

[67] Noel Cressie. “The origins of kriging”. In: *Mathematical geology* 22 (1990), pp. 239–252.

[68] A. Stein and L. C. A. Corsten. “Universal Kriging and Cokriging as a Regression Procedure”. In: *Biometrics* 47.2 (1991), pp. 575–587. ISSN: 0006341X, 15410420. URL: `http://www.jstor.org/stable/2532147` (visited on 06/06/2025).

[69] Jean Duchon. “Splines minimizing rotation-invariant semi-norms in Sobolev spaces”. In: *Constructive Theory of Functions of Several Variables*. Ed. by Walter Schempp and Karl Zeller. Berlin, Heidelberg: Springer Berlin Heidelberg, 1977, pp. 85–100. ISBN: 978-3-540-37496-1.

[70] Wolfgang Keller and Andrzej Borkowski. “Thin plate spline interpolation”. In: *Journal of Geodesy* 93 (2019), pp. 1251–1269.

[71] Bob Glahn et al. “The gridding of MOS”. In: *Weather and forecasting* 24.2 (2009), pp. 520–529.

[72] Daniel Vassallo, Raghavendra Krishnamurthy, and Harindra JS Fernando. “Decreasing wind speed extrapolation error via domain-specific feature extraction and selection”. In: *Wind Energy Science* 5.3 (2020), pp. 959–975.

[73] C. M. Leme Beu and E. Landulfo. “Machine-learning-based estimate of the wind speed over complex terrain using the long short-term memory (LSTM) recurrent neural network”. In: *Wind Energy Science* 9.6 (2024), pp. 1431–1450. DOI: `10.5194/wes-9-1431-2024`. URL: `https://wes.copernicus.org/articles/9/1431/2024/`.

[74] Pratik Nag, Ying Sun, and Brian J Reich. “Spatio-temporal DeepKriging for interpolation and probabilistic forecasting”. In: *Spatial Statistics* 57 (2023), p. 100773.

A

Experiment Results for Temporal Extrapolation

A.1. Meso vs. LES as MCP Reference Data

See Table A.1.

Table A.1: Performance comparison of methods on Meso and LES scales (Mean ± Std)

Method	Scale	MAE	RMSE	R^2
Identity	Meso	1.1989 ± 0.0384	1.5615 ± 0.0477	0.6675 ± 0.0535
	LES	1.2212 ± 0.0853	1.5935 ± 0.0899	0.6488 ± 0.0868
LR	Meso	1.1230 ± 0.0544	1.4578 ± 0.0682	0.7087 ± 0.0558
	LES	1.1195 ± 0.0512	1.4608 ± 0.0657	0.7086 ± 0.0494
Ridge	Meso	1.1231 ± 0.0544	1.4577 ± 0.0681	0.7087 ± 0.0559
	LES	1.1198 ± 0.0513	1.4609 ± 0.0658	0.7086 ± 0.0494
MLP	Meso	1.1247 ± 0.0519	1.4583 ± 0.0658	0.7083 ± 0.0570
	LES	1.1193 ± 0.0517	1.4603 ± 0.0669	0.7088 ± 0.0496
SVR	Meso	1.1165 ± 0.0506	1.4571 ± 0.0601	0.7094 ± 0.0532
	LES	1.1156 ± 0.0506	1.4593 ± 0.0654	0.7090 ± 0.0508
XGB	Meso	1.1220 ± 0.0527	1.4568 ± 0.0653	0.7091 ± 0.0559
	LES	1.1175 ± 0.0488	1.4581 ± 0.0612	0.7097 ± 0.0493

A.2. Additional Input Features

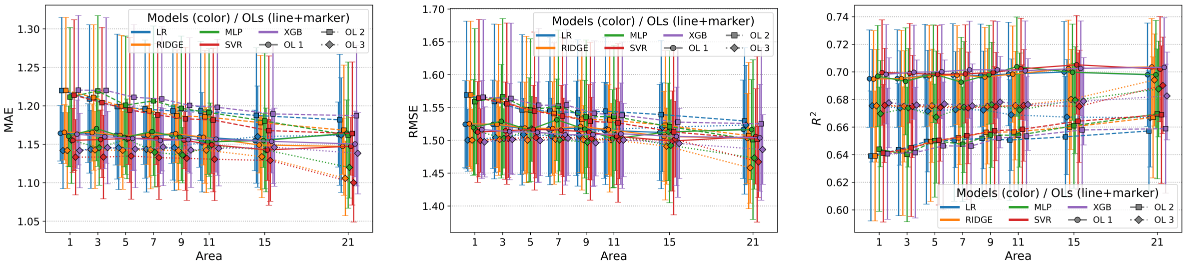
See Tables A.2, A.3, A.4, A.5 and A.6.

A.3. Additional Area

See Tables A.7, A.8 and A.9. Also see Figure A.1.

A.4. Additional Time Steps

See Tables A.10, A.11 and A.12.



(a) MAE

(b) RMSE

(c) R^2

Figure A.1: Performance of the baseline models when including spatial information from increasingly larger areas around the observation locations. Results are shown for all models and observation locations (OLs) in terms of MAE, RMSE, and R^2, with markers representing mean values and error bars showing the standard deviation.

Table A.2: Linear Regression (LR) Model Performance by Input Feature (IF) Set Grouped by Observation Location (OL). WS: Wind Speed; RT: Resolved Turbulence; WDIR: Wind Direction (decomposed); TOD: Time of Day; DOY: Day of Year. Performance is reported as Mean ± Standard Deviation from cross-validation.

		Linear Regression		
OL	**IF Set**	**MAE**	**RMSE**	$\mathbf{R^2}$
1	WS	1.1644 ± 0.0361	1.5245 ± 0.0663	0.6951 ± 0.0353
	WS, RT	1.1623 ± 0.0391	1.5214 ± 0.0703	0.6961 ± 0.0375
	WS, WDIR	1.1422 ± 0.0333	1.4965 ± 0.0628	0.7067 ± 0.0289
	WS, TOD	1.1562 ± 0.0261	1.5113 ± 0.0580	0.7003 ± 0.0344
	WS, DOY	1.1683 ± 0.0275	1.5353 ± 0.0523	0.6910 ± 0.0312
	WS, RT, WDIR	1.1415 ± 0.0336	1.4957 ± 0.0646	0.7069 ± 0.0303
	WS, RT, TOD	1.1560 ± 0.0278	1.5108 ± 0.0596	0.7004 ± 0.0354
	WS, RT, DOY	1.1630 ± 0.0292	1.5295 ± 0.0552	0.6933 ± 0.0325
	WS, WDIR, TOD	**1.1359 ± 0.0274**	**1.4882 ± 0.0580**	**0.7099 ± 0.0288**
	WS, WDIR, DOY	1.1475 ± 0.0297	1.5057 ± 0.0552	0.7031 ± 0.0272
	WS, TOD, DOY	1.1616 ± 0.0180	1.5231 ± 0.0442	0.6958 ± 0.0313
	WS, RT, WDIR, TOD	1.1363 ± 0.0271	1.4885 ± 0.0577	0.7097 ± 0.0291
	WS, RT, WDIR, DOY	1.1449 ± 0.0304	1.5037 ± 0.0573	0.7039 ± 0.0280
	WS, RT, TOD, DOY	1.1581 ± 0.0195	1.5202 ± 0.0449	0.6969 ± 0.0316
	WS, WDIR, TOD, DOY	1.1417 ± 0.0232	1.4977 ± 0.0506	0.7062 ± 0.0277
	WS, RT, WDIR, TOD, DOY	1.1412 ± 0.0232	1.4973 ± 0.0503	0.7063 ± 0.0277
2	WS	1.2199 ± 0.0950	1.5690 ± 0.1125	0.6391 ± 0.0471
	WS, RT	1.2201 ± 0.0947	1.5696 ± 0.1122	0.6388 ± 0.0472
	WS, WDIR	1.1933 ± 0.0741	1.5423 ± 0.0966	0.6519 ± 0.0366
	WS, TOD	1.2092 ± 0.0830	1.5532 ± 0.1039	0.6464 ± 0.0435
	WS, DOY	1.2112 ± 0.0787	1.5733 ± 0.0931	0.6375 ± 0.0391
	WS, RT, WDIR	1.1929 ± 0.0736	1.5423 ± 0.0960	0.6520 ± 0.0360
	WS, RT, TOD	1.2097 ± 0.0829	1.5539 ± 0.1038	0.6461 ± 0.0434
	WS, RT, DOY	1.2102 ± 0.0776	1.5731 ± 0.0924	0.6376 ± 0.0391
	WS, WDIR, TOD	1.1842 ± 0.0643	1.5298 ± 0.0882	**0.6577 ± 0.0334**
	WS, WDIR, DOY	1.1909 ± 0.0686	1.5495 ± 0.0893	0.6490 ± 0.0324
	WS, TOD, DOY	1.2033 ± 0.0691	1.5598 ± 0.0897	0.6436 ± 0.0386
	WS, RT, WDIR, TOD	1.1842 ± 0.0637	**1.5297 ± 0.0881**	**0.6577 ± 0.0330**
	WS, RT, WDIR, DOY	1.1913 ± 0.0685	1.5499 ± 0.0893	0.6488 ± 0.0323
	WS, RT, TOD, DOY	1.2025 ± 0.0689	1.5598 ± 0.0896	0.6436 ± 0.0385
	WS, WDIR, TOD, DOY	**1.1841 ± 0.0609**	1.5388 ± 0.0835	0.6538 ± 0.0311
	WS, RT, WDIR, TOD, DOY	1.1847 ± 0.0608	1.5393 ± 0.0837	0.6535 ± 0.0312
3	WS	1.1418 ± 0.0495	1.5003 ± 0.0474	0.6755 ± 0.0408
	WS, RT	1.1429 ± 0.0498	1.5010 ± 0.0475	0.6752 ± 0.0405
	WS, WDIR	1.1172 ± 0.0536	1.4769 ± 0.0500	0.6858 ± 0.0365
	WS, TOD	1.1417 ± 0.0472	1.4959 ± 0.0493	0.6773 ± 0.0419
	WS, DOY	1.1350 ± 0.0567	1.4975 ± 0.0427	0.6762 ± 0.0435
	WS, RT, WDIR	1.1171 ± 0.0574	1.4766 ± 0.0520	0.6860 ± 0.0363
	WS, RT, TOD	1.1427 ± 0.0474	1.4964 ± 0.0494	0.6771 ± 0.0416
	WS, RT, DOY	1.1357 ± 0.0571	1.4980 ± 0.0426	0.6760 ± 0.0435
	WS, WDIR, TOD	1.1146 ± 0.0574	1.4716 ± 0.0551	0.6879 ± 0.0388
	WS, WDIR, DOY	1.1141 ± 0.0550	1.4769 ± 0.0425	0.6853 ± 0.0405
	WS, TOD, DOY	1.1335 ± 0.0544	1.4906 ± 0.0454	0.6791 ± 0.0446
	WS, RT, WDIR, TOD	1.1153 ± 0.0592	1.4717 ± 0.0558	0.6878 ± 0.0391
	WS, RT, WDIR, DOY	1.1145 ± 0.0581	1.4762 ± 0.0434	0.6856 ± 0.0406
	WS, RT, TOD, DOY	1.1343 ± 0.0547	1.4911 ± 0.0450	0.6789 ± 0.0446
	WS, WDIR, TOD, DOY	**1.1124 ± 0.0587**	**1.4697 ± 0.0477**	**0.6882 ± 0.0423**
	WS, RT, WDIR, TOD, DOY	1.1128 ± 0.0609	1.4698 ± 0.0479	0.6880 ± 0.0428

Table A.3: Ridge Regression Model Performance by Input Feature (IF) Set Grouped by Observation Location (OL). WS: Wind Speed; RT: Resolved Turbulence; WDIR: Wind Direction (decomposed); TOD: Time of Day; DOY: Day of Year. Performance is reported as Mean ± Standard Deviation from cross-validation.

Ridge Regression				
OL	**IF Set**	**MAE**	**RMSE**	$\mathbf{R^2}$
1	WS	1.1656 ± 0.0360	1.5252 ± 0.0655	0.6948 ± 0.0350
	WS, RT	1.1629 ± 0.0394	1.5213 ± 0.0698	0.6962 ± 0.0372
	WS, WDIR	1.1425 ± 0.0332	1.4968 ± 0.0626	0.7066 ± 0.0288
	WS, TOD	1.1573 ± 0.0261	1.5121 ± 0.0574	0.7000 ± 0.0341
	WS, DOY	1.1733 ± 0.0279	1.5399 ± 0.0498	0.6894 ± 0.0290
	WS, RT, WDIR	1.1417 ± 0.0336	1.4959 ± 0.0644	0.7068 ± 0.0302
	WS, RT, TOD	1.1565 ± 0.0285	1.5110 ± 0.0595	0.7003 ± 0.0349
	WS, RT, DOY	1.1666 ± 0.0298	1.5331 ± 0.0530	0.6920 ± 0.0305
	WS, WDIR, TOD	**1.1367 ± 0.0275**	**1.4890 ± 0.0576**	**0.7096 ± 0.0283**
	WS, WDIR, DOY	1.1495 ± 0.0305	1.5080 ± 0.0544	0.7023 ± 0.0259
	WS, TOD, DOY	1.1650 ± 0.0185	1.5269 ± 0.0422	0.6944 ± 0.0296
	WS, RT, WDIR, TOD	1.1367 ± 0.0271	1.4891 ± 0.0573	0.7095 ± 0.0287
	WS, RT, WDIR, DOY	1.1465 ± 0.0307	1.5057 ± 0.0561	0.7032 ± 0.0269
	WS, RT, TOD, DOY	1.1606 ± 0.0211	1.5229 ± 0.0437	0.6960 ± 0.0301
	WS, WDIR, TOD, DOY	1.1438 ± 0.0244	1.5002 ± 0.0500	0.7053 ± 0.0265
	WS, RT, WDIR, TOD, DOY	1.1427 ± 0.0240	1.4992 ± 0.0497	0.7056 ± 0.0268
2	WS	1.2199 ± 0.0950	1.5690 ± 0.1124	0.6391 ± 0.0471
	WS, RT	1.2202 ± 0.0947	1.5696 ± 0.1121	0.6388 ± 0.0472
	WS, WDIR	1.1933 ± 0.0741	1.5423 ± 0.0966	0.6519 ± 0.0366
	WS, TOD	1.2093 ± 0.0830	1.5533 ± 0.1039	0.6464 ± 0.0435
	WS, DOY	1.2151 ± 0.0759	1.5762 ± 0.0887	0.6364 ± 0.0371
	WS, RT, WDIR	1.1929 ± 0.0736	1.5423 ± 0.0960	0.6520 ± 0.0360
	WS, RT, TOD	1.2098 ± 0.0830	1.5539 ± 0.1036	0.6461 ± 0.0434
	WS, RT, DOY	1.2132 ± 0.0757	1.5754 ± 0.0894	0.6367 ± 0.0375
	WS, WDIR, TOD	1.1842 ± 0.0643	1.5298 ± 0.0882	**0.6577 ± 0.0334**
	WS, WDIR, DOY	1.1922 ± 0.0681	1.5508 ± 0.0882	0.6485 ± 0.0314
	WS, TOD, DOY	1.2062 ± 0.0683	1.5631 ± 0.0860	0.6423 ± 0.0367
	WS, RT, WDIR, TOD	**1.1842 ± 0.0637**	**1.5297 ± 0.0881**	**0.6577 ± 0.0330**
	WS, RT, WDIR, DOY	1.1926 ± 0.0681	1.5511 ± 0.0882	0.6483 ± 0.0314
	WS, RT, TOD, DOY	1.2046 ± 0.0683	1.5620 ± 0.0873	0.6428 ± 0.0371
	WS, WDIR, TOD, DOY	1.1855 ± 0.0609	1.5403 ± 0.0825	0.6531 ± 0.0302
	WS, RT, WDIR, TOD, DOY	1.1856 ± 0.0608	1.5403 ± 0.0832	0.6531 ± 0.0306
3	WS	1.1419 ± 0.0494	1.5004 ± 0.0473	0.6754 ± 0.0408
	WS, RT	1.1429 ± 0.0497	1.5008 ± 0.0473	0.6753 ± 0.0403
	WS, WDIR	1.1174 ± 0.0536	1.4771 ± 0.0499	0.6858 ± 0.0365
	WS, TOD	1.1417 ± 0.0473	1.4955 ± 0.0490	0.6775 ± 0.0413
	WS, DOY	1.1388 ± 0.0566	1.4996 ± 0.0434	0.6754 ± 0.0434
	WS, RT, WDIR	1.1176 ± 0.0574	1.4768 ± 0.0519	0.6859 ± 0.0363
	WS, RT, TOD	1.1427 ± 0.0475	1.4959 ± 0.0490	0.6773 ± 0.0409
	WS, RT, DOY	1.1397 ± 0.0568	1.5001 ± 0.0429	0.6751 ± 0.0434
	WS, WDIR, TOD	1.1150 ± 0.0574	1.4717 ± 0.0552	0.6879 ± 0.0385
	WS, WDIR, DOY	1.1142 ± 0.0550	1.4768 ± 0.0425	0.6853 ± 0.0404
	WS, TOD, DOY	1.1366 ± 0.0542	1.4928 ± 0.0455	0.6782 ± 0.0443
	WS, RT, WDIR, TOD	1.1156 ± 0.0594	1.4716 ± 0.0558	0.6879 ± 0.0387
	WS, RT, WDIR, DOY	1.1148 ± 0.0582	1.4763 ± 0.0435	0.6856 ± 0.0406
	WS, RT, TOD, DOY	1.1375 ± 0.0545	1.4934 ± 0.0447	0.6779 ± 0.0445
	WS, WDIR, TOD, DOY	**1.1128 ± 0.0587**	**1.4699 ± 0.0477**	**0.6881 ± 0.0423**
	WS, RT, WDIR, TOD, DOY	1.1133 ± 0.0609	1.4701 ± 0.0479	0.6879 ± 0.0430

Table A.4: Multi Layer Perceptron (MLP) Model Performance by Input Feature (IF) Set Grouped by Observation Location (OL). WS: Wind Speed; RT: Resolved Turbulence; WDIR: Wind Direction (decomposed); TOD: Time of Day; DOY: Day of Year. Performance is reported as Mean ± Standard Deviation from cross-validation.

		Multi Layer Perceptron		
OL	**IF Set**	**MAE**	**RMSE**	$\mathbf{R}^2$
1	WS	1.1606 ± 0.0391	1.5196 ± 0.0729	0.6969 ± 0.0367
	WS, RT	1.1467 ± 0.0559	1.4939 ± 0.0893	0.7066 ± 0.0412
	WS, WDIR	1.1133 ± 0.0335	1.4656 ± 0.0726	0.7181 ± 0.0334
	WS, TOD	1.1484 ± 0.0359	1.4931 ± 0.0715	0.7068 ± 0.0388
	WS, DOY	1.1750 ± 0.0565	1.5340 ± 0.0860	0.6924 ± 0.0286
	WS, RT, WDIR	1.0989 ± 0.0392	1.4391 ± 0.0740	0.7277 ± 0.0362
	WS, RT, TOD	1.1399 ± 0.0360	1.4832 ± 0.0705	0.7111 ± 0.0355
	WS, RT, DOY	1.1662 ± 0.0254	1.5227 ± 0.0619	0.6954 ± 0.0385
	WS, WDIR, TOD	1.1226 ± 0.0234	1.4654 ± 0.0495	0.7189 ± 0.0249
	WS, WDIR, DOY	1.1340 ± 0.0822	1.4813 ± 0.1153	0.7114 ± 0.0448
	WS, TOD, DOY	1.1514 ± 0.0297	1.5009 ± 0.0521	0.7051 ± 0.0260
	WS, RT, WDIR, TOD	**1.0738 ± 0.0285**	**1.4081 ± 0.0670**	**0.7395 ± 0.0327**
	WS, RT, WDIR, DOY	1.1077 ± 0.0671	1.4505 ± 0.1090	0.7231 ± 0.0429
	WS, RT, TOD, DOY	1.1461 ± 0.0480	1.4876 ± 0.0813	0.7094 ± 0.0364
	WS, WDIR, TOD, DOY	1.0985 ± 0.0436	1.4397 ± 0.0706	0.7279 ± 0.0321
	WS, RT, WDIR, TOD, DOY	1.0936 ± 0.0462	1.4274 ± 0.0858	0.7314 ± 0.0412
2	WS	1.2113 ± 0.0905	1.5588 ± 0.1112	0.6439 ± 0.0450
	WS, RT	1.2184 ± 0.1026	1.5688 ± 0.1200	0.6393 ± 0.0474
	WS, WDIR	1.1734 ± 0.0681	1.5141 ± 0.1047	0.6636 ± 0.0441
	WS, TOD	1.2098 ± 0.0972	1.5434 ± 0.1236	0.6509 ± 0.0470
	WS, DOY	1.1972 ± 0.0816	1.5496 ± 0.1078	0.6484 ± 0.0407
	WS, RT, WDIR	1.1474 ± 0.0848	1.4812 ± 0.1102	0.6784 ± 0.0416
	WS, RT, TOD	1.2023 ± 0.1025	1.5357 ± 0.1279	0.6543 ± 0.0481
	WS, RT, DOY	1.1972 ± 0.0790	1.5504 ± 0.1065	0.6475 ± 0.0449
	WS, WDIR, TOD	1.1446 ± 0.0584	1.4739 ± 0.0879	0.6815 ± 0.0370
	WS, WDIR, DOY	1.1473 ± 0.0944	1.4843 ± 0.1172	0.6771 ± 0.0439
	WS, TOD, DOY	1.1920 ± 0.0980	1.5270 ± 0.1222	0.6584 ± 0.0454
	WS, RT, WDIR, TOD	1.1271 ± 0.0704	**1.4554 ± 0.0996**	0.6894 ± 0.0387
	WS, RT, WDIR, DOY	1.1533 ± 0.0780	1.4899 ± 0.1022	0.6748 ± 0.0392
	WS, RT, TOD, DOY	1.1745 ± 0.0884	1.5225 ± 0.1199	0.6597 ± 0.0495
	WS, WDIR, TOD, DOY	1.1408 ± 0.0622	1.4687 ± 0.0922	0.6838 ± 0.0376
	WS, RT, WDIR, TOD, DOY	**1.1258 ± 0.0776**	1.4521 ± 0.1084	**0.6905 ± 0.0423**
3	WS	1.1553 ± 0.0563	1.5144 ± 0.0502	0.6697 ± 0.0385
	WS, RT	1.1420 ± 0.0420	1.5014 ± 0.0452	0.6755 ± 0.0349
	WS, WDIR	1.1005 ± 0.0441	1.4512 ± 0.0505	0.6972 ± 0.0311
	WS, TOD	1.1343 ± 0.0522	1.4837 ± 0.0625	0.6834 ± 0.0354
	WS, DOY	1.1365 ± 0.0370	1.5005 ± 0.0188	0.6740 ± 0.0493
	WS, RT, WDIR	1.0925 ± 0.0555	1.4399 ± 0.0531	0.7022 ± 0.0281
	WS, RT, TOD	1.1393 ± 0.0546	1.4955 ± 0.0625	0.6777 ± 0.0415
	WS, RT, DOY	1.1277 ± 0.0487	1.4814 ± 0.0422	0.6836 ± 0.0390
	WS, WDIR, TOD	1.0746 ± 0.0432	1.4101 ± 0.0425	**0.7136 ± 0.0333**
	WS, WDIR, DOY	1.1271 ± 0.0579	1.4888 ± 0.0446	0.6784 ± 0.0541
	WS, TOD, DOY	1.1477 ± 0.0566	1.4954 ± 0.0443	0.6773 ± 0.0421
	WS, RT, WDIR, TOD	**1.0887 ± 0.0531**	**1.4327 ± 0.0504**	0.7048 ± 0.0310
	WS, RT, WDIR, DOY	1.1283 ± 0.0689	1.4935 ± 0.0601	0.6762 ± 0.0589
	WS, RT, TOD, DOY	1.1352 ± 0.0510	1.4920 ± 0.0389	0.6781 ± 0.0471
	WS, WDIR, TOD, DOY	1.1202 ± 0.0600	1.4726 ± 0.0505	0.6854 ± 0.0536
	WS, RT, WDIR, TOD, DOY	1.1294 ± 0.0885	1.4906 ± 0.0844	0.6748 ± 0.0776

Table A.5: Support Vector Regression (SVR) Model Performance by Input Feature (IF) Set Grouped by Observation Location (OL). WS: Wind Speed; RT: Resolved Turbulence; WDIR: Wind Direction (decomposed); TOD: Time of Day; DOY: Day of Year. Performance is reported as Mean ± Standard Deviation from cross-validation.

		Support Vector Regression (SVR)		
OL	**IF Set**	**MAE**	**RMSE**	$\mathbf{R^2}$
1	WS	1.1551 ± 0.0417	1.5127 ± 0.0769	0.6995 ± 0.0382
	WS, RT	1.1439 ± 0.0475	1.5025 ± 0.0889	0.7031 ± 0.0419
	WS, WDIR	1.1109 ± 0.0365	1.4619 ± 0.0778	0.7190 ± 0.0378
	WS, TOD	1.1299 ± 0.0348	1.4800 ± 0.0765	0.7124 ± 0.0360
	WS, DOY	1.1933 ± 0.0516	1.5689 ± 0.0764	0.6777 ± 0.0325
	WS, RT, WDIR	1.0871 ± 0.0455	1.4406 ± 0.0902	0.7265 ± 0.0415
	WS, RT, TOD	1.1286 ± 0.0393	1.4757 ± 0.0827	0.7138 ± 0.0383
	WS, RT, DOY	1.1918 ± 0.0525	1.5737 ± 0.0854	0.6749 ± 0.0399
	WS, WDIR, TOD	1.0739 ± 0.0309	**1.4158 ± 0.0691**	**0.7368 ± 0.0319**
	WS, WDIR, DOY	1.1071 ± 0.0432	1.4558 ± 0.0849	0.7217 ± 0.0361
	WS, TOD, DOY	1.1513 ± 0.0439	1.5098 ± 0.0767	0.7006 ± 0.0378
	WS, RT, WDIR, TOD	**1.0714 ± 0.0400**	1.4165 ± 0.0844	0.7359 ± 0.0379
	WS, RT, WDIR, DOY	1.1037 ± 0.0532	1.4549 ± 0.0953	0.7216 ± 0.0396
	WS, RT, TOD, DOY	1.1574 ± 0.0436	1.5173 ± 0.0685	0.6978 ± 0.0354
	WS, WDIR, TOD, DOY	1.0886 ± 0.0334	1.4270 ± 0.0632	0.7330 ± 0.0285
	WS, RT, WDIR, TOD, DOY	1.0928 ± 0.0402	1.4412 ± 0.0787	0.7271 ± 0.0352
2	WS	1.2137 ± 0.0981	1.5642 ± 0.1198	0.6411 ± 0.0499
	WS, RT	1.2100 ± 0.1126	1.5614 ± 0.1364	0.6422 ± 0.0550
	WS, WDIR	1.1540 ± 0.0773	1.4972 ± 0.1118	0.6712 ± 0.0438
	WS, TOD	1.1959 ± 0.0948	1.5396 ± 0.1244	0.6525 ± 0.0477
	WS, DOY	1.2252 ± 0.1079	1.5897 ± 0.1353	0.6286 ± 0.0605
	WS, RT, WDIR	1.1422 ± 0.0790	1.4909 ± 0.1159	0.6739 ± 0.0448
	WS, RT, TOD	1.1929 ± 0.1037	1.5323 ± 0.1332	0.6556 ± 0.0510
	WS, RT, DOY	1.2009 ± 0.0788	1.5605 ± 0.0995	0.6429 ± 0.0436
	WS, WDIR, TOD	1.1274 ± 0.0684	1.4662 ± 0.1013	0.6848 ± 0.0396
	WS, WDIR, DOY	1.1320 ± 0.0808	1.4749 ± 0.1125	0.6807 ± 0.0442
	WS, TOD, DOY	1.2108 ± 0.1045	1.5663 ± 0.1326	0.6396 ± 0.0585
	WS, RT, WDIR, TOD	1.1169 ± 0.0749	1.4504 ± 0.1050	0.6913 ± 0.0410
	WS, RT, WDIR, DOY	1.1339 ± 0.0755	1.4802 ± 0.1055	0.6785 ± 0.0430
	WS, RT, TOD, DOY	1.1829 ± 0.0847	1.5291 ± 0.1112	0.6571 ± 0.0452
	WS, WDIR, TOD, DOY	1.1166 ± 0.0764	1.4549 ± 0.1046	0.6889 ± 0.0446
	WS, RT, WDIR, TOD, DOY	**1.1109 ± 0.0760**	**1.4440 ± 0.1025**	**0.6937 ± 0.0427**
3	WS	1.1332 ± 0.0490	1.4996 ± 0.0485	0.6757 ± 0.0414
	WS, RT	1.1330 ± 0.0486	1.4981 ± 0.0422	0.6759 ± 0.0444
	WS, WDIR	1.0858 ± 0.0620	1.4540 ± 0.0792	0.6967 ± 0.0290
	WS, TOD	1.1305 ± 0.0595	1.4896 ± 0.0636	0.6805 ± 0.0382
	WS, DOY	1.1249 ± 0.0446	1.4940 ± 0.0377	0.6771 ± 0.0479
	WS, RT, WDIR	1.0777 ± 0.0579	1.4478 ± 0.0590	0.6987 ± 0.0307
	WS, RT, TOD	1.1342 ± 0.0570	1.4955 ± 0.0634	0.6780 ± 0.0379
	WS, RT, DOY	1.1268 ± 0.0442	1.4962 ± 0.0292	0.6756 ± 0.0513
	WS, WDIR, TOD	1.0773 ± 0.0517	**1.4322 ± 0.0479**	**0.7047 ± 0.0332**
	WS, WDIR, DOY	1.1096 ± 0.0570	1.4852 ± 0.0472	0.6793 ± 0.0598
	WS, TOD, DOY	1.1484 ± 0.0679	1.5210 ± 0.0629	0.6654 ± 0.0514
	WS, RT, WDIR, TOD	**1.0813 ± 0.0586**	1.4429 ± 0.0521	0.6997 ± 0.0388
	WS, RT, WDIR, DOY	1.1145 ± 0.0747	1.4863 ± 0.0630	0.6778 ± 0.0678
	WS, RT, TOD, DOY	1.1348 ± 0.0528	1.5042 ± 0.0410	0.6726 ± 0.0494
	WS, WDIR, TOD, DOY	1.1086 ± 0.0744	1.4802 ± 0.0705	0.6804 ± 0.0685
	WS, RT, WDIR, TOD, DOY	1.1090 ± 0.0810	1.4773 ± 0.0732	0.6816 ± 0.0690

Table A.6: XGBoost (XGB) Model Performance by Input Feature (IF) Set Grouped by Observation Location (OL). WS: Wind Speed; RT: Resolved Turbulence; WDIR: Wind Direction (decomposed); TOD: Time of Day; DOY: Day of Year. Performance is reported as Mean ± Standard Deviation from cross-validation.

		XGBoost (XGB)		
OL	**IF Set**	**MAE**	**RMSE**	$\mathbf{R^2}$
1	WS	1.1631 ± 0.0444	1.5164 ± 0.0745	0.6983 ± 0.0360
	WS, RT	1.1551 ± 0.0461	1.5134 ± 0.0869	0.6992 ± 0.0391
	WS, WDIR	1.1129 ± 0.0380	1.4652 ± 0.0738	0.7184 ± 0.0329
	WS, TOD	1.1439 ± 0.0313	1.4895 ± 0.0715	0.7091 ± 0.0329
	WS, DOY	1.1930 ± 0.0576	1.5524 ± 0.0866	0.6838 ± 0.0381
	WS, RT, WDIR	1.0957 ± 0.0369	1.4461 ± 0.0750	0.7258 ± 0.0309
	WS, RT, TOD	1.1285 ± 0.0351	1.4737 ± 0.0762	0.7147 ± 0.0363
	WS, RT, DOY	1.1857 ± 0.0675	1.5421 ± 0.0977	0.6883 ± 0.0379
	WS, WDIR, TOD	1.0929 ± 0.0222	1.4316 ± 0.0577	0.7313 ± 0.0278
	WS, WDIR, DOY	1.1141 ± 0.0436	1.4640 ± 0.0789	0.7189 ± 0.0328
	WS, TOD, DOY	1.1777 ± 0.0263	1.5333 ± 0.0545	0.6922 ± 0.0271
	WS, RT, WDIR, TOD	**1.0727 ± 0.0237**	**1.4119 ± 0.0665**	**0.7384 ± 0.0304**
	WS, RT, WDIR, DOY	1.1303 ± 0.0405	1.4763 ± 0.0783	0.7145 ± 0.0305
	WS, RT, TOD, DOY	1.1595 ± 0.0473	1.5100 ± 0.0787	0.7013 ± 0.0316
	WS, WDIR, TOD, DOY	1.1101 ± 0.0402	1.4541 ± 0.0666	0.7233 ± 0.0248
	WS, RT, WDIR, TOD, DOY	1.0968 ± 0.0300	1.4357 ± 0.0653	0.7300 ± 0.0268
2	WS	1.2203 ± 0.0972	1.5656 ± 0.1181	0.6407 ± 0.0475
	WS, RT	1.2176 ± 0.1041	1.5648 ± 0.1233	0.6411 ± 0.0489
	WS, WDIR	1.1660 ± 0.0806	1.5049 ± 0.1108	0.6683 ± 0.0406
	WS, TOD	1.2079 ± 0.0855	1.5470 ± 0.1114	0.6492 ± 0.0455
	WS, DOY	1.2406 ± 0.0790	1.5960 ± 0.0992	0.6275 ± 0.0381
	WS, RT, WDIR	1.1593 ± 0.0701	1.5044 ± 0.0988	0.6690 ± 0.0343
	WS, RT, TOD	1.1928 ± 0.0938	1.5347 ± 0.1168	0.6547 ± 0.0459
	WS, RT, DOY	1.2357 ± 0.0805	1.5919 ± 0.0988	0.6290 ± 0.0406
	WS, WDIR, TOD	1.1516 ± 0.0656	1.4834 ± 0.0948	0.6780 ± 0.0343
	WS, WDIR, DOY	1.1613 ± 0.0884	1.5055 ± 0.1104	0.6684 ± 0.0390
	WS, TOD, DOY	1.2206 ± 0.0783	1.5691 ± 0.1028	0.6390 ± 0.0447
	WS, RT, WDIR, TOD	**1.1346 ± 0.0813**	**1.4687 ± 0.1113**	**0.6840 ± 0.0401**
	WS, RT, WDIR, DOY	1.1612 ± 0.0765	1.5109 ± 0.1027	0.6659 ± 0.0377
	WS, RT, TOD, DOY	1.2228 ± 0.0886	1.5661 ± 0.1114	0.6407 ± 0.0434
	WS, WDIR, TOD, DOY	1.1626 ± 0.0682	1.5014 ± 0.0964	0.6701 ± 0.0354
	WS, RT, WDIR, TOD, DOY	1.1358 ± 0.0763	1.4698 ± 0.0974	0.6834 ± 0.0386
3	WS	1.1418 ± 0.0485	1.4975 ± 0.0534	0.6776 ± 0.0326
	WS, RT	1.1468 ± 0.0457	1.5021 ± 0.0504	0.6755 ± 0.0336
	WS, WDIR	1.0923 ± 0.0549	1.4414 ± 0.0508	0.7012 ± 0.0307
	WS, TOD	1.1353 ± 0.0493	1.4889 ± 0.0615	0.6814 ± 0.0324
	WS, DOY	1.1425 ± 0.0492	1.4976 ± 0.0464	0.6770 ± 0.0368
	WS, RT, WDIR	1.0949 ± 0.0600	1.4488 ± 0.0499	0.6979 ± 0.0332
	WS, RT, TOD	1.1498 ± 0.0490	1.5061 ± 0.0616	0.6735 ± 0.0370
	WS, RT, DOY	1.1689 ± 0.0717	1.5220 ± 0.0774	0.6676 ± 0.0308
	WS, WDIR, TOD	**1.0791 ± 0.0671**	**1.4311 ± 0.0684**	**0.7058 ± 0.0294**
	WS, WDIR, DOY	1.1087 ± 0.0749	1.4679 ± 0.0850	0.6906 ± 0.0326
	WS, TOD, DOY	1.1474 ± 0.0651	1.5021 ± 0.0791	0.6760 ± 0.0330
	WS, RT, WDIR, TOD	1.0810 ± 0.0671	1.4334 ± 0.0630	0.7043 ± 0.0342
	WS, RT, WDIR, DOY	1.1069 ± 0.0587	1.4549 ± 0.0545	0.6955 ± 0.0324
	WS, RT, TOD, DOY	1.1501 ± 0.0583	1.5033 ± 0.0651	0.6748 ± 0.0375
	WS, WDIR, TOD, DOY	1.0964 ± 0.0701	1.4477 ± 0.0768	0.6989 ± 0.0316
	WS, RT, WDIR, TOD, DOY	1.0900 ± 0.0636	1.4401 ± 0.0644	0.7021 ± 0.0289

Table A.7: Performance comparison on Observation Location 1 (Mean ± Std)

Method	**Size (n)**	**MAE**	**RMSE**	**R^2**
LR	1	1.1644 ± 0.0361	1.5245 ± 0.0663	0.6951 ± 0.0353
	3	1.1644 ± 0.0382	1.5243 ± 0.0674	0.6951 ± 0.0354
	5	1.1616 ± 0.0385	1.5183 ± 0.0705	0.6973 ± 0.0372
	7	1.1617 ± 0.0378	1.5172 ± 0.0726	0.6977 ± 0.0372
	9	1.1622 ± 0.0371	1.5181 ± 0.0720	0.6974 ± 0.0373
	11	1.1594 ± 0.0394	1.5153 ± 0.0731	0.6985 ± 0.0375
	15	1.1549 ± 0.0394	1.5120 ± 0.0733	0.6999 ± 0.0367
	21	1.1626 ± 0.0422	1.5167 ± 0.0750	0.6979 ± 0.0375
	31	1.1620 ± 0.0383	1.5203 ± 0.0735	0.6968 ± 0.0350
Ridge	1	1.1656 ± 0.0360	1.5252 ± 0.0655	0.6948 ± 0.0350
	3	1.1652 ± 0.0372	1.5243 ± 0.0661	0.6952 ± 0.0349
	5	1.1615 ± 0.0386	1.5187 ± 0.0699	0.6972 ± 0.0366
	7	1.1621 ± 0.0383	1.5179 ± 0.0713	0.6976 ± 0.0359
	9	1.1638 ± 0.0379	1.5212 ± 0.0706	0.6963 ± 0.0359
	11	1.1591 ± 0.0361	1.5162 ± 0.0684	0.6984 ± 0.0349
	15	1.1491 ± 0.0364	1.5060 ± 0.0688	0.7023 ± 0.0352
	21	1.1473 ± 0.0364	1.5028 ± 0.0685	0.7036 ± 0.0351
	31	1.1371 ± 0.0326	1.4928 ± 0.0654	0.7077 ± 0.0330
	41	1.1322 ± 0.0307	1.4868 ± 0.0652	0.7101 ± 0.0323
	51	1.1253 ± 0.0248	1.4780 ± 0.0612	0.7135 ± 0.0317
	61	1.1159 ± 0.0219	1.4681 ± 0.0551	0.7173 ± 0.0310
	71	1.0971 ± 0.0229	1.4469 ± 0.0504	0.7256 ± 0.0284
	81	1.0904 ± 0.0220	1.4417 ± 0.0507	0.7276 ± 0.0280
	91	1.0829 ± 0.0209	1.4332 ± 0.0486	0.7309 ± 0.0264
	101	1.0793 ± 0.0219	1.4286 ± 0.0499	0.7326 ± 0.0259
	111	1.0738 ± 0.0193	1.4175 ± 0.0467	0.7366 ± 0.0269
MLP	1	1.1606 ± 0.0391	1.5196 ± 0.0729	0.6969 ± 0.0367
	3	1.1704 ± 0.0421	1.5290 ± 0.0865	0.6931 ± 0.0388
	5	1.1589 ± 0.0383	1.5157 ± 0.0795	0.6986 ± 0.0354
	7	1.1665 ± 0.0264	1.5313 ± 0.0621	0.6926 ± 0.0325
	9	1.1591 ± 0.0340	1.5170 ± 0.0744	0.6980 ± 0.0358
	11	1.1531 ± 0.0423	1.5023 ± 0.0748	0.7038 ± 0.0360
	15	1.1548 ± 0.0349	1.5128 ± 0.0744	0.6998 ± 0.0349
	21	1.1644 ± 0.0431	1.5163 ± 0.0762	0.6980 ± 0.0393
	31	1.1487 ± 0.0320	1.5024 ± 0.0644	0.7044 ± 0.0290
SVR	1	1.1551 ± 0.0417	1.5127 ± 0.0769	0.6995 ± 0.0382
	3	1.1561 ± 0.0384	1.5145 ± 0.0741	0.6990 ± 0.0361
	5	1.1576 ± 0.0340	1.5166 ± 0.0671	0.6981 ± 0.0354
	7	1.1577 ± 0.0341	1.5155 ± 0.0682	0.6985 ± 0.0356
	9	1.1553 ± 0.0370	1.5125 ± 0.0724	0.6996 ± 0.0373
	11	1.1493 ± 0.0373	1.5060 ± 0.0735	0.7022 ± 0.0367
	15	1.1427 ± 0.0397	1.4990 ± 0.0762	0.7051 ± 0.0358
	21	1.1473 ± 0.0375	1.5059 ± 0.0726	0.7020 ± 0.0382
	31	1.1352 ± 0.0296	1.4967 ± 0.0642	0.7058 ± 0.0350
XGB	1	1.1631 ± 0.0444	1.5164 ± 0.0745	0.6983 ± 0.0360
	3	1.1598 ± 0.0416	1.5132 ± 0.0734	0.6997 ± 0.0341
	5	1.1576 ± 0.0420	1.5123 ± 0.0765	0.7002 ± 0.0341
	7	1.1549 ± 0.0431	1.5084 ± 0.0766	0.7015 ± 0.0353
	9	1.1561 ± 0.0420	1.5087 ± 0.0791	0.7015 ± 0.0351
	11	1.1566 ± 0.0428	1.5099 ± 0.0773	0.7009 ± 0.0360
	15	1.1538 ± 0.0411	1.5059 ± 0.0751	0.7026 ± 0.0343
	21	1.1503 ± 0.0428	1.5039 ± 0.0813	0.7033 ± 0.0359
	31	1.1351 ± 0.0462	1.4900 ± 0.0841	0.7089 ± 0.0347

Table A.8: Performance comparison on Observation Location 2 (Mean ± Std)

Method	Size (n)	MAE	RMSE	R^2
LR	1	1.2199 ± 0.0950	1.5690 ± 0.1125	0.6391 ± 0.0471
	3	1.2089 ± 0.0964	1.5592 ± 0.1178	0.6436 ± 0.0478
	5	1.1989 ± 0.0941	1.5464 ± 0.1146	0.6496 ± 0.0454
	7	1.1957 ± 0.0898	1.5444 ± 0.1122	0.6505 ± 0.0443
	9	1.1933 ± 0.0899	1.5417 ± 0.1114	0.6518 ± 0.0436
	11	1.1932 ± 0.0906	1.5444 ± 0.1118	0.6506 ± 0.0437
	15	1.1874 ± 0.0880	1.5392 ± 0.1130	0.6529 ± 0.0438
	21	1.1818 ± 0.0851	1.5295 ± 0.1114	0.6572 ± 0.0436
	31	1.1899 ± 0.0879	1.5334 ± 0.1105	0.6554 ± 0.0433
Ridge	1	1.2199 ± 0.0950	1.5690 ± 0.1124	0.6391 ± 0.0471
	3	1.2108 ± 0.0945	1.5617 ± 0.1151	0.6426 ± 0.0466
	5	1.1991 ± 0.0924	1.5469 ± 0.1114	0.6495 ± 0.0434
	7	1.1935 ± 0.0861	1.5408 ± 0.1077	0.6522 ± 0.0425
	9	1.1867 ± 0.0860	1.5320 ± 0.1077	0.6562 ± 0.0425
	11	1.1857 ± 0.0864	1.5310 ± 0.1098	0.6565 ± 0.0436
	15	1.1781 ± 0.0874	1.5229 ± 0.1138	0.6599 ± 0.0459
	21	1.1683 ± 0.0846	1.5071 ± 0.1114	0.6670 ± 0.0439
	31	1.1640 ± 0.0891	1.4991 ± 0.1151	0.6702 ± 0.0467
	41	1.1524 ± 0.0894	1.4885 ± 0.1175	0.6749 ± 0.0465
	51	1.1487 ± 0.0853	1.4810 ± 0.1146	0.6781 ± 0.0458
	61	1.1423 ± 0.0822	1.4723 ± 0.1119	0.6819 ± 0.0453
	71	1.1329 ± 0.0801	1.4613 ± 0.1108	0.6865 ± 0.0453
	81	1.1273 ± 0.0792	1.4528 ± 0.1109	0.6901 ± 0.0452
	91	1.1151 ± 0.0773	1.4398 ± 0.1100	0.6955 ± 0.0447
	101	1.1103 ± 0.0768	1.4350 ± 0.1077	0.6977 ± 0.0435
	111	1.1061 ± 0.0758	1.4303 ± 0.1063	0.6997 ± 0.0425
MLP	1	1.2113 ± 0.0905	1.5588 ± 0.1112	0.6439 ± 0.0450
	3	1.2193 ± 0.0979	1.5664 ± 0.1191	0.6402 ± 0.0488
	5	1.2008 ± 0.0898	1.5459 ± 0.1087	0.6502 ± 0.0410
	7	1.2064 ± 0.0972	1.5522 ± 0.1157	0.6475 ± 0.0415
	9	1.1967 ± 0.0940	1.5356 ± 0.1171	0.6543 ± 0.0465
	11	1.1913 ± 0.0829	1.5364 ± 0.1080	0.6540 ± 0.0440
	15	1.1802 ± 0.0882	1.5208 ± 0.1160	0.6610 ± 0.0453
	21	1.1619 ± 0.0949	1.5010 ± 0.1216	0.6695 ± 0.0470
	31	1.1667 ± 0.0963	1.4999 ± 0.1222	0.6696 ± 0.0502
SVR	1	1.2137 ± 0.0981	1.5642 ± 0.1198	0.6411 ± 0.0499
	3	1.2042 ± 0.0977	1.5557 ± 0.1228	0.6450 ± 0.0499
	5	1.1946 ± 0.0942	1.5446 ± 0.1204	0.6503 ± 0.0468
	7	1.1881 ± 0.0928	1.5370 ± 0.1197	0.6538 ± 0.0464
	9	1.1831 ± 0.0929	1.5290 ± 0.1201	0.6573 ± 0.0465
	11	1.1815 ± 0.0934	1.5266 ± 0.1208	0.6584 ± 0.0467
	15	1.1678 ± 0.0970	1.5127 ± 0.1262	0.6642 ± 0.0495
	21	1.1639 ± 0.0930	1.5016 ± 0.1262	0.6690 ± 0.0499
	31	1.1522 ± 0.0937	1.4897 ± 0.1244	0.6738 ± 0.0513
XGB	1	1.2203 ± 0.0972	1.5656 ± 0.1181	0.6407 ± 0.0475
	3	1.2199 ± 0.0946	1.5637 ± 0.1147	0.6415 ± 0.0474
	5	1.2110 ± 0.0947	1.5534 ± 0.1170	0.6464 ± 0.0463
	7	1.2090 ± 0.0936	1.5536 ± 0.1156	0.6462 ± 0.0465
	9	1.2005 ± 0.0880	1.5410 ± 0.1104	0.6520 ± 0.0443
	11	1.1980 ± 0.0885	1.5382 ± 0.1114	0.6534 ± 0.0431
	15	1.1894 ± 0.0907	1.5277 ± 0.1145	0.6579 ± 0.0451
	21	1.1872 ± 0.0913	1.5251 ± 0.1163	0.6589 ± 0.0467
	31	1.1818 ± 0.0940	1.5154 ± 0.1195	0.6632 ± 0.0469

Table A.9: Performance comparison on Observation Location 3 (Mean ± Std)

Method	**Size (n)**	**MAE**	**RMSE**	**R^2**
LR	1	1.1418 ± 0.0495	1.5003 ± 0.0474	0.6755 ± 0.0408
	3	1.1433 ± 0.0507	1.5041 ± 0.0527	0.6739 ± 0.0414
	5	1.1454 ± 0.0543	1.5088 ± 0.0553	0.6716 ± 0.0441
	7	1.1433 ± 0.0555	1.5070 ± 0.0573	0.6726 ± 0.0420
	9	1.1455 ± 0.0557	1.5074 ± 0.0583	0.6724 ± 0.0427
	11	1.1526 ± 0.0646	1.5166 ± 0.0717	0.6689 ± 0.0405
	15	1.1599 ± 0.0658	1.5198 ± 0.0675	0.6672 ± 0.0418
	21	1.1618 ± 0.0745	1.5233 ± 0.0910	0.6668 ± 0.0358
	31	1.1452 ± 0.0803	1.5037 ± 0.0927	0.6751 ± 0.0381
Ridge	1	1.1419 ± 0.0494	1.5004 ± 0.0473	0.6754 ± 0.0408
	3	1.1426 ± 0.0502	1.5015 ± 0.0496	0.6750 ± 0.0403
	5	1.1422 ± 0.0507	1.5003 ± 0.0495	0.6754 ± 0.0410
	7	1.1426 ± 0.0506	1.5017 ± 0.0516	0.6748 ± 0.0421
	9	1.1417 ± 0.0510	1.5012 ± 0.0518	0.6750 ± 0.0418
	11	1.1423 ± 0.0552	1.5008 ± 0.0555	0.6755 ± 0.0394
	15	1.1338 ± 0.0533	1.4911 ± 0.0579	0.6800 ± 0.0363
	21	1.1056 ± 0.0485	1.4581 ± 0.0537	0.6941 ± 0.0334
	31	1.0754 ± 0.0564	1.4247 ± 0.0555	0.7081 ± 0.0317
	41	1.0663 ± 0.0617	1.4152 ± 0.0600	0.7121 ± 0.0307
	51	1.0624 ± 0.0614	1.4108 ± 0.0582	0.7138 ± 0.0308
	61	1.0561 ± 0.0607	1.4001 ± 0.0556	0.7180 ± 0.0308
	71	1.0519 ± 0.0569	1.3929 ± 0.0498	0.7207 ± 0.0319
	81	1.0484 ± 0.0593	1.3857 ± 0.0513	0.7236 ± 0.0318
	91	1.0447 ± 0.0585	1.3829 ± 0.0511	0.7247 ± 0.0317
	101	1.0374 ± 0.0568	1.3745 ± 0.0492	0.7279 ± 0.0318
	111	1.0310 ± 0.0539	1.3680 ± 0.0427	0.7302 ± 0.0337
MLP	1	1.1553 ± 0.0563	1.5144 ± 0.0502	0.6697 ± 0.0385
	3	1.1454 ± 0.0487	1.5030 ± 0.0452	0.6740 ± 0.0430
	5	1.1570 ± 0.0518	1.5174 ± 0.0495	0.6672 ± 0.0479
	7	1.1396 ± 0.0452	1.5007 ± 0.0475	0.6744 ± 0.0475
	9	1.1359 ± 0.0360	1.4971 ± 0.0388	0.6761 ± 0.0454
	11	1.1478 ± 0.0557	1.5076 ± 0.0539	0.6724 ± 0.0414
	15	1.1426 ± 0.0544	1.4922 ± 0.0696	0.6795 ± 0.0387
	21	1.1199 ± 0.0399	1.4731 ± 0.0475	0.6876 ± 0.0347
	31	1.0931 ± 0.0441	1.4467 ± 0.0499	0.6984 ± 0.0368
SVR	1	1.1332 ± 0.0490	1.4996 ± 0.0485	0.6757 ± 0.0414
	3	1.1333 ± 0.0543	1.5029 ± 0.0598	0.6749 ± 0.0371
	5	1.1347 ± 0.0561	1.5041 ± 0.0614	0.6742 ± 0.0389
	7	1.1325 ± 0.0543	1.5044 ± 0.0598	0.6739 ± 0.0405
	9	1.1317 ± 0.0547	1.5026 ± 0.0630	0.6749 ± 0.0393
	11	1.1305 ± 0.0538	1.5002 ± 0.0587	0.6757 ± 0.0405
	15	1.1289 ± 0.0530	1.5017 ± 0.0577	0.6750 ± 0.0407
	21	1.1001 ± 0.0511	1.4671 ± 0.0543	0.6902 ± 0.0350
	31	1.0646 ± 0.0507	1.4248 ± 0.0482	0.7077 ± 0.0332
XGB	1	1.1418 ± 0.0485	1.4975 ± 0.0534	0.6776 ± 0.0326
	3	1.1457 ± 0.0472	1.5014 ± 0.0536	0.6760 ± 0.0327
	5	1.1437 ± 0.0502	1.4996 ± 0.0536	0.6768 ± 0.0319
	7	1.1464 ± 0.0503	1.5032 ± 0.0567	0.6754 ± 0.0304
	9	1.1424 ± 0.0486	1.4961 ± 0.0504	0.6781 ± 0.0332
	11	1.1446 ± 0.0487	1.5005 ± 0.0560	0.6763 ± 0.0333
	15	1.1461 ± 0.0507	1.4952 ± 0.0535	0.6787 ± 0.0316
	21	1.1383 ± 0.0528	1.4860 ± 0.0530	0.6826 ± 0.0318
	31	1.1214 ± 0.0544	1.4726 ± 0.0580	0.6884 ± 0.0301

Table A.10: Model Performance Across Increasing Input Time Windows (tw) for OL 1

Method	**Time Window (tw)**	**MAE**	**RMSE**	**R^2**
LR	1	1.1644 ± 0.0361	1.5245 ± 0.0663	0.6951 ± 0.0353
	3	1.1225 ± 0.0267	1.4621 ± 0.0483	0.7194 ± 0.0317
	5	1.1124 ± 0.0288	1.4445 ± 0.0459	0.7263 ± 0.0300
	7	1.1073 ± 0.0321	1.4367 ± 0.0482	0.7294 ± 0.0295
	9	1.1064 ± 0.0329	1.4347 ± 0.0507	0.7302 ± 0.0295
	11	1.1051 ± 0.0326	1.4335 ± 0.0518	0.7306 ± 0.0298
	13	1.1051 ± 0.0332	1.4333 ± 0.0531	0.7306 ± 0.0302
	15	1.1046 ± 0.0342	1.4322 ± 0.0552	0.7310 ± 0.0307
	17	1.1045 ± 0.0346	1.4320 ± 0.0565	0.7310 ± 0.0309
	19	1.1039 ± 0.0346	1.4314 ± 0.0566	0.7311 ± 0.0315
	21	1.1038 ± 0.0358	1.4309 ± 0.0575	0.7312 ± 0.0320
RIDGE	1	1.1656 ± 0.0360	1.5252 ± 0.0655	0.6948 ± 0.0350
	3	1.1240 ± 0.0259	1.4638 ± 0.0477	0.7188 ± 0.0312
	5	1.1143 ± 0.0284	1.4462 ± 0.0451	0.7258 ± 0.0293
	7	1.1076 ± 0.0320	1.4367 ± 0.0480	0.7294 ± 0.0293
	9	1.1068 ± 0.0330	1.4346 ± 0.0505	0.7302 ± 0.0292
	11	1.1058 ± 0.0326	1.4337 ± 0.0517	0.7305 ± 0.0295
	13	1.1057 ± 0.0333	1.4334 ± 0.0531	0.7306 ± 0.0299
	15	1.1049 ± 0.0343	1.4322 ± 0.0552	0.7310 ± 0.0305
	17	1.1048 ± 0.0346	1.4319 ± 0.0564	0.7311 ± 0.0307
	19	1.1042 ± 0.0346	1.4313 ± 0.0566	0.7311 ± 0.0313
	21	1.1040 ± 0.0356	1.4308 ± 0.0573	0.7313 ± 0.0317
MLP	1	1.1606 ± 0.0391	1.5196 ± 0.0729	0.6969 ± 0.0367
	3	1.1223 ± 0.0285	1.4572 ± 0.0511	0.7219 ± 0.0262
	5	1.0856 ± 0.0370	1.4087 ± 0.0631	0.7393 ± 0.0335
	7	1.0888 ± 0.0331	1.4092 ± 0.0540	0.7395 ± 0.0304
	9	1.0846 ± 0.0406	1.4006 ± 0.0587	0.7431 ± 0.0270
	11	1.0814 ± 0.0389	1.4001 ± 0.0639	0.7429 ± 0.0300
	13	1.0920 ± 0.0537	1.4117 ± 0.0820	0.7383 ± 0.0353
	15	1.1031 ± 0.0401	1.4234 ± 0.0663	0.7340 ± 0.0339
	17	1.0864 ± 0.0340	1.4063 ± 0.0588	0.7408 ± 0.0290
	19	1.0842 ± 0.0453	1.4102 ± 0.0682	0.7388 ± 0.0333
	21	1.0903 ± 0.0374	1.4121 ± 0.0673	0.7378 ± 0.0352
SVR	1	1.1551 ± 0.0417	1.5127 ± 0.0769	0.6995 ± 0.0382
	3	1.1039 ± 0.0350	1.4405 ± 0.0650	0.7273 ± 0.0353
	5	1.0859 ± 0.0334	1.4174 ± 0.0560	0.7362 ± 0.0322
	7	1.0734 ± 0.0337	1.3945 ± 0.0529	0.7446 ± 0.0314
	9	1.0795 ± 0.0330	1.4058 ± 0.0584	0.7404 ± 0.0330
	11	1.0721 ± 0.0366	1.3935 ± 0.0581	0.7451 ± 0.0316
	13	1.0681 ± 0.0380	1.3885 ± 0.0610	0.7468 ± 0.0323
	15	1.0684 ± 0.0408	1.3892 ± 0.0649	0.7465 ± 0.0330
	17	1.0640 ± 0.0409	1.3848 ± 0.0693	0.7480 ± 0.0339
	19	1.0631 ± 0.0368	1.3845 ± 0.0631	0.7481 ± 0.0327
	21	1.0609 ± 0.0423	1.3846 ± 0.0717	0.7480 ± 0.0340
XGB	1	1.1631 ± 0.0444	1.5164 ± 0.0745	0.6983 ± 0.0360
	3	1.1128 ± 0.0266	1.4462 ± 0.0552	0.7255 ± 0.0312
	5	1.0922 ± 0.0355	1.4148 ± 0.0573	0.7376 ± 0.0293
	7	1.0811 ± 0.0395	1.4003 ± 0.0642	0.7429 ± 0.0295
	9	1.0826 ± 0.0417	1.3992 ± 0.0649	0.7434 ± 0.0291
	11	1.0748 ± 0.0415	1.3908 ± 0.0645	0.7464 ± 0.0294
	13	1.0770 ± 0.0411	1.3942 ± 0.0653	0.7452 ± 0.0289
	15	1.0727 ± 0.0415	1.3876 ± 0.0665	0.7475 ± 0.0300
	17	1.0687 ± 0.0399	1.3849 ± 0.0659	0.7486 ± 0.0287
	19	1.0693 ± 0.0431	1.3887 ± 0.0676	0.7472 ± 0.0289
	21	1.0665 ± 0.0398	1.3816 ± 0.0632	0.7496 ± 0.0293

Table A.11: Model Performance Across Increasing Input Time Windows (tw) for OL 2

Method	**Time Window (tw)**	**MAE**	**RMSE**	**R^2**
LR	1	1.2199 ± 0.0950	1.5690 ± 0.1125	0.6391 ± 0.0471
	3	1.1533 ± 0.0493	1.4770 ± 0.0552	0.6795 ± 0.0392
	5	1.1319 ± 0.0369	1.4449 ± 0.0434	0.6930 ± 0.0397
	7	1.1240 ± 0.0352	1.4333 ± 0.0440	0.6979 ± 0.0392
	9	1.1204 ± 0.0380	1.4300 ± 0.0465	0.6994 ± 0.0384
	11	1.1190 ± 0.0410	1.4277 ± 0.0484	0.7003 ± 0.0382
	13	1.1176 ± 0.0426	1.4268 ± 0.0507	0.7007 ± 0.0388
	15	1.1179 ± 0.0430	1.4272 ± 0.0519	0.7005 ± 0.0395
	17	1.1182 ± 0.0443	1.4276 ± 0.0531	0.7000 ± 0.0401
	19	1.1186 ± 0.0458	1.4280 ± 0.0549	0.6999 ± 0.0403
	21	1.1179 ± 0.0464	1.4281 ± 0.0565	0.6999 ± 0.0402
RIDGE	1	1.2199 ± 0.0950	1.5690 ± 0.1124	0.6391 ± 0.0471
	3	1.1551 ± 0.0520	1.4789 ± 0.0579	0.6788 ± 0.0390
	5	1.1344 ± 0.0402	1.4474 ± 0.0453	0.6921 ± 0.0384
	7	1.1249 ± 0.0362	1.4336 ± 0.0441	0.6979 ± 0.0383
	9	1.1215 ± 0.0387	1.4299 ± 0.0465	0.6995 ± 0.0377
	11	1.1198 ± 0.0417	1.4276 ± 0.0485	0.7004 ± 0.0375
	13	1.1186 ± 0.0435	1.4267 ± 0.0508	0.7009 ± 0.0379
	15	1.1191 ± 0.0445	1.4273 ± 0.0524	0.7005 ± 0.0386
	17	1.1191 ± 0.0451	1.4273 ± 0.0533	0.7003 ± 0.0391
	19	1.1194 ± 0.0462	1.4275 ± 0.0546	0.7002 ± 0.0392
	21	1.1189 ± 0.0474	1.4275 ± 0.0564	0.7003 ± 0.0391
MLP	1	1.2113 ± 0.0905	1.5588 ± 0.1112	0.6439 ± 0.0450
	3	1.1424 ± 0.0548	1.4610 ± 0.0648	0.6860 ± 0.0430
	5	1.1247 ± 0.0537	1.4293 ± 0.0665	0.6992 ± 0.0437
	7	1.0962 ± 0.0401	1.3982 ± 0.0552	0.7121 ± 0.0415
	9	1.0947 ± 0.0359	1.3986 ± 0.0524	0.7121 ± 0.0395
	11	1.0910 ± 0.0326	1.3920 ± 0.0519	0.7149 ± 0.0387
	13	1.0927 ± 0.0360	1.3953 ± 0.0569	0.7127 ± 0.0457
	15	1.1012 ± 0.0416	1.3963 ± 0.0603	0.7130 ± 0.0404
	17	1.0912 ± 0.0292	1.3937 ± 0.0514	0.7138 ± 0.0404
	19	1.0871 ± 0.0486	1.3914 ± 0.0707	0.7145 ± 0.0440
	21	1.0891 ± 0.0281	1.3898 ± 0.0471	0.7158 ± 0.0368
SVR	1	1.2137 ± 0.0981	1.5642 ± 0.1198	0.6411 ± 0.0499
	3	1.1453 ± 0.0469	1.4661 ± 0.0582	0.6839 ± 0.0419
	5	1.1119 ± 0.0340	1.4201 ± 0.0528	0.7027 ± 0.0445
	7	1.0954 ± 0.0307	1.3984 ± 0.0499	0.7119 ± 0.0418
	9	1.0894 ± 0.0280	1.3918 ± 0.0477	0.7145 ± 0.0425
	11	1.0807 ± 0.0316	1.3802 ± 0.0509	0.7194 ± 0.0402
	13	1.0841 ± 0.0366	1.3843 ± 0.0528	0.7178 ± 0.0409
	15	1.0816 ± 0.0391	1.3798 ± 0.0564	0.7196 ± 0.0405
	17	1.0733 ± 0.0364	1.3728 ± 0.0571	0.7220 ± 0.0423
	19	1.0735 ± 0.0354	1.3744 ± 0.0568	0.7214 ± 0.0418
	21	1.0730 ± 0.0407	1.3764 ± 0.0608	0.7207 ± 0.0415
XGB	1	1.2203 ± 0.0972	1.5656 ± 0.1181	0.6407 ± 0.0475
	3	1.1552 ± 0.0590	1.4749 ± 0.0677	0.6806 ± 0.0391
	5	1.1214 ± 0.0298	1.4274 ± 0.0421	0.7003 ± 0.0388
	7	1.1089 ± 0.0221	1.4078 ± 0.0405	0.7085 ± 0.0376
	9	1.0941 ± 0.0327	1.3927 ± 0.0512	0.7147 ± 0.0380
	11	1.0928 ± 0.0351	1.3911 ± 0.0496	0.7155 ± 0.0359
	13	1.0846 ± 0.0349	1.3802 ± 0.0522	0.7197 ± 0.0384
	15	1.0832 ± 0.0340	1.3766 ± 0.0528	0.7211 ± 0.0384
	17	1.0791 ± 0.0345	1.3728 ± 0.0541	0.7224 ± 0.0390
	19	1.0839 ± 0.0388	1.3806 ± 0.0576	0.7191 ± 0.0406
	21	1.0817 ± 0.0384	1.3769 ± 0.0559	0.7206 ± 0.0408

A.5. Joint Multi-Location Predicting

See Table A.13

A.6. Autoencoded Windspeed-Field

See Tables A.14, A.15 and A.16.

A.7. Combining Strategies

See Tables A.17.

Table A.12: Model Performance Across Increasing Input Time Windows (tw) for OL 3

Method	Time Window (tw)	MAE	RMSE	R^2
LR	1	1.1418 ± 0.0495	1.5003 ± 0.0474	0.6755 ± 0.0408
	3	1.0918 ± 0.0501	1.4373 ± 0.0513	0.7020 ± 0.0405
	5	1.0861 ± 0.0528	1.4212 ± 0.0606	0.7086 ± 0.0423
	7	1.0841 ± 0.0513	1.4145 ± 0.0607	0.7112 ± 0.0436
	9	1.0843 ± 0.0523	1.4129 ± 0.0637	0.7117 ± 0.0456
	11	1.0846 ± 0.0542	1.4122 ± 0.0687	0.7118 ± 0.0480
	13	1.0858 ± 0.0532	1.4128 ± 0.0698	0.7114 ± 0.0489
	15	1.0852 ± 0.0552	1.4125 ± 0.0711	0.7115 ± 0.0490
	17	1.0862 ± 0.0556	1.4137 ± 0.0715	0.7107 ± 0.0491
	19	1.0872 ± 0.0555	1.4146 ± 0.0713	0.7099 ± 0.0493
	21	1.0878 ± 0.0546	1.4159 ± 0.0710	0.7095 ± 0.0489
RIDGE	1	1.1419 ± 0.0494	1.5004 ± 0.0473	0.6754 ± 0.0408
	3	1.0925 ± 0.0506	1.4379 ± 0.0514	0.7018 ± 0.0401
	5	1.0877 ± 0.0518	1.4221 ± 0.0594	0.7083 ± 0.0418
	7	1.0851 ± 0.0511	1.4147 ± 0.0604	0.7112 ± 0.0430
	9	1.0855 ± 0.0518	1.4129 ± 0.0631	0.7118 ± 0.0447
	11	1.0860 ± 0.0539	1.4124 ± 0.0681	0.7118 ± 0.0471
	13	1.0873 ± 0.0521	1.4129 ± 0.0691	0.7115 ± 0.0480
	15	1.0866 ± 0.0542	1.4125 ± 0.0704	0.7116 ± 0.0481
	17	1.0877 ± 0.0545	1.4135 ± 0.0706	0.7109 ± 0.0480
	19	1.0889 ± 0.0548	1.4147 ± 0.0706	0.7100 ± 0.0482
	21	1.0901 ± 0.0542	1.4160 ± 0.0703	0.7096 ± 0.0478
MLP	1	1.1553 ± 0.0563	1.5144 ± 0.0502	0.6697 ± 0.0385
	3	1.0736 ± 0.0394	1.4165 ± 0.0454	0.7108 ± 0.0363
	5	1.0694 ± 0.0596	1.4133 ± 0.0820	0.7121 ± 0.0437
	7	1.0657 ± 0.0410	1.4010 ± 0.0479	0.7167 ± 0.0410
	9	1.0676 ± 0.0509	1.4077 ± 0.0696	0.7124 ± 0.0552
	11	1.0630 ± 0.0422	1.3861 ± 0.0543	0.7222 ± 0.0451
	13	1.0684 ± 0.0576	1.3996 ± 0.0678	0.7167 ± 0.0479
	15	1.0636 ± 0.0499	1.3863 ± 0.0656	0.7217 ± 0.0488
	17	1.0639 ± 0.0309	1.3889 ± 0.0394	0.7201 ± 0.0464
	19	1.0713 ± 0.0559	1.3982 ± 0.0642	0.7164 ± 0.0486
	21	1.0789 ± 0.0514	1.4100 ± 0.0583	0.7114 ± 0.0491
SVR	1	1.1332 ± 0.0490	1.4996 ± 0.0485	0.6757 ± 0.0414
	3	1.0773 ± 0.0569	1.4268 ± 0.0630	0.7064 ± 0.0399
	5	1.0550 ± 0.0536	1.4016 ± 0.0715	0.7168 ± 0.0398
	7	1.0459 ± 0.0496	1.3848 ± 0.0651	0.7232 ± 0.0415
	9	1.0385 ± 0.0480	1.3724 ± 0.0586	0.7279 ± 0.0423
	11	1.0457 ± 0.0532	1.3805 ± 0.0675	0.7244 ± 0.0458
	13	1.0474 ± 0.0541	1.3860 ± 0.0688	0.7220 ± 0.0471
	15	1.0511 ± 0.0603	1.3866 ± 0.0761	0.7216 ± 0.0484
	17	1.0478 ± 0.0516	1.3782 ± 0.0664	0.7243 ± 0.0491
	19	1.0478 ± 0.0535	1.3795 ± 0.0682	0.7231 ± 0.0511
	21	1.0579 ± 0.0484	1.3920 ± 0.0661	0.7181 ± 0.0523
XGB	1	1.1418 ± 0.0485	1.4975 ± 0.0534	0.6776 ± 0.0326
	3	1.0863 ± 0.0541	1.4356 ± 0.0713	0.7037 ± 0.0337
	5	1.0710 ± 0.0528	1.4117 ± 0.0674	0.7137 ± 0.0313
	7	1.0585 ± 0.0459	1.3983 ± 0.0655	0.7190 ± 0.0316
	9	1.0617 ± 0.0469	1.3947 ± 0.0680	0.7207 ± 0.0307
	11	1.0571 ± 0.0491	1.3877 ± 0.0630	0.7230 ± 0.0343
	13	1.0609 ± 0.0449	1.3902 ± 0.0610	0.7224 ± 0.0299
	15	1.0559 ± 0.0497	1.3809 ± 0.0709	0.7259 ± 0.0325
	17	1.0552 ± 0.0456	1.3798 ± 0.0661	0.7253 ± 0.0382
	19	1.0587 ± 0.0471	1.3850 ± 0.0695	0.7233 ± 0.0355
	21	1.0576 ± 0.0412	1.3823 ± 0.0640	0.7235 ± 0.0419

Table A.13: Joint Multi-Location Prediction Performance (Single Model Trained on OL 1, 2 and 3).

OL	**Method**	**MAE**	**RMSE**	**R^2**
1	LR	1.1702 ± 0.0611	1.5260 ± 0.0836	0.6975 ± 0.0274
	Ridge	1.1713 ± 0.0614	1.5268 ± 0.0837	0.6972 ± 0.0274
	MLP	1.1489 ± 0.0600	1.5063 ± 0.0886	0.7051 ± 0.0296
	SVR	1.1564 ± 0.0524	1.5177 ± 0.0803	0.7009 ± 0.0256
	XGB	1.1621 ± 0.0603	1.5152 ± 0.0891	0.7017 ± 0.0289
2	LR	1.2337 ± 0.0973	1.5876 ± 0.1189	0.6359 ± 0.0394
	Ridge	1.2340 ± 0.0969	1.5874 ± 0.1182	0.6360 ± 0.0390
	MLP	1.2344 ± 0.1062	1.5850 ± 0.1317	0.6369 ± 0.0455
	SVR	1.2313 ± 0.0970	1.5872 ± 0.1196	0.6360 ± 0.0404
	XGB	1.2316 ± 0.1010	1.5806 ± 0.1234	0.6390 ± 0.0417
3	LR	1.1393 ± 0.0409	1.5056 ± 0.0634	0.6823 ± 0.0169
	Ridge	1.1392 ± 0.0413	1.5051 ± 0.0636	0.6825 ± 0.0170
	MLP	1.1383 ± 0.0324	1.4914 ± 0.0536	0.6885 ± 0.0058
	SVR	1.1327 ± 0.0504	1.5012 ± 0.0764	0.6840 ± 0.0229
	XGB	1.1331 ± 0.0290	1.4898 ± 0.0494	0.6890 ± 0.0102

Table A.14: Model Performance Metrics for OL 1 using CAE Encoding

Model	**LD**	**MAE**	**RMSE**	R^2
lr	*None*	1.164 ± (0.04)	1.524 ± (0.07)	0.695 ± (0.04)
	8	1.131 ± (0.04)	1.481 ± (0.06)	0.712 ± (0.03)
	16	1.101 ± (0.03)	1.442 ± (0.06)	0.727 ± (0.03)
	32	1.097 ± (0.04)	1.436 ± (0.06)	0.730 ± (0.03)
	64	1.089 ± (0.03)	1.424 ± (0.06)	0.735 ± (0.03)
	128	1.074 ± (0.03)	1.409 ± (0.06)	0.740 ± (0.03)
ridge	*None*	1.166 ± (0.04)	1.525 ± (0.07)	0.695 ± (0.04)
	8	1.132 ± (0.03)	1.482 ± (0.06)	0.712 ± (0.03)
	16	1.103 ± (0.03)	1.442 ± (0.05)	0.727 ± (0.03)
	32	1.098 ± (0.03)	1.436 ± (0.06)	0.730 ± (0.03)
	64	1.090 ± (0.03)	1.424 ± (0.06)	0.734 ± (0.03)
	128	1.074 ± (0.03)	1.409 ± (0.06)	0.740 ± (0.03)
mlp	*None*	1.161 ± (0.04)	1.520 ± (0.07)	0.697 ± (0.04)
	8	1.063 ± (0.06)	1.396 ± (0.09)	0.744 ± (0.04)
	16	1.065 ± (0.04)	1.393 ± (0.07)	0.745 ± (0.03)
	32	1.067 ± (0.04)	1.396 ± (0.08)	0.744 ± (0.03)
	64	1.069 ± (0.03)	1.396 ± (0.06)	0.745 ± (0.03)
	128	1.083 ± (0.03)	1.420 ± (0.07)	0.735 ± (0.03)
svr	*None*	1.155 ± (0.04)	1.513 ± (0.08)	0.700 ± (0.04)
	8	1.067 ± (0.03)	1.408 ± (0.06)	0.740 ± (0.03)
	16	1.062 ± (0.03)	1.393 ± (0.07)	0.745 ± (0.03)
	32	1.052 ± (0.03)	1.385 ± (0.06)	0.748 ± (0.03)
	64	1.063 ± (0.04)	1.396 ± (0.08)	0.745 ± (0.03)
	128	1.061 ± (0.03)	1.399 ± (0.07)	0.743 ± (0.03)
xgb	*None*	1.163 ± (0.04)	1.516 ± (0.07)	0.698 ± (0.04)
	8	1.055 ± (0.04)	1.384 ± (0.07)	0.749 ± (0.03)
	16	1.047 ± (0.03)	1.378 ± (0.06)	0.751 ± (0.03)
	32	1.039 ± (0.02)	1.367 ± (0.06)	0.755 ± (0.03)
	64	1.039 ± (0.03)	1.363 ± (0.07)	0.756 ± (0.03)
	128	**1.035** ± (0.03)	**1.358** ± (0.06)	**0.758** ± (0.03)

Table A.15: Model Performance Metrics for OL 2 using CAE Encoding

Model	**LD**	**MAE**	**RMSE**	R^2
lr	*None*	1.220 ± (0.10)	1.569 ± (0.11)	0.639 ± (0.05)
	8	1.179 ± (0.10)	1.523 ± (0.13)	0.660 ± (0.05)
	16	1.156 ± (0.11)	1.489 ± (0.15)	0.674 ± (0.06)
	32	1.151 ± (0.11)	1.485 ± (0.14)	0.676 ± (0.05)
	64	1.139 ± (0.11)	1.470 ± (0.14)	0.683 ± (0.05)
	128	1.127 ± (0.11)	1.457 ± (0.14)	0.688 ± (0.05)
ridge	*None*	1.220 ± (0.10)	1.569 ± (0.11)	0.639 ± (0.05)
	8	1.179 ± (0.10)	1.523 ± (0.13)	0.660 ± (0.05)
	16	1.157 ± (0.11)	1.490 ± (0.15)	0.674 ± (0.05)
	32	1.152 ± (0.11)	1.485 ± (0.14)	0.676 ± (0.05)
	64	1.140 ± (0.11)	1.470 ± (0.14)	0.683 ± (0.05)
	128	1.128 ± (0.11)	1.458 ± (0.14)	0.688 ± (0.05)
mlp	*None*	1.211 ± (0.09)	1.559 ± (0.11)	0.644 ± (0.05)
	8	1.111 ± (0.11)	1.444 ± (0.15)	0.693 ± (0.06)
	16	1.123 ± (0.10)	1.453 ± (0.13)	0.690 ± (0.05)
	32	1.136 ± (0.10)	1.472 ± (0.13)	0.681 ± (0.05)
	64	1.105 ± (0.11)	1.434 ± (0.14)	0.698 ± (0.05)
	128	1.124 ± (0.12)	1.457 ± (0.15)	0.688 ± (0.06)
svr	*None*	1.214 ± (0.10)	1.564 ± (0.12)	0.641 ± (0.05)
	8	1.114 ± (0.09)	1.454 ± (0.12)	0.690 ± (0.05)
	16	1.117 ± (0.11)	1.447 ± (0.14)	0.693 ± (0.05)
	32	1.110 ± (0.11)	1.444 ± (0.14)	0.694 ± (0.05)
	64	1.117 ± (0.12)	1.451 ± (0.16)	0.691 ± (0.06)
	128	1.116 ± (0.11)	1.450 ± (0.15)	0.691 ± (0.05)
xgb	*None*	1.220 ± (0.10)	1.566 ± (0.12)	0.641 ± (0.05)
	8	1.099 ± (0.09)	1.425 ± (0.12)	0.702 ± (0.05)
	16	1.098 ± (0.10)	1.423 ± (0.13)	0.702 ± (0.05)
	32	1.093 ± (0.09)	1.417 ± (0.13)	0.705 ± (0.05)
	64	1.101 ± (0.10)	1.423 ± (0.14)	0.702 ± (0.05)
	128	**1.085** ± (0.09)	**1.410** ± (0.12)	**0.708** ± (0.05)

Table A.16: Model Performance Metrics for OL 3 using CAE Encoding

Model	LD	MAE	RMSE	R^2
lr	*None*	1.142 ± (0.05)	1.500 ± (0.05)	0.675 ± (0.04)
	8	1.111 ± (0.05)	1.464 ± (0.05)	0.692 ± (0.03)
	16	1.093 ± (0.04)	1.443 ± (0.05)	0.700 ± (0.04)
	32	1.081 ± (0.04)	1.433 ± (0.05)	0.704 ± (0.04)
	64	1.084 ± (0.04)	1.425 ± (0.04)	0.707 ± (0.03)
	128	1.062 ± (0.04)	1.406 ± (0.04)	0.715 ± (0.03)
ridge	*None*	1.142 ± (0.05)	1.500 ± (0.05)	0.675 ± (0.04)
	8	1.111 ± (0.05)	1.464 ± (0.05)	0.692 ± (0.03)
	16	1.095 ± (0.04)	1.443 ± (0.05)	0.700 ± (0.04)
	32	1.081 ± (0.03)	1.431 ± (0.05)	0.705 ± (0.04)
	64	1.084 ± (0.04)	1.425 ± (0.04)	0.707 ± (0.03)
	128	1.063 ± (0.04)	1.406 ± (0.04)	0.715 ± (0.03)
mlp	*None*	1.155 ± (0.06)	1.514 ± (0.05)	0.670 ± (0.04)
	8	1.056 ± (0.05)	1.402 ± (0.06)	0.718 ± (0.03)
	16	1.075 ± (0.03)	1.414 ± (0.05)	0.712 ± (0.03)
	32	1.069 ± (0.04)	1.414 ± (0.05)	0.713 ± (0.03)
	64	1.076 ± (0.04)	1.415 ± (0.05)	0.712 ± (0.03)
	128	1.086 ± (0.05)	1.430 ± (0.04)	0.705 ± (0.03)
svr	*None*	1.133 ± (0.05)	1.500 ± (0.05)	0.676 ± (0.04)
	8	1.059 ± (0.05)	1.411 ± (0.05)	0.714 ± (0.03)
	16	1.077 ± (0.05)	1.428 ± (0.06)	0.707 ± (0.03)
	32	1.079 ± (0.05)	1.432 ± (0.06)	0.706 ± (0.03)
	64	1.068 ± (0.04)	1.409 ± (0.05)	0.714 ± (0.03)
	128	**1.048** ± (0.04)	1.399 ± (0.04)	0.718 ± (0.03)
xgb	*None*	1.142 ± (0.05)	1.498 ± (0.05)	0.678 ± (0.03)
	8	1.061 ± (0.06)	1.413 ± (0.07)	0.714 ± (0.02)
	16	1.065 ± (0.05)	1.408 ± (0.06)	0.716 ± (0.02)
	32	1.056 ± (0.05)	1.403 ± (0.05)	0.717 ± (0.02)
	64	1.065 ± (0.06)	1.413 ± (0.07)	0.714 ± (0.02)
	128	1.053 ± (0.05)	**1.399** ± (0.05)	**0.719** ± (0.03)

Table A.17: Performance of Combined Strategies (CS) across all Observation Locations (OL). CS1 is applied to all models, while CS2 is applied only to LR and Ridge.

			MAE		RMSE		R^2	
OL	**Method**	**Strat.**	**Value**	**Imp.%**	**Value**	**Imp.%**	**Value**	**Imp.%**
OL 1	**LR**	CS1	1.0293 ± 0.0368	11.60	1.3356 ± 0.0579	12.39	0.7656 ± 0.0309	10.14
		CS2	1.1757 ± 0.0377	-0.98	1.5353 ± 0.0655	-0.71	0.6910 ± 0.0372	-0.58
	Ridge	CS1	1.0264 ± 0.0328	11.94	1.3322 ± 0.0545	12.65	0.7670 ± 0.0288	10.39
		CS2	0.9983 ± 0.0302	14.35	1.3063 ± 0.0438	14.35	0.7760 ± 0.0272	11.68
	MLP	CS1	0.9942 ± 0.0332	14.34	1.2954 ± 0.0545	14.76	0.7797 ± 0.0278	11.88
	SVR	CS1	0.9892 ± 0.0394	14.36	1.2911 ± 0.0689	14.65	0.7802 ± 0.0361	11.53
	XGB	CS1	0.9624 ± 0.0346	17.25	1.2563 ± 0.0581	17.16	0.7927 ± 0.0278	13.52
OL 2	**LR**	CS1	1.0453 ± 0.0280	14.31	1.3330 ± 0.0472	15.04	0.7384 ± 0.0368	15.54
		CS2	1.1913 ± 0.0515	2.34	1.5340 ± 0.0729	2.23	0.6550 ± 0.0390	2.50
	Ridge	CS1	1.0409 ± 0.0221	14.67	1.3278 ± 0.0405	15.37	0.7407 ± 0.0339	15.90
		CS2	1.0184 ± 0.0256	16.52	1.3082 ± 0.0420	16.62	0.7481 ± 0.0343	17.06
	MLP	CS1	1.0143 ± 0.0350	16.26	1.3026 ± 0.0608	16.44	0.7506 ± 0.0326	16.56
	SVR	CS1	1.0039 ± 0.0257	17.29	1.2896 ± 0.0492	17.55	0.7549 ± 0.0358	17.76
	XGB	CS1	0.9778 ± 0.0324	19.87	1.2585 ± 0.0484	19.62	0.7662 ± 0.0371	19.59
OL 3	**LR**	CS1	1.0280 ± 0.0656	9.97	1.3509 ± 0.0740	9.96	0.7374 ± 0.0346	9.17
		CS2	1.3881 ± 0.1073	-21.57	1.8078 ± 0.1605	-20.50	0.5286 ± 0.0898	-21.74
	Ridge	CS1	1.0243 ± 0.0602	10.30	1.3446 ± 0.0669	10.38	0.7404 ± 0.0289	9.61
		CS2	1.0046 ± 0.0494	12.03	1.3242 ± 0.0619	11.74	0.7468 ± 0.0390	10.56
	MLP	CS1	0.9996 ± 0.0506	13.48	1.3127 ± 0.0408	13.32	0.7514 ± 0.0343	12.21
	SVR	CS1	0.9912 ± 0.0741	12.53	1.3060 ± 0.0777	12.91	0.7545 ± 0.0331	11.67
	XGB	CS1	0.9744 ± 0.0758	14.66	1.2854 ± 0.0894	14.16	0.7622 ± 0.0344	12.48
OL 4	**LR**	CS1	0.9610	9.89	1.2752	9.18	0.6907	10.33
	Ridge	CS1	0.9528	10.66	1.2635	10.01	0.6966	11.27
	MLP	CS1	0.9517	10.56	1.2678	9.43	0.6972	10.87
	SVR	CS1	0.9309	13.05	1.2394	12.21	0.7076	13.68
	XGB	CS1	0.9269	13.39	1.2431	12.05	0.7080	13.65
OL 5	**LR**	CS1	0.9923	6.29	1.2916	6.09	0.7805	4.02
	Ridge	CS1	0.9869	6.80	1.2854	6.54	0.7827	4.32
	MLP	CS1	0.9651	8.67	1.2608	8.12	0.7912	5.29
	SVR	CS1	0.9618	8.99	1.2591	8.28	0.7915	5.40
	XGB	CS1	0.9506	10.62	1.2409	10.43	0.7980	6.67
OL 6	**LR**	CS1	0.9928	6.90	1.2842	6.31	0.7855	4.14
	Ridge	CS1	0.9866	7.50	1.2773	6.80	0.7884	4.50
	MLP	CS1	0.9765	7.55	1.2604	7.44	0.7937	4.80
	SVR	CS1	0.9602	9.26	1.2435	8.82	0.7984	5.48
	XGB	CS1	0.9330	12.52	1.2156	11.46	0.8073	7.09

B

Experiment Results for Spatial Extrapolation

B.1. Wind Speed enhanced Inverse Distance Weighting

See Table B.1.

B.2. Wind Direction enhanced Inverse Distance Weighting

See Table B.2.

B.3. Elevation enhanced Inverse Distance Weighting

See Table B.3.

B.4. Tree Canopy Height enhanced Inverse Distance Weighting

See Table B.4.

Table B.1: Detailed Results for WS-IDW with varying α for the 2 SE mast scenario on Site 1 and on Site 2, and the 3 SE mast scenario on Site 2.

Site 2, 2 Masts			**Site 2, 3 Masts**			**Site 1, 2 Masts**		
α	**MAE**	**RMSE**	α	**MAE**	**RMSE**	α	**MAE**	**RMSE**
0.00	0.8202	1.0963	0.00	0.7679	1.0268	0.00	1.1038	1.4734
0.10	0.8201	1.0962	0.10	0.7675	1.0263	0.10	1.1037	1.4732
0.20	0.8200	1.0961	0.20	0.7671	1.0257	0.20	1.1035	1.4730
0.30	0.8198	1.0959	0.30	0.7665	1.0249	0.30	1.1034	1.4728
0.40	0.8196	1.0956	0.40	0.7658	1.0238	0.40	1.1031	1.4725
0.50	0.8193	1.0952	0.50	0.7648	1.0224	0.50	1.1028	1.4720
0.60	0.8189	1.0947	0.60	0.7633	1.0203	0.60	1.1023	1.4713
0.70	0.8181	1.0938	0.70	0.7610	1.0171	0.70	1.1016	1.4702
0.80	0.8167	1.0920	0.80	0.7567	1.0113	0.80	1.1001	1.4681
0.90	0.8127	1.0874	0.90	0.7467	0.9987	0.90	1.0961	1.4626
0.91	0.8119	1.0865	0.91	0.7449	0.9965	0.91	1.0953	1.4615
0.92	0.8109	1.0854	0.92	0.7429	0.9941	0.92	1.0944	1.4603
0.93	0.8098	1.0841	0.93	0.7406	0.9914	0.93	1.0931	1.4587
0.94	0.8083	1.0825	0.94	0.7380	0.9883	0.94	1.0916	1.4568
0.95	0.8064	1.0806	0.95	0.7350	0.9848	0.95	1.0897	1.4544
0.96	0.8040	1.0783	0.96	0.7316	0.9810	0.96	1.0871	1.4513
0.97	0.8008	1.0756	0.97	0.7276	0.9773	0.97	1.0835	1.4472
0.975	0.7990	1.0743	0.975	0.7246	0.9750	0.975	1.0812	1.4447
0.980	0.7970	1.0733	0.980	0.7244	0.9750	0.980	1.0785	1.4418
0.982	0.7964	1.0732	0.982	0.7240	0.9751	0.982	1.0772	1.4405
0.984	0.7959	1.0733	0.984	0.7241	0.9755	0.984	1.0759	1.4392
0.986	0.7957	1.0738	0.986	0.7247	0.9766	0.986	1.0745	1.4379
0.988	0.7959	1.0748	0.988	0.7262	0.9784	0.988	1.0730	1.4365
0.990	0.7969	1.0767	0.990	0.7290	0.9815	0.990	1.0714	1.4353
0.992	0.7992	1.0799	0.992	0.7333	0.9866	0.992	1.0701	1.4342
0.994	0.8033	1.0852	0.994	0.7398	0.9949	0.994	1.0692	1.4337
0.996	0.8100	1.0940	0.996	0.7504	1.0091	0.996	1.0688	1.4342
0.998	0.8224	1.1091	0.998	0.7691	1.0348	0.998	1.0708	1.4375
1.00	0.8466	1.1376	1.00	0.8078	1.0896	1.00	1.0781	1.4486

Table B.2: Detailed Results for WD-IDW with varying α for the 2 SE mast scenario on Site 1 and on Site 2, and the 3 SE mast scenario on Site 2.

Site 2, 2 Masts

α	MAE	RMSE
0.00	0.8202	1.0963
0.05	0.8193	1.0955
0.10	0.8185	1.0948
0.15	0.8176	1.0940
0.20	0.8168	1.0932
0.25	0.8160	1.0925
0.30	0.8152	1.0919
0.35	0.8144	1.0913
0.40	0.8137	1.0909
0.45	0.8132	1.0906
0.50	0.8129	1.0904
0.55	0.8127	1.0905
0.60	0.8126	1.0908
0.65	0.8129	1.0915
0.70	0.8138	1.0928
0.75	0.8153	1.0948
0.80	0.8179	1.0980
0.85	0.8223	1.1033
0.90	0.8293	1.1123
0.95	0.8423	1.1288
1.00	0.8713	1.1648

Site 2, 3 Masts

α	MAE	RMSE
0.00	0.7679	1.0268
0.05	0.7667	1.0258
0.10	0.7657	1.0249
0.15	0.7649	1.0243
0.20	0.7644	1.0240
0.25	0.7640	1.0239
0.30	0.7639	1.0241
0.35	0.7641	1.0246
0.40	0.7648	1.0256
0.45	0.7658	1.0270
0.50	0.7671	1.0289
0.55	0.7689	1.0315
0.60	0.7712	1.0349
0.65	0.7741	1.0392
0.70	0.7777	1.0448
0.75	0.7825	1.0520
0.80	0.7888	1.0614
0.85	0.7969	1.0738
0.90	0.8078	1.0907
0.95	0.8249	1.1155
1.00	0.8626	1.1666

Site 1, 2 Masts

α	MAE	RMSE
0.00	1.1038	1.4734
0.05	1.1034	1.4727
0.10	1.1031	1.4721
0.15	1.1029	1.4716
0.20	1.1027	1.4712
0.25	1.1026	1.4708
0.30	1.1024	1.4705
0.35	1.1021	1.4701
0.40	1.1018	1.4697
0.45	1.1015	1.4692
0.50	1.1011	1.4688
0.55	1.1006	1.4682
0.60	1.1001	1.4676
0.65	1.0994	1.4668
0.70	1.0986	1.4660
0.75	1.0975	1.4650
0.80	1.0961	1.4638
0.85	1.0943	1.4624
0.90	1.0920	1.4610
0.91	1.0916	1.4608
0.92	1.0911	1.4607
0.93	1.0907	1.4606
0.94	1.0905	1.4607
0.95	1.0905	1.4611
0.96	1.0908	1.4618
0.97	1.0913	1.4633
0.98	1.0926	1.4662
0.99	1.0958	1.4719
1.00	1.1069	1.4876

Table B.3: Detailed Results for Elevation-IDW with varying α for the 2 SE mast and 3 SE mast scenarios on Site 2.

Site 2, 2 Masts

α	MAE	RMSE
0.00	0.8202	1.0963
0.05	0.8199	1.0960
0.10	0.8196	1.0957
0.15	0.8193	1.0953
0.20	0.8191	1.0949
0.25	0.8188	1.0946
0.30	0.8185	1.0942
0.35	0.8183	1.0939
0.40	0.8182	1.0936
0.45	0.8181	1.0934
0.50	0.8181	1.0934
0.55	0.8183	1.0936
0.60	0.8190	1.0943
0.65	0.8201	1.0958
0.70	0.8223	1.0984
0.75	0.8261	1.1031
0.80	0.8329	1.1112
0.85	0.8444	1.1252
0.90	0.8628	1.1485
0.95	0.8881	1.1836
1.00	0.9115	1.2205

Site 2, 3 Masts

α	MAE	RMSE
0.00	0.7679	1.0268
0.05	0.7678	1.0263
0.10	0.7678	1.0258
0.15	0.7679	1.0254
0.20	0.7681	1.0251
0.25	0.7686	1.0249
0.30	0.7693	1.0250
0.35	0.7702	1.0252
0.40	0.7712	1.0257
0.45	0.7724	1.0265
0.50	0.7740	1.0277
0.55	0.7760	1.0293
0.60	0.7784	1.0314
0.65	0.7813	1.0342
0.70	0.7850	1.0377
0.75	0.7896	1.0424
0.80	0.7957	1.0488
0.85	0.8050	1.0585
0.90	0.8199	1.0748
0.95	0.8471	1.1065
1.00	0.9092	1.1913

Table B.4: Detailed Results for TCH-IDW with varying α for the 2 SE mast and 3 SE mast scenarios on Site 2.

Site 2, 2 Masts

α	MAE	RMSE
0.00	0.8202	1.0963
0.05	0.8201	1.0963
0.10	0.8201	1.0962
0.15	0.8200	1.0962
0.20	0.8199	1.0961
0.25	0.8198	1.0960
0.30	0.8198	1.0959
0.35	0.8196	1.0958
0.40	0.8195	1.0957
0.45	0.8194	1.0955
0.50	0.8192	1.0954
0.55	0.8190	1.0951
0.60	0.8188	1.0949
0.65	0.8185	1.0946
0.70	0.8181	1.0942
0.75	0.8176	1.0938
0.80	0.8170	1.0932
0.85	0.8162	1.0925
0.90	0.8154	1.0920
0.91	0.8154	1.0921
0.92	0.8155	1.0923
0.93	0.8159	1.0928
0.94	0.8165	1.0936
0.95	0.8176	1.0952
0.96	0.8197	1.0979
0.97	0.8240	1.1031
0.98	0.8335	1.1138
0.99	0.8574	1.1402
1.00	0.9155	1.2145

Site 2, 3 Masts

α	MAE	RMSE
0.00	0.7679	1.0268
0.05	0.7679	1.0268
0.10	0.7678	1.0267
0.15	0.7678	1.0267
0.20	0.7678	1.0266
0.25	0.7677	1.0266
0.30	0.7677	1.0265
0.35	0.7676	1.0264
0.40	0.7675	1.0263
0.45	0.7675	1.0262
0.50	0.7674	1.0261
0.55	0.7673	1.0260
0.60	0.7673	1.0259
0.65	0.7673	1.0258
0.70	0.7673	1.0257
0.75	0.7675	1.0257
0.80	0.7681	1.0259
0.85	0.7694	1.0267
0.90	0.7730	1.0299
0.925	0.7778	1.0346
0.95	0.7887	1.0467
0.975	0.8191	1.0871
1.00	0.9494	1.2897