Approximation, Visualization, and Short-Term Forecasting of the Current Production of Wind Turbines in Brandenburg Based on Public Data: Down to the Level of Individual Turbines

Christian Daniel Matthaei
(20-478935-22)

Supervisor: Prof. Dr. Raad Tareaf
Prof. Dr. Bastian Halecker

Study Program: Coding and Software Engineering
XU Exponential University of Applied Sciences

This thesis is submitted for the degree of
Bachelor of Science (B.Sc.)

November 16, 2025

Signed Declaration of Authorship

I hereby declare that the work written and presented in this thesis is entirely my own. Where I have consulted the works of others, this is always clearly stated, with the consulted sources being individually named in the bibliography. The work at hand has not been presented for a degree at any other educational institution. It does not include material that to a substantial extent has been part of any other assignment for the duration of my studies at XU Exponential University and has not yet been published.

Christian Daniel Matthaei
November 16, 2025

Disclosure Statement

While drafting this manuscript, the author utilized Cursor for the purpose of rephrasing text. The specific material produced by these AI tools is elaborated in Appendix B. Before the final submission, the author has meticulously examined the AI-generated content and assumes complete accountability for the contents of the submitted thesis.

Christian Daniel Matthaei
November 16, 2025

Abstract

Wind power has become a dominant contributor to Germany's renewable electricity generation, yet transparency remains limited to municipal-level aggregation due to inaccessible individual turbine production data. This thesis develops a methodology to approximate, visualize, and forecast wind power production at the individual turbine level using exclusively public data sources.

Conducted within the RITS project framework, this research synthesizes meteorological measurements from 32 Deutscher Wetterdienst (DWD) weather stations, metadata from 4,796 Brandenburg wind turbines, and manufacturer power curves through a multi-stage pipeline. The methodology applies inverse distance weighting for spatial interpolation, power law extrapolation to hub height, and turbine-specific power curves for production estimation. Two deep learning architectures — Bidirectional LSTM and PatchTST — predict wind conditions over a 12-hour horizon, with forecasts propagated through the estimation pipeline.

The Bidirectional LSTM achieves mean absolute errors of 0.88 m/s for wind speed and 38.73° for wind direction, representing 38.4% and 9.5% improvements over the baseline model - a persistence model that propagates the most recent observation forward. Spatial interpolation and power production estimates are validated through reasonability analysis, confirming physically plausible patterns. A web-based prototype operationalizes the methodology, providing regional and turbine-level visualization of current approximations and 12-hour forecasts.

This work demonstrates that turbine-level wind power transparency is achievable without proprietary data. While limitations include the absence of empirical validation against actual measurements and idealized operational assumptions, the modular framework establishes a foundation for public engagement with renewable energy systems and provides a reproducible approach adaptable to other regions.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
BiLSTM	Bidirectional Long Short-Term Memory
CC BY 4.0	Creative Commons Attribution 4.0
CNN	Convolutional Neural Network
CNN-LSTM	Convolutional Neural Network-Long Short-Term Memory
DBM	Deep Boltzmann Machine
DBN	Deep Belief Network
DWD	Deutscher Wetterdienst
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
ESN	Echo State Network
EWF	Energy, Water, and Food
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
IDW	Inverse Distance Weighting
ILB	Investitionsbank des Landes Brandenburg
LSTM	Long Short-Term Memory

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MaStR Marktstammdatenregister

ML Machine Learning

MLP Multi-Layer Perceptron

MSE Mean Squared Error

PatchTST Patch Time Series Transformer

RITS Resilient Infrastructure Technology Suite

RMSE Root Mean Squared Error

RNN Recurrent Neural Network

SDK Software Development Kit

TCN Temporal Convolutional Network

UTC Coordinated Universal Time

VMD Variational Mode Decomposition

WAsP Wind Atlas Analysis and Application Program

WPD Wavelet Packet Decomposition

XU XU Exponential University

Chapter 1

Introduction

The transition toward renewable energy sources is accelerating globally, positioning them as a critical pillar of modern energy systems. In Germany, renewables accounted for 54.4% of electricity generation in 2024, reflecting a 3 percentage point increase from the previous year. This translates to an absolute output of 284 billion kWh from renewable sources. Among these, wind energy has emerged as the dominant contributor, generating 138.9 billion kWh and representing 48.9% of all renewable electricity production. The sector's continued growth is evidenced by the installation of 3,337 MW of new wind capacity during 2024. Moreover, approvals for an additional 15,000 MW of wind capacity signal a substantial acceleration in deployment for the coming years [44].

Despite this rapid expansion, the renewable energy transition remains largely abstract and challenging for the general public to comprehend. Existing initiatives such as E.ON's Energy Monitor offer valuable insights into how renewable sources contribute to local electricity supply, yet it presents data aggregated at the municipal level [19], precluding granular analysis of individual solar or wind installations. Additionally, the accuracy of such systems is limited by their reliance on standardized profiles to estimate production and consumption patterns within municipalities. Consequently, these monitoring tools depend on real-world measurement data from municipal energy systems to provide their data [27].

A comprehensive literature review conducted during this research revealed no existing framework or model capable of accurately approximating wind turbine production across arbitrary geographical locations. The absence of a standardized methodology poses considerable obstacles for reliable energy output forecasting under diverse meteorological and operational conditions throughout Germany. Compounding these challenges is the ownership structure of wind turbines, which are predominantly held by private entities and individuals [49]. This fragmented owner-

ship landscape prevents standardized data access protocols, thereby hindering efforts to compile comprehensive production datasets.

The inherent variability of renewable energy generation, predominantly influenced by meteorological conditions, introduces substantial fluctuations in electricity supply [15]. As renewable penetration increases, grid operators face "increased grid operation complexity" [37], thereby underscoring the critical need for robust forecasting capabilities in renewable energy production.

1.1 Resilient Infrastructure Technology Suite

The Resilient Infrastructure Technology Suite (RITS) represents an ILB-funded research initiative at XU Exponential University dedicated to developing a comprehensive process model and technology suite that facilitates Brandenburg's transition toward resilient infrastructure systems. With a focus on energy, water, and food infrastructure (EWF systems), RITS endeavors to establish flexible and sustainable systems capable of absorbing environmental and systemic disruptions. The initiative convenes a multistakeholder network comprising startups, research institutions, infrastructure operators, businesses, and public administration to develop a digital, data-driven platform featuring a "digital twin" of Brandenburg's infrastructure landscape. Employing advanced technologies such as artificial intelligence and blockchain, RITS creates interoperable solutions, cultivates resilient infrastructure networks, and facilitates stakeholder collaboration toward building a decentralized ecosystem for sustained resilience [53].

This thesis is conducted within the framework of the RITS project and aims to develop a methodology that integrates seamlessly with the broader RITS initiative. Consequently, the geographical scope of this research is confined to the federal state of Brandenburg.

1.2 Research Question and Thesis Structure

The central research question guiding this thesis is: *How can the current energy production of wind turbines in Brandenburg be approximated, visualized, and short-term forecasted using public data down to the level of individual turbines?*

To address this question, the thesis proceeds as follows. Chapter 2 introduces fundamental concepts of wind energy, spatial interpolation techniques, and time series forecasting methods. Chapter 3 surveys existing research on wind power forecasting and spatial extrapolation methodologies. Chapter 4 outlines the research design,

characterizing the study as both conceptual and quantitative with dual objectives: approximating current turbine production and enabling short-term wind forecasting. Chapter 5 describes the three primary data sources: meteorological measurements from 32 Deutscher Wetterdienst (DWD) weather stations, metadata from 4,796 active wind turbines registered in the Marktstammdatenregister (MaStR), and manufacturer-specific power curves. Chapter 6 presents the technical methodology, encompassing three forecasting approaches for wind conditions: a persistence baseline model and two deep learning architectures (Bidirectional LSTM and PatchTST). Subsequently, the chapter details the spatial interpolation of wind measurements to turbine locations using Inverse Distance Weighting (IDW), vertical extrapolation to hub height via the power law method, and the translation of wind speeds into electrical power output through turbine-specific power curves. Chapter 7 evaluates model performance and discusses the reasonability of production estimates. Chapter 8 details the web-based prototype comprising a FastAPI backend and Next.js frontend. Chapter 9 interprets the results presented in Chapter 7, identifies key limitations of the proposed methodology, discusses practical implications for wind energy forecasting, and outlines directions for future research. Finally, Chapter 10 consolidates the research findings and reinforces the central contribution: demonstrating that individual wind turbine production can be approximated using exclusively public data sources.

Chapter 2

Theoretical Background

2.1 Wind Energy Fundamentals

2.1.1 Wind Energy and Wind Turbines

Wind energy constitutes a fundamental pillar of the global renewable energy transition. As defined by the U.S. Department of Energy, wind power harnesses the kinetic energy of moving air to generate electricity through wind turbines, which convert the rotational motion of turbine blades into electrical energy [46].

The operational principle of a wind turbine relies on aerodynamic forces acting upon the rotor blades. When air flows across the blade surfaces, it generates both *lift* and *drag* forces, with lift being the dominant component that induces rotational motion. This rotation is subsequently transmitted to a generator — either directly or through a mechanical gearbox — where the mechanical energy is transformed into electrical power. Wind turbines can be deployed in various configurations depending on site characteristics and energy requirements. Large-scale installations are typically organized as wind farms that feed power directly into the electrical grid, whereas smaller distributed turbines serve localized applications such as individual residences, agricultural facilities, or remote off-grid locations. Furthermore, installations can be categorized by their placement as land-based, offshore, or distributed systems positioned near the point of consumption [45].

2.1.2 Power Curves

A wind power curve is a graphical representation that illustrates the relationship between wind speed and electrical power output of a wind turbine. This curve serves as an essential tool for estimating the energy generation potential of a turbine at a given site. The power curve exhibits a characteristic shape defined by three

critical wind speed thresholds that govern turbine operation. The first threshold is the *cut-in speed*, typically ranging from 3 to 4 m/s, at which the turbine initiates power generation. As wind speed increases beyond the cut-in point, power output rises progressively until reaching the *rated power*, generally achieved at wind speeds between 12 and 17 m/s. Once the rated power is attained, control mechanisms — predominantly blade pitch adjustment — regulate the output to maintain a constant power level despite further increases in wind speed. Finally, at the *cut-out speed* of approximately 25 m/s, the turbine is shut down as a protective measure to prevent structural damage from excessive wind forces [5].

2.1.3 Generic Power Curves and Wind Speed Distribution

In situations where manufacturer-specific power curves are not available, generic power curves can be constructed through statistical analysis of wind speed data. The Weibull distribution has emerged as the standard probabilistic model for characterizing wind speed distributions, providing a crucial foundation for analyzing and modeling generic wind turbine power curves. The Weibull probability density function is mathematically expressed as:

$$f(v) = \frac{k}{A}\left(\frac{v}{A}\right)^{k-1} \exp\left[-\left(\frac{v}{A}\right)^{k}\right]$$

where v denotes the wind speed, A represents the scale parameter, which is proportional to the mean wind speed, and k denotes the shape parameter. The shape parameter k characterizes the variability of wind conditions at a given location. Values of k approaching 1 indicate highly variable wind regimes with significant speed fluctuations, whereas larger values of k — typically up to approximately 3 — correspond to more stable and consistent wind conditions [41].

2.1.4 Wake Effects

In operational wind farms, turbine energy production is significantly influenced by aerodynamic interactions between individual machines. When a wind turbine extracts energy from the flowing air, it generates a *wake* region downstream characterized by reduced wind speeds and increased turbulence. Under certain conditions, this wake effect can theoretically decrease wind speeds for downstream turbines by up to 30%. However, modern wind farm layouts account for wake effects during the planning phase through strategic turbine placement, typically limiting actual production losses to approximately 10% in real-world installations. The magnitude of wake effects is determined by several factors, including inter-turbine spacing, wind direction, and wind speed [20].

2.1.5 Wind Vector Components

The rigorous mathematical analysis of wind direction necessitates its decomposition into Cartesian vector components. In meteorological practice, horizontal wind velocity is represented through two orthogonal components: the *zonal component u*, oriented in the east-west direction, and the *meridional component v*, oriented in the north-south direction. Given a wind speed u_i and a meteorological wind direction θ_i expressed in degrees — defined according to the meteorological convention as the direction from which the wind originates — the Cartesian components are derived through conversion to radians and application of a sign convention that transforms the source direction into the direction of flow:

$$u = -u_i \times \sin\left(\frac{2\pi\,\theta_i}{360}\right), \quad v = -u_i \times \cos\left(\frac{2\pi\,\theta_i}{360}\right).$$

The negative signs in these expressions accommodate the meteorological convention. While θ_i specifies the source direction of the wind, the vector components must represent the direction toward which the air mass moves. Consequently, the transformation incorporates negation to ensure alignment with standard vector orientation conventions. An important characteristic of this decomposition is that each directional vector is implicitly weighted by its magnitude [22]. Conversely, the wind speed magnitude can be recovered from the components using the relationship $u_i = \sqrt{u^2 + v^2}$ [18].

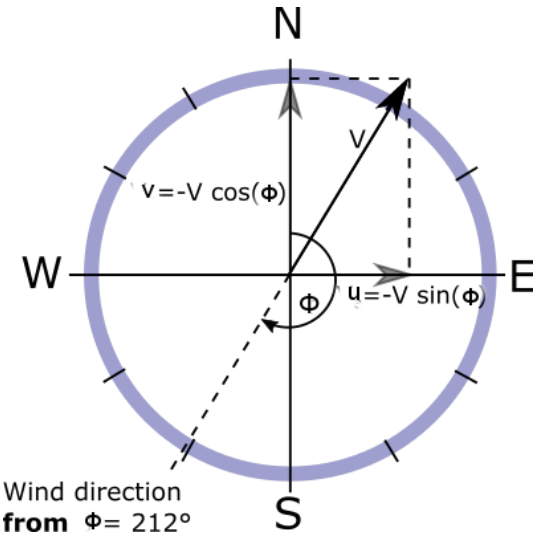

Fig. 2.1 Visualization of u and v components

2.2 Spatial Interpolation and Extrapolation

2.2.1 Spatial Interpolation

The estimation of values at unmeasured locations from sparse observations necessitates the application of interpolation techniques. In its most general definition, interpolation refers to the addition of a calculated value into a series based on the values preceding and following it [3]. When applied to geographical contexts, spatial interpolation constitutes the process of utilizing data points with known values to estimate values at locations where direct measurements are unavailable [38].

This methodology is widely employed to generate continuous surfaces — such as maps of precipitation, elevation, or population density — from a limited number of discrete measurement points. The necessity for such techniques arises when practical constraints, including financial limitations, site accessibility, or logistical challenges, preclude comprehensive data collection across all locations of interest. Through the application of interpolation methods, discrete point observations can be transformed into continuous raster surfaces, thereby enabling comprehensive spatial analysis and modeling across an entire study area [38].

2.2.2 Inverse Distance Weighting

Among the various spatial interpolation methods available, Inverse Distance Weighting (IDW) represents a deterministic approach for estimating unknown values at unsampled locations based on the values observed at nearby measurement points. The method employs a distance-based weighting scheme in which the estimated value $\hat{z}(x)$ at location x is calculated as a weighted mean of measured values:

$$\hat{z}(x) = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i}$$

where $w_i = |x - x_i|^{-\beta}$ represents the weights, which vary inversely with the distance between the prediction location and the measurement points [23].

The fundamental assumption underlying IDW is the principle of *spatial autocorrelation*: locations that are closer in space tend to exhibit more similar values than those that are farther apart. As a consequence of this assumption, nearby observations exert a proportionally stronger influence on the predicted value than more distant observations [23].

The weighting behavior of the interpolation is governed by the power parameter $\beta \geq 0$, typically set to 1 or 2, which controls the rate at which the influence of an observation decreases with distance. When $\beta = 1$, corresponding to inverse

distance weighting, the influence decays gradually, resulting in smoother interpolated surfaces. When $\beta = 2$, corresponding to inverse squared distance weighting, the influence diminishes more rapidly, thereby producing greater emphasis on proximate observations. Higher values of β accentuate local effects, which is particularly advantageous for variables exhibiting strong spatial variability, such as precipitation or temperature [23].

Practical implementation of IDW necessitates decisions regarding the number of neighboring points n to include in the calculation, which determines whether local or global weighting is applied. The optimal values of n and β are typically determined through cross-validation procedures. A special case occurs when the target location coincides exactly with an observation point; in this instance, the measured value is returned directly to circumvent numerical instability arising from infinite weights [23].

Despite its widespread adoption, IDW exhibits inherent limitations that warrant consideration. The quality of interpolation may deteriorate when the spatial distribution of observation points is irregular, thereby introducing increased uncertainty in sparsely sampled regions. Furthermore, the method imposes a constraint whereby minimum and maximum values can only occur at observation points rather than in the interpolated space between them, meaning that interpolated fields are bounded by the range of measured values [38].

2.2.3 Vertical Extrapolation

Whereas interpolation estimates values within the spatial domain bounded by observations, extrapolation extends predictions beyond the range of measured data. In its broadest sense, extrapolation constitutes the process of utilizing existing information to infer or predict outcomes in domains where direct observations are absent [2]. Within the context of wind energy applications, vertical extrapolation refers to the technique employed to estimate wind speeds at turbine hub heights from near-surface wind measurements, thereby facilitating accurate assessment of wind energy resources when direct measurements at operational heights are unavailable [35].

2.2.4 Power Law for Vertical Wind Speed Extrapolation

The power law constitutes an empirical method extensively employed to estimate wind speeds at varying elevations, particularly within wind energy applications where measurements are typically obtained only at lower heights. This approach enables the extrapolation of wind speed from a known reference elevation to the hub

height of a wind turbine, thereby facilitating both energy potential assessment and strategic siting decisions [51].

The vertical wind speed profile is mathematically expressed through a power-law relationship between two heights:

$$u_2 = u_1 \left(\frac{z_2}{z_1} \right)^{\alpha}$$

where u_1 and u_2 denote wind speeds at heights z_1 and z_2, respectively, and α represents the wind shear exponent, which characterizes the rate at which wind speed varies with elevation. In cases where measurements at two distinct heights are available, the shear exponent can be empirically determined through the following relationship [51]:

$$\alpha = \frac{\ln(u_2/u_1)}{\ln(z_2/z_1)}$$

Within practical wind energy studies, the power law is frequently preferred due to its computational simplicity and acceptable accuracy in situations where comprehensive data or detailed thermodynamic information are unavailable. However, it is important to recognize that the exponent α is not a universal constant but rather varies as a function of surface roughness, atmospheric stability, and terrain complexity. Atmospheric conditions exert a particularly strong influence on vertical wind shear behavior. Under stable atmospheric conditions — characterized by weak vertical mixing and cool surface temperatures — the shear exponent tends to assume larger values, thereby indicating more pronounced wind shear. Conversely, unstable atmospheric conditions — featuring strong vertical mixing and convective turbulence — produce smaller values of α, which correspond to more uniform vertical wind speed profiles [51].

2.3 Time Series Forecasting

2.3.1 Machine Learning Fundamentals

The forecasting methodology employed in this research leverages machine learning techniques to predict future wind conditions based on historical observations. At its core, machine learning represents the capacity of computational systems to process and evaluate data beyond explicitly programmed algorithms through contextualized inference [14]. More comprehensively defined, Machine Learning (ML) constitutes a subfield of Artificial Intelligence (AI) that enables computational systems to emulate human learning processes. ML systems operate by autonomously executing tasks

and iteratively enhancing their performance through learning from data and prior experiences, without requiring explicit programming for each specific scenario [24].

A machine learning system comprises three fundamental components. First, a decision process whereby an algorithm generates predictions or classifications based on input data, which may be either labeled or unlabeled. Second, a loss or error function that quantifies the accuracy of predictions by comparing them against known outcomes. Third, an optimization process through which model parameters — commonly referred to as weights — are iteratively adjusted to minimize prediction errors and enhance overall performance [24].

Machine learning facilitates the discovery of complex patterns within large-scale datasets, often requiring minimal manual intervention, thereby enhancing efficiency and enabling personalization across numerous application domains. However, the effectiveness of ML models is fundamentally dependent on the availability of large, high-quality datasets. Poor quality or biased input data can substantially degrade model performance, a phenomenon commonly characterized by the principle "garbage in, garbage out." Furthermore, models are susceptible to various challenges including instability, overfitting, and latent biases if not subjected to rigorous validation procedures [24].

2.3.2 Time Series Forecasting Concepts

A time series is defined as a sequence of observations recorded at regular temporal intervals, where the chronological ordering of data points is essential to the analysis. Time series forecasting is concerned with predicting future values based on historical observations. However, unlike other predictive modeling approaches such as regression, time series forecasting employs past values of the same variable as inputs to predict its future states, thereby rendering time itself an intrinsic predictor [28].

Time series data typically comprise several key components. These include trend, representing the long-term directional movement of the series; seasonality, characterized by repeating short-term cycles such as daily, monthly, or yearly patterns; cyclic behavior, manifested as irregular long-term fluctuations exemplified by economic cycles; and noise, representing random and unpredictable variation. A fundamental characteristic distinguishing time series data from other data types is temporal dependency. Time series analysis operates under the assumption of temporal dependency between successive data points, meaning that adjacent observations exhibit correlation — a property formally known as autocorrelation [28].

2.3.3 Forecast Evaluation Metrics

The rigorous assessment of forecasting model performance necessitates the application of quantitative error metrics. This research employs two standard metrics for evaluation: Mean Absolute Error and Root Mean Squared Error.

Mean Absolute Error (MAE) is a statistical metric that measures the accuracy of predictive models by quantifying the average magnitude of prediction errors. This metric calculates the mean of the absolute differences between predicted and actual values, thereby disregarding the directional sign of individual errors. The MAE is mathematically expressed as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where n represents the number of observations, y_i denotes the actual values, and $\hat{y}_i$ represents the predicted values. The resulting metric represents the average absolute deviation of predictions from the true values and is expressed in the same units as the target variable, facilitating direct interpretation [6].

Root Mean Squared Error (RMSE) constitutes a widely employed evaluation metric in regression analysis that quantifies the typical magnitude of prediction errors. This metric expresses the extent to which predicted values deviate, on average, from the actual observed values. RMSE is mathematically defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where y_i represents the actual value, $\hat{y}_i$ denotes the predicted value, and n indicates the number of observations. The squaring of errors prior to averaging assigns disproportionately higher penalties to large deviations, thereby making RMSE particularly appropriate when substantial prediction errors are considered especially problematic. Like MAE, RMSE quantifies prediction accuracy in the same units as the response variable, rendering it intuitive and directly comparable to the practical context of the predictions. RMSE is computed as the square root of the Mean Squared Error (MSE).

While MSE is extensively employed as a loss function during model training through optimization algorithms such as gradient descent, RMSE is predominantly utilized in post-training evaluation to report model performance in interpretable units [7].

2.3.4 Baseline Models and Model Comparison

The rigorous evaluation of sophisticated forecasting algorithms necessitates comparison against simpler reference approaches to establish their relative merit. A baseline model is defined as an intentionally simple predictive model employed to establish a minimum performance threshold against which more complex approaches can be assessed. Such models serve as benchmarks for determining whether advanced algorithms deliver meaningful performance improvements that justify their increased computational and conceptual complexity. By establishing a performance floor that sophisticated models must exceed, baseline models play a crucial role in preventing unnecessary algorithmic complexity. They provide a clear reference point for assessing whether additional model sophistication yields commensurate benefits in predictive accuracy [25].

Several categories of baseline models are commonly employed in forecasting applications. Random baselines generate predictions through stochastic processes and are particularly useful when no prior information about the prediction task is available. Majority class baselines consistently predict the most frequently occurring class and are especially relevant for classification tasks involving imbalanced datasets. Simple heuristic baselines employ elementary domain-specific rules, such as classifying text sentiment based on the ratio of positive to negative words [25].

2.3.5 Persistence Model for Weather Forecasting

The persistence model constitutes a straightforward baseline approach that assumes future values will remain constant at the most recently observed value. Mathematically, this is expressed as

$$P(t+d) = P(t),$$

where $d \in \{1, 2, \ldots, 12\}$ represents the forecast timestep and $P(t)$ denotes the value at the last known time t. This model demonstrates strong performance in very-short-term forecasting scenarios. However, its predictive accuracy degrades progressively as the forecast horizon extends [50]. The rationale for employing persistence as a baseline model derives from the inherent autocorrelation characteristic of atmospheric variables. Since meteorological conditions in the near future tend to closely resemble current conditions, simple extrapolation provides a non-trivial reference point against which more sophisticated forecasting methods can be evaluated.

2.3.6 Long Short-Term Memory Model

Long Short-Term Memory (LSTM) networks constitute a specialized architecture designed to address the limitations of traditional recurrent neural networks in learning temporal patterns across extended sequences. Introduced by Hochreiter and Schmidhuber in 1997, LSTMs represent a particular variant of Recurrent Neural Networks (RNNs) specifically engineered to learn and retain information over long sequential dependencies. Unlike conventional feedforward neural networks that process each input independently, LSTMs maintain an internal memory of previous information through recurrent connections that link past outputs to subsequent inputs. This architectural feature enables the network to make predictions that depend on contextual information from earlier positions in the sequence [33].

The fundamental concept underlying LSTM architecture is the *cell state*, denoted as C_t, which functions analogously to a conveyor belt traversing the entire sequence and facilitating information flow with minimal transformation. The principal advantage of LSTMs over standard RNNs lies in their capacity to mitigate the long-term dependency problem — the inherent difficulty that conventional RNNs encounter when attempting to retain information from distant time steps as the temporal gap between relevant inputs increases [33].

The propagation of information through the cell state is regulated by a system of gating mechanisms that control the information flow by determining which information to retain, which to discard, and what new information to incorporate. The forget gate determines which components of the previous cell state should be discarded, while the input gate, in conjunction with candidate values, determines what new information should be stored in the cell state. The cell state is subsequently updated by combining the filtered previous state with the new candidate information. Finally, the output gate regulates which portion of the cell state is exposed in the output. These coordinated mechanisms enable LSTMs to preserve relevant information over extended periods, discard irrelevant data, and selectively update their internal memory, thereby rendering them particularly effective for sequential tasks such as language modeling, machine translation, and speech recognition [33].

2.3.7 Bidirectional LSTM Model

Bidirectional Long Short-Term Memory (BiLSTM) networks represent an architectural extension of conventional LSTM networks that processes sequential data in both forward and backward temporal directions. The architecture comprises two distinct LSTM layers operating in parallel: a forward layer that traverses the input sequence from the initial to the final time step, and a backward layer that processes

the sequence in reverse chronological order. This bidirectional processing paradigm enables the network to capture contextual information from both preceding and succeeding time steps, rendering it particularly advantageous for tasks requiring comprehensive sequence context, such as sentiment analysis [21].

The final output representation at each time step t is obtained by combining the outputs from both directional layers according to the formula

$$p_t = p_t^f + p_t^b$$

where p_t denotes the final output vector, p_t^f represents the forward LSTM output, and p_t^b represents the backward LSTM output. Through this integration of bidirectional representations, BiLSTMs generate enriched contextual embeddings that provide more comprehensive information for subsequent processing layers or classification modules [21].

2.3.8 PatchTST Model

PatchTST (Patch Time Series Transformer) constitutes a Transformer-based architecture specifically developed for multivariate time series forecasting and self-supervised representation learning [32].

The architecture is founded upon two fundamental design principles. First, the *patching mechanism* segments time series data into subseries-level patches, thereby generating tokens that preserve local semantic information within each patch. Second, the *channel-independence principle* processes each variable (channel) of the multivariate series independently through a shared Transformer backbone, facilitating parameter sharing across series while preventing undesirable channel mixing [32].

The patching mechanism confers significant computational advantages through token count reduction. Specifically, this mechanism reduces the number of input tokens from L to approximately $\frac{L}{S}$, where L denotes the look-back window length and S represents the stride length. This reduction consequently decreases the quadratic time and memory complexity of the self-attention mechanism from $O(N^2)$ to $O\left(\left(\frac{L}{S}\right)^2\right)$. The patching approach thereby enables the model to capture local temporal patterns, substantially reduce computational requirements, and leverage longer historical contexts for improved forecasting accuracy [32].

Within the channel-independent configuration, each univariate time series $x^{(i)} = (x_1^{(i)}, \ldots, x_L^{(i)})$ generates its corresponding prediction

$$\hat{x}^{(i)} = (\hat{x}_{L+1}^{(i)}, \ldots, \hat{x}_{L+T}^{(i)})$$

The overall loss function is computed as the mean squared error averaged across all M channels:

$$\mathscr{L} = \frac{1}{M} \sum_{i=1}^{M} \| \hat{x}_{L+1:L+T}^{(i)} - x_{L+1:L+T}^{(i)} \|_2^2$$

To enhance training stability, the model employs instance normalization, which normalizes each time series to zero mean and unit variance, thereby mitigating distributional discrepancies between training and test data [32].

PatchTST additionally supports self-supervised learning through a patch-level masking strategy, wherein randomly selected portions of the input are masked and the model is trained to reconstruct the masked content, thereby encouraging the acquisition of abstract and transferable representations. Empirical evaluations demonstrate that PatchTST achieves state-of-the-art performance in forecasting tasks and exhibits strong capabilities in transfer learning and representation learning scenarios, thereby confirming the efficacy of Transformer architectures for long-term time series forecasting applications [32].

Chapter 3

Related Work

This chapter examines the body of research pertinent to the two central aims of this thesis: the forecasting of wind conditions at meteorological stations and the approximation of turbine-level power production via spatial inter- and extrapolation.

3.1 Machine Learning for Wind Forecasting

Machine learning applications in short-term wind forecasting have undergone substantial development over the past decades. In a systematic review encompassing 23 studies from 1983 to 2023, [1] examine wind *nowcasting* methodologies across prediction horizons spanning one minute to one week. The analysis reveals that neural network architectures, especially contemporary deep learning models including LSTM, Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM), Graph Neural Network (GNN), Temporal Convolutional Network (TCN), and ensemble configurations, systematically surpass traditional statistical methods in predicting short-term wind speed and direction. Mean performance across the reviewed studies yields MAE of 0.56 m/s, MSE of 1.10 m/s, and MAPE of 6.72%. For high-resolution forecasts employing 1- to 10-minute intervals, errors decrease markedly, with the most effective ensemble approach achieving MAE of 0.02 m/s and MAPE of 0.57% at one-minute lead times. A key finding relevant to this thesis is that incorporating additional meteorological variables beyond wind-specific features provides no systematic accuracy gains, whereas finer temporal resolution and sophisticated deep learning architectures consistently improve forecast performance. The review further notes that joint prediction of wind speed and direction remains underexplored, and the absence of standardized evaluation metrics hinders meaningful cross-study comparison.

Building on this foundation, [50] investigate deep learning approaches with specific focus on hybrid model architectures. Such hybrid frameworks commonly integrate signal decomposition or denoising methods — including Empirical Mode Decomposition (EMD)/Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Variational Mode Decomposition (VMD), and Discrete Wavelet Transform (DWT)/Wavelet Packet Decomposition (WPD) — with feature selection mechanisms and deep learning predictors such as LSTM, Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), Echo State Network (ESN), and ConvLSTM variants. Their analysis demonstrates that hybrid models consistently outperform single-architecture baselines, with error metrics including MAE, RMSE, and MAPE showing notable reductions when decomposition techniques are coupled with LSTM or CNN-LSTM configurations. The authors identify ongoing challenges encompassing nonstationarity, spatiotemporal dependencies, measurement noise, overfitting tendencies, and constrained interpretability, while highlighting probabilistic forecasting and attention-based Transformer models as promising research directions.

A concrete implementation of the hybrid paradigm is presented by [40], who develop a framework integrating Boruta wrapper-based feature selection with a Bidirectional LSTM network. The feature selection stage performs dimensionality reduction by identifying the most predictive meteorological variables, thereby alleviating overfitting risks and computational complexity associated with high-dimensional inputs. The Bi-LSTM component then models temporal dependencies bidirectionally, enhancing the network's ability to capture uncertainty and nonlinear dynamics inherent in wind behavior. Evaluated on an hourly European meteorological dataset containing 31 features across six years, the approach selected 13 salient variables—encompassing temperature, radiation, humidity, and soil properties — and attained RMSE of 0.784, MAE of 0.530, MSE of 0.615, and $R^2 = 0.8766$, surpassing MLP, standard LSTM, and baseline Bi-LSTM benchmarks. This work underscores the efficacy of integrating feature selection with deep learning for short-term wind forecasting and suggests that similar preprocessing strategies could enhance LSTM or Transformer-based models when approximating turbine production from publicly available station measurements.

Recent advances are exemplified by [52], who introduce a heterogeneous PatchTST-GRU sequence-to-sequence architecture for multi-step wind power forecasting. Their model combines historical turbine observations with refined numerical weather predictions via a fusion attention mechanism. The PatchTST module employs Transformer-based processing to extract long-range temporal patterns, while the

GRU component efficiently handles sequential dependencies. Results demonstrate that integrating patch-based temporal segmentation with recurrent networks yields robust multi-horizon wind forecasts. Critically, the authors emphasize that forecast accuracy hinges on reconciling the spatial resolution gap between large-scale meteorological inputs and site-specific turbine conditions — a challenge that parallels the present research objective of forecasting wind conditions from spatially sparse station networks.

3.2 Spatial Interpolation and Extrapolation

Estimating wind conditions at turbine sites from spatially sparse meteorological station observations necessitates reliable horizontal interpolation and vertical extrapolation techniques. [36] examine vertical extrapolation methods for wind resource characterization and energy yield estimation. Drawing on measurements from 21 meteorological masts distributed across varied terrain types, the study contrasts the logarithmic-based WAsP profile against the power law approach. Results indicate that the logarithmic formulation demonstrates lower predictive reliability, while the power law — employing wind shear coefficients derived from the two highest measurement levels and stratified by 30° directional sectors — yields superior accuracy and diminishes uncertainty in annual energy production forecasts. A central conclusion is that extrapolation fidelity depends critically on measurement height configuration and directional shear parameterization, whereas finer discretization through additional wind speed bins offers negligible performance gains. Nonetheless, site-specific terrain characteristics and atmospheric stability variations introduce persistent challenges that constrain transferability across heterogeneous landscapes.

Turning to horizontal interpolation, [17] assess various techniques for projecting mesoscale weather model data to specific sites, examining IDW, linear and cubic interpolation, and nearest neighbor approaches. The evaluation reveals that although all horizontal methods exhibit comparable error magnitudes, IDW configured with 16 neighboring grid points delivers optimal results, attaining the lowest MAE near 2.5 m/s when paired with straightforward vertical interpolation schemes such as linear height adjustment. Concurrently, the study exposes fundamental interpolation limitations, with error floors of MAE above 2.4 m/s and RMSE above 3.3 m/s, and observes degraded performance in topographically complex mountainous regions attributable to the lack of terrain-adaptive interpolation strategies. These results underscore both the utility and inherent constraints of IDW when spatially extending wind observations from limited measurement networks.

Chapter 4

Research Design and Approach

4.1 Research Approach

Conducted within the framework of the RITS project, this research employs a *hybrid methodological approach* that integrates both conceptual and quantitative dimensions. The research addresses two interconnected objectives. First, it seeks to develop a methodology for approximating wind turbine power output using exclusively publicly available datasets. Second, it aims to enable short-term forecasting of turbine power production by incorporating weather predictions at measurement stations, thereby extending the approximation framework beyond historical reconstruction to future time horizons.

The quantitative dimension of this research manifests through its reliance on empirical data sources — specifically, meteorological measurements from weather stations, turbine metadata from public registries, and manufacturer-provided power curves — to derive production estimates via computational methods. Concurrently, the conceptual dimension emerges through the development of a generalizable framework that, while implemented and validated within the Brandenburg context, can be adapted to alternative geographical regions where comparable public data sources are available. This methodological duality reflects the study's commitment to both rigorous data-driven analysis and transferable system design.

The research constitutes *applied research*, addressing the practical challenge of enhancing transparency in wind energy production while improving forecasting capabilities for public and stakeholder engagement. Rather than pursuing an optimized production system, the overarching objective is to establish a *proof of concept* that demonstrates methodological feasibility. Accordingly, each component of the methodology described in Section 4.2 emphasizes simplicity and modularity, thereby facilitating future refinement or replacement with more sophisticated techniques.

Throughout this thesis, priority is accorded to straightforward, interpretable methods that establish feasibility while explicitly acknowledging avenues for methodological enhancement in subsequent research.

4.2 Research Design

The methodological design of this research draws inspiration from existing transparency initiatives, particularly the E.ON Energy Monitor, which provides visualization of renewable energy production aggregated at the municipal level. Extending beyond this municipal-scale approach, the present study seeks to achieve turbine-level granularity for wind power production in Brandenburg through a multi-stage computational pipeline that synthesizes publicly accessible data sources.

The computational methodology is structured as a sequential processing pipeline comprising four principal stages. In the first stage, meteorological data — encompassing wind speed, wind direction — are acquired from publicly available measurement stations operated by DWD, as comprehensively detailed in Chapter 5. The second stage applies spatial intrapolation techniques, specifically IDW, to translate station-level measurements to individual turbine locations. In the third stage, horizontally intrapolated wind speeds undergo vertical extrapolation from the measurement height of approximately 10 meters above ground to the hub height of each turbine using the power law method. The fourth and final stage applies turbine-specific power curves to the extrapolated hub-height wind speeds, thereby deriving instantaneous power production estimates for each installation by translating meteorological conditions into electrical output based on manufacturer specifications.

Beyond historical approximation, the methodology incorporates a predictive component to enable short-term forecasting capabilities. Machine learning models are trained to predict wind speed and direction at measurement stations over a 12-hour forecast horizon. These predicted meteorological conditions are then propagated through the identical spatial intrapolation and power estimation pipeline, thereby generating production forecasts at the individual turbine level.

Both historical approximations and prospective forecasts are operationalized through an interactive web-based visualization platform. This dashboard presents aggregated production metrics across all wind turbines in Brandenburg while simultaneously providing access to turbine-specific information, including historical production trajectories for the preceding 12 hours and forecasted outputs for the subsequent 12 hours.

While the framework is developed and validated specifically within the Brandenburg context, its modular architectural design facilitates transferability to alternative geographical regions, contingent upon the availability of comparable public data infrastructure.

4.3 Evaluation Strategy

4.3.1 Power Production Estimation

Quantitative validation of the power production estimation component faces severe limitations due to the unavailability of ground-truth production data from individual wind turbines. As elaborated in Chapter 1, turbine-level production data remains inaccessible to the public, primarily attributable to the fragmented ownership landscape of wind installations, wherein the majority of turbines are operated by private entities or individual proprietors. This data unavailability precludes direct validation against measured production values within the scope of this research. In the absence of quantitative validation data, the power production estimates are therefore evaluated through *reasonability analysis*, which assesses whether the derived estimates conform to expected physical and operational constraints.

4.3.2 Weather Forecasting

The weather forecasting component, in contrast, permits rigorous quantitative evaluation through standard machine learning validation protocols. The evaluation framework utilizes a held-out test dataset that is temporally disjoint from the training data, thereby enabling unbiased quantitative assessment of forecast accuracy via established error metrics. Chapter 6 provides a comprehensive exposition of the evaluation methodology.

Chapter 5

Data Sources

5.1 Deutscher Wetterdienst

As Germany's national meteorological service, the Deutscher Wetterdienst (DWD) operates under the Federal Ministry of Transport, delivering comprehensive weather and climate services for both public welfare and economic sectors. The "Gesetz über den Deutschen Wetterdienst" defines its mandate, establishing responsibilities that encompass both scientific research and public information dissemination [8].

A legislative amendment enacted on July 25, 2017, mandates the DWD to make weather and climate information freely available to the public. This open data initiative provides unrestricted access to weather forecasts, radar imagery, real-time measurements, observational data, and historical climate records through the dedicated portal at https://opendata.dwd.de [9]. The real-time meteorological measurements available through this platform serve as the primary data source for the present research.

5.1.1 Meteorological Measurement Stations

Across Germany, the DWD maintains a network of approximately 400 active meteorological measurement stations. These stations demonstrate considerable variability in temporal resolution and measured parameters, meaning that not all stations capture the complete suite of meteorological variables [12].

To align with the geographical scope of this research, data collection focused exclusively on stations located within Brandenburg and its immediate border regions. The selection process applied three criteria: 10-minute temporal resolution, active operational status, and continuous measurement of both wind speed and wind direction. Applying these criteria identified 18 stations within Brandenburg and 14 stations in

adjacent border regions, resulting in a total of 32 measurement stations. The spatial distribution of these selected stations is illustrated in Figure 5.1, derived from the comprehensive station inventory provided by the DWD [31].

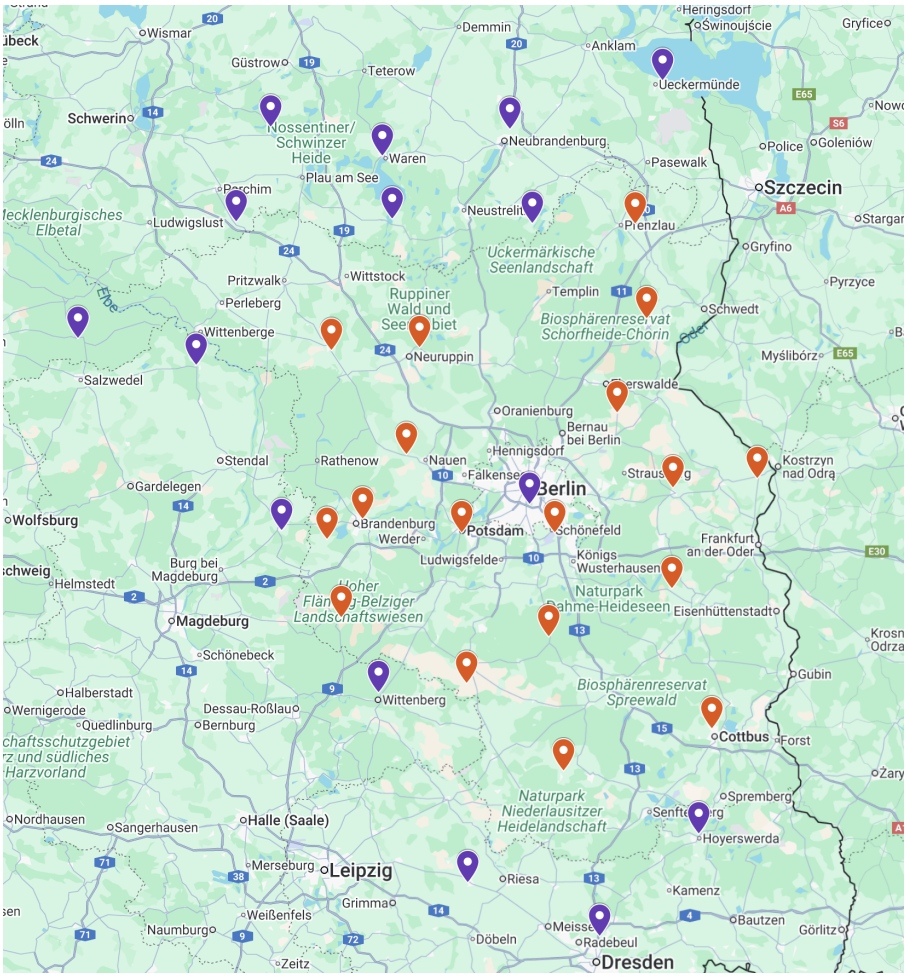

Fig. 5.1 Spatial distribution of the 33 selected DWD measurement stations in Brandenburg and adjacent border regions

Incorporating stations from border regions improves spatial coverage and enhances interpolation accuracy, especially for wind turbines situated near Brandenburg's boundaries. Nevertheless, Brandenburg's location at Germany's eastern border presents a limitation: the absence of measurement stations beyond the German-Polish border may introduce interpolation errors for turbines in eastern Brandenburg. Section 9.2 discusses this constraint and its implications for production estimation accuracy.

5.1.2 Meteorological Measurements

The DWD structures station measurements into three distinct temporal datasets, each representing a different stage of quality assurance. The *historical* dataset comprises data that has undergone complete operational quality control procedures and receives

annual updates, with all values considered final and immutable. The *recent* dataset covers the preceding 500 days and undergoes daily updates, though values remain subject to revision as quality control processes continue. The *now* dataset delivers the most current 24 hours of observations with updates occurring multiple times per hour, representing preliminary measurements that have not yet undergone quality validation [13].

This research utilized data from all three temporal datasets as provided, without explicitly incorporating quality control flags into the analysis. All timestamps follow Coordinated Universal Time (UTC) convention since the year 2000, maintaining a uniform temporal resolution of 10 minutes [13].

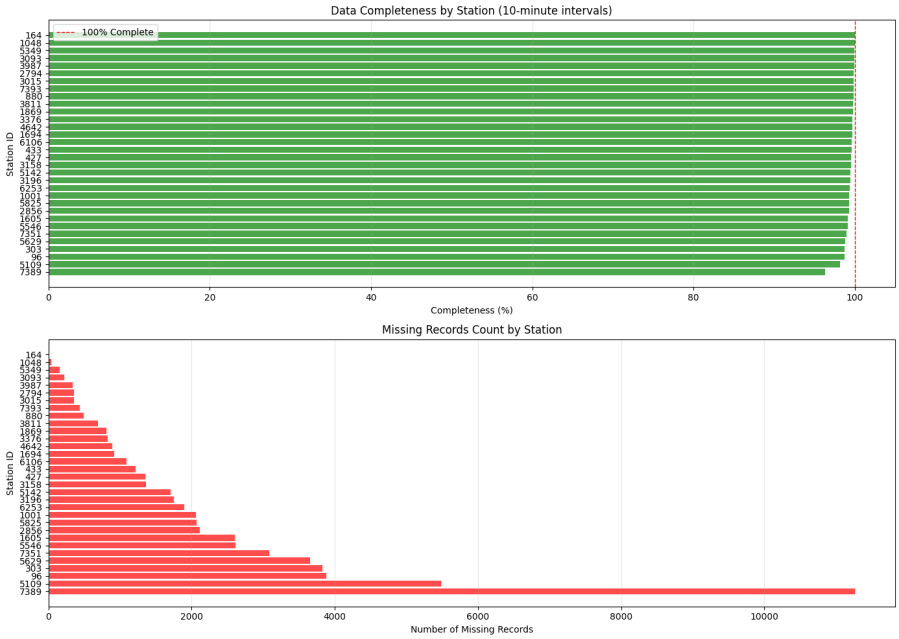

Fig. 5.2 Data completeness for the 32 selected DWD measurement stations

Data completeness exhibits variation across both stations and time intervals, as illustrated in Figure 5.2. Certain stations demonstrate substantial proportions of missing values. Across the entire dataset spanning from January 1, 2020, to October 5, 2025, the overall completeness reaches 99.20%, corresponding to 77,777 missing values. Station 7389 exhibits the lowest completeness at 96.28%. To address missing observations at specific 10-minute intervals, a backward-filling strategy was employed, propagating the most recent valid measurement forward to subsequent missing timestamps. While pragmatic, this approach introduces quality degradation when addressing extended temporal gaps.

The DWD provides access to multiple meteorological datasets beyond wind measurements. Although this research primarily employs the wind dataset, preliminary experiments evaluated forecasting models trained on temperature and precipitation

datasets as well. The superior performance of models trained exclusively on wind data corroborates the findings of [1].

Wind Parameters

The wind measurement dataset delivers 10-minute averaged values for both wind speed (expressed in m/s) and wind direction (expressed in degrees) [13]. Measurements are conducted at a nominal height of 10 meters above ground level, though actual measurement heights may vary within a tolerance of ±2 meters [10]. To maintain computational simplicity, this research assumes a standardized measurement height of 10 meters across all stations [13]. Table 5.1 provides a summary of these wind parameters .

Table 5.1 DWD wind measurement parameters

Parameter	Unit	Description
Wind speed	m/s	10-minute average wind speed
Wind direction	degrees	10-minute average wind direction

Given that this research focuses on predicting wind speed and direction over a 60-minute forecast horizon, the raw 10-minute data undergoes temporal aggregation to 60-minute intervals. This aggregation employs arithmetic mean values for both parameters.

5.2 Marktstammdatenregister

The Marktstammdatenregister (MaStR) constitutes a centralized, official registry operated by the Bundesnetzagentur (Federal Network Agency), delivering a comprehensive and standardized database for Germany's electricity and gas markets. This publicly accessible registry encompasses all energy generation and storage facilities, irrespective of subsidy status or commissioning date. Legal obligations require facility owners to register their installations in the MaStR and maintain up-to-date information. A critical characteristic of this database is its exclusive focus on *Stammdaten* (master data) — static metadata describing facility characteristics — while explicitly excluding *Bewegungsdaten* (transactional data) such as time-series production measurements [30].

A complete database extract was acquired for this research on August 15, 2025. The MaStR distributes data in XML format, necessitating parsing operations based on schema definitions specified in accompanying XSD files [29]. Although the MaStR data model incorporates numerous optional fields, only five fields remain mandatory for all registered facilities: *EinheitMastrNummer* (unique facility identifier),

DatumLetzteAktualisierung (last update date), *Bruttoleistung* (gross rated power), *Nettonennleistung* (net rated power), and *Technologie* (technology type).

The wind turbine dataset contains extensive metadata fields, thirteen of which were selected for utilization in this research. Table 5.2 presents an overview of these selected fields alongside their descriptions.

Table 5.2 MaStR wind turbine metadata fields utilized in this research

Field Name	**Description**
EinheitMastrNummer	Unique facility identifier
DatumLetzteAktualisierung	Last update date
Laengengrad	Longitude
Breitengrad	Latitude
DatumEndgueltigeStilllegung	Decommissioning date
Bruttoleistung	Gross rated power
Nettonennleistung	Net rated power
Hersteller	Manufacturer
Technologie	Technology type
Typenbezeichnung	Model designation
Nabenhoehe	Hub height
Rotordurchmesser	Rotor diameter
Land	Federal state

Dataset preparation encompassed multiple filtering and cleaning procedures applied sequentially. First, the *Land* field served as a filter criterion to retain exclusively installations located in Brandenburg. This filtering operation revealed three outliers that, despite Brandenburg registration, exhibited geographical coordinates positioned beyond state boundaries, leading to their exclusion from the analysis. Given the optional nature of the *Land* field, complete coverage of all Brandenburg turbines cannot be guaranteed. Second, 25 turbines lacking geographical coordinates (*Laengengrad* and *Breitengrad*) were eliminated, as spatial coordinates constitute essential prerequisites for the interpolation methodology. Third, the dataset was restricted to active turbines, identified through the absence of entries in the *DatumEndgueltigeStilllegung* field.

Certain turbines demonstrate missing values for *Hersteller* and *Typenbezeichnung*, both representing requirements for power curve assignment. Section 6.2 details the imputation strategy employed to address these missing values. Upon completion of all filtering and cleaning procedures, the final dataset encompasses 4,796 active wind turbines distributed across Brandenburg.

Although the MaStR additionally provides information concerning operational constraints and shutdown requirements, these parameters remained excluded from the

present analysis to avoid the substantial increase in model complexity their incorporation would necessitate.

5.3 Wind Power Curves

As elaborated in Section 2.1, power curves define the functional relationship between wind speed and electrical power output for wind turbines, thereby enabling instantaneous production estimation from meteorological measurements. These manufacturer-specific curves serve as essential instruments for converting spatially and vertically extrapolated wind speeds into turbine-level power generation estimates.

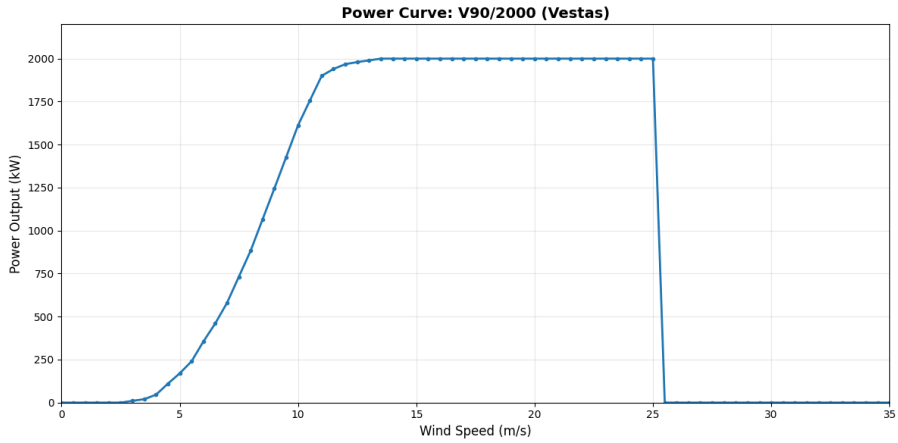

Fig. 5.3 Example power curve for a wind turbine (Vestas V90/2000) from The Wind Power database

This research draws power curve data primarily from "The Wind Power," a comprehensive database encompassing all aspects of the global wind energy sector. The platform delivers both quantitative and qualitative information on wind turbines, maintaining currency through data updates performed at minimum twice annually. The database catalogs power curves for 896 distinct turbine models, systematically organized by manufacturer and model designation [42]. Individual power curves specify electrical output in kilowatts (kW) at 0.5 m/s wind speed increments, thereby providing sufficient granularity for production estimation purposes. Figure 5.3 illustrates an exemplary power curve from this database.

Although maintained as a commercial resource rather than open data, database access can be secured through a one-time payment of 125 Euros. The complete power curve dataset was procured for this research to maximize coverage of the turbine inventory documented in the MaStR. Nevertheless, the power curve database does not cover all manufacturer-model combinations present within the MaStR dataset. For turbines

lacking specific power curves, a standardized generic power curve serves as an approximation.

Chapter 6

Methodology

6.1 Weather Forecasting

The weather forecasting component is designed to predict wind speed and direction at each of the 32 measurement stations over a 12-hour forecast horizon with hourly temporal resolution. This task represents a multivariate time series forecasting problem in which historical meteorological observations are utilized to generate predictions at multiple spatial locations simultaneously.

Three distinct forecasting approaches are implemented to establish performance benchmarks and evaluate the trade-offs between model complexity and predictive accuracy. The first approach employs a persistence baseline model that serves as a simple reference point for assessing forecast skill. The second and third approaches utilize deep learning architectures—specifically, a bidirectional LSTM network and a PatchTST model — both of which represent contemporary state-of-the-art methodologies for temporal sequence modeling and time series forecasting.

6.1.1 Dataset

The forecasting models are trained and evaluated using DWD meteorological measurements, as detailed in Section 5.1.2. The dataset covers the period from January 1, 2020, to October 5, 2025 and comprises observations from 32 measurement stations at a uniform hourly temporal resolution. Since the data is already preprocessed, no additional preprocessing is necessary.

To facilitate rigorous model evaluation, the dataset is divided into temporally disjoint training and test partitions. The training partition spans from January 1, 2020, through April 1, 2025, thereby providing over five years of historical observations for model learning. The test partition covers the period from April 1, 2025, through

October 5, 2025, ensuring that evaluation is conducted on genuinely unseen future observations. This temporal partitioning strategy is applied uniformly across all three forecasting approaches to enable fair and consistent model comparison.

6.1.2 Persistence Baseline Model

A persistence model is employed as the baseline for assessing forecasting performance. This approach operates under the assumption that meteorological conditions remain unchanged throughout the forecast horizon. Consequently, the most recently observed values are propagated forward across all 12 future timesteps, corresponding to the 12-hour forecast period at hourly intervals. For each measurement station, the wind speed and wind direction recorded at time t are replicated as predictions for all subsequent hours $t+1, t+2, \ldots, t+12$.

Three principal considerations motivate the selection of the persistence model as the baseline. First, it establishes a realistic performance floor for short-term forecasting, given that meteorological conditions demonstrate temporal autocorrelation over hourly timescales. Second, the model's inherent simplicity ensures interpretability and facilitates transparent communication of forecast skill relative to a naive forecasting strategy. Third, the absence of training requirements or parameter tuning enables straightforward implementation while providing a computationally efficient reference point for benchmarking more sophisticated models.

6.1.3 Bidirectional LSTM

Bidirectional Long Short-Term Memory (BiLSTM) networks have demonstrated robust performance in time series forecasting applications through their capacity to capture temporal dependencies in both forward and backward directions across sequential data.

The BiLSTM architecture employed in this research is configured to generate simultaneous forecasts of wind speed and direction for all 32 measurement stations over the 12-hour prediction horizon. The model utilizes the dataset described in Chapter 5, with wind direction encoded as the sine and cosine of the direction angle to circumvent the *angular discontinuity problem* at 0°/360°. Alternative representations using Cartesian wind components (u and v) were evaluated during preliminary experiments but yielded inferior performance, leading to the adoption of the trigonometric encoding approach.

The network architecture comprises four main components. First, a *station embedding layer* learns a unique dense vector representation for each of the 32 measurement stations. These embeddings capture location-specific characteristics, including local

climate patterns and geographical context, and are concatenated with the meteorological input features at each temporal step.

Second, a *bidirectional LSTM layer* processes the augmented input sequence. The bidirectional architecture enables the network to extract temporal dependencies from both past-to-present (forward) and present-to-past (backward) directions, thereby providing richer contextual information compared to unidirectional processing. The forward and backward hidden states at the final timestep are concatenated to construct a comprehensive representation of the complete input sequence.

Third, a stack of *fully connected (feedforward) layers* with ReLU activation functions and dropout regularization transforms the concatenated LSTM output. These layers facilitate the learning of complex non-linear relationships between historical meteorological conditions and future wind patterns. Dropout regularization is employed to mitigate overfitting through the random deactivation of neurons during the training process.

Fourth, a final *linear projection layer* maps the learned representation to the output space, generating predictions for all 12 future timesteps simultaneously through direct multi-step forecasting rather than iterative autoregressive prediction.

Model training follows a supervised learning paradigm with MSE serving as the loss function, optimized using the Adam algorithm. Key hyperparameters encompass the LSTM hidden dimension, number of LSTM layers, dropout rate, learning rate, batch size, and station embedding dimensionality. Hyperparameter tuning is performed through manual experimentation rather than systematic grid search or Bayesian optimization, prioritizing computational efficiency in light of the substantial dataset size and extended training durations. Various configurations of sequence length, hidden dimensions, layer count, dropout rates, learning rates, and batch sizes are assessed based on validation set performance. The specific hyperparameter values adopted for the final model configuration are documented in Chapter 7.

6.1.4 PatchTST

The PatchTST architecture employed in this research is configured to generate simultaneous forecasts of wind speed and direction for all 32 measurement stations over the 12-hour prediction horizon, analogous to the BiLSTM approach. The model utilizes the dataset described in Chapter 5, with wind direction encoded as the sine and cosine of the direction angle to circumvent the *angular discontinuity problem* at 0°/360°. Alternative representations using Cartesian wind components (u and v) were evaluated during preliminary experiments but yielded inferior performance, leading to the adoption of the trigonometric encoding approach.

The network architecture comprises five principal components. First, a *patch embedding layer* segments the continuous input sequence into overlapping temporal patches of fixed length (e.g., 6 timesteps). Each patch is projected into a latent feature space through a linear transformation. This patching mechanism reduces the effective sequence length, thereby improving computational efficiency while enabling the model to process local temporal patterns as higher-level semantic tokens.

Second, a *station embedding layer* learns a unique dense vector representation for each of the 32 measurement stations, capturing location-specific characteristics. This embedding is prepended to the sequence of patch embeddings as a special "station token," enabling the model to condition its forecasts on station identity.

Third, the resulting token sequence is processed by a stack of *Transformer encoder layers*. Each encoder layer comprises multi-head self-attention mechanisms and position-wise feedforward sublayers. The self-attention mechanism enables the model to learn dependencies between different temporal segments across the entire input history, capturing both short-term and long-term relationships without the sequential processing constraints inherent in recurrent architectures such as LSTMs. This parallel processing capability constitutes a key advantage of the Transformer architecture.

Fourth, the final hidden representation of the station token serves as a condensed summary of the complete input sequence. This summary is processed through a *projection head* comprising linear and non-linear transformations with dropout regularization.

Fifth, the projection head maps the learned representation to the output space, generating predictions for all 12 future timesteps simultaneously through direct multi-step forecasting.

Model training follows a supervised learning paradigm with MSE serving as the loss function, optimized using the AdamW algorithm, which incorporates weight decay for enhanced regularization. Optional temporal smoothness regularization can be applied to encourage stable forecasts and prevent abrupt discontinuities in predicted trajectories. Key hyperparameters encompass patch length, patch stride, number of Transformer encoder layers, embedding dimension, number of attention heads, and dropout rates. These hyperparameters are tuned through manual experimentation, prioritizing computational efficiency. The specific hyperparameter values adopted for the final model configuration are documented in Chapter 7.

6.1.5 Model Evaluation

Model evaluation is performed using a time series forecasting protocol applied to the held-out test partition spanning from April 1, 2025, through October 5, 2025. Evaluation timestamps are selected at 12-hour intervals for each of the 32 measurement stations, yielding multiple independent forecast instances per station. At each evaluation timestamp t, the model processes 24 hours of historical observations and generates predictions for the subsequent 12-hour period.

Forecast accuracy is quantified by comparing predicted values against actual observed measurements at each forecast horizon step. For wind speed predictions, two error metrics are computed: MAE and RMSE, both expressed in meters per second. For wind direction predictions, *circular angular error* is calculated to appropriately account for the periodic nature of angular measurements, ensuring correct handling of the 0°/360° wraparound. This circular error metric determines the minimum angular distance between predicted and observed directions, thereby avoiding artificially inflated errors near the angular discontinuity.

Performance metrics are aggregated across all evaluation timestamps and stations to derive overall model performance statistics. Additionally, metrics are analyzed as a function of forecast horizon to characterize how prediction accuracy degrades with increasing lead time.

6.2 Extrapolation and Power Production Estimation

6.2.1 Horizontal Spatial Interpolation

Horizontal spatial interpolation is employed to transfer wind measurements from the 32 DWD measurement stations to the geographical locations of individual wind turbines distributed across Brandenburg. This approach operates under the assumption that each turbine is spatially surrounded by multiple measurement stations, thereby enabling interpolation through proximity-weighted averaging.

IDW is selected as the interpolation method based on its computational simplicity, efficiency, and established application in meteorological contexts. As detailed in Chapter 2, IDW estimates values at unmeasured locations as weighted averages of surrounding observations, with weights inversely proportional to the distance from each observation point. To appropriately handle the circular nature of wind direction, interpolation is performed on the Cartesian wind components u (east-west) and v (north-south) rather than directly on direction angles. Following interpolation, these components are converted back to wind speed and direction at each turbine location.

This approach rests on several simplifying assumptions:

- Brandenburg's predominantly flat topography [43] permits the neglect of orographic effects that would otherwise necessitate terrain-aware interpolation methods.
- Terrain roughness and local land cover variations are not explicitly modeled, representing a potential source of interpolation error that could be addressed in future research.
- All measurement stations and turbine locations are assumed to be at equivalent ground elevations for the purpose of horizontal interpolation.

The IDW power parameter (commonly denoted β) is set to 2, a standard value that balances spatial smoothness with sensitivity to nearby observations. Optimal tuning of this parameter would require validation against actual turbine-level wind measurements, which are unavailable for this research. Future work incorporating ground-truth data could refine this parameter through empirical comparison.

6.2.2 Vertical Extrapolation to Hub Height

Following horizontal interpolation, wind speeds at turbine locations remain at the measurement height of approximately 10 meters above ground level [10]. Given that wind turbines operate with tower heights usually around 150 meters [16], vertical extrapolation is required to estimate wind speeds at the turbine's operational altitude. The turbine-specific hub heights required for this calculation are obtained from the MaStR metadata.

The power law method is employed for vertical wind speed extrapolation. This empirical relationship models the vertical increase in wind speed with height above ground, accounting for surface friction effects that diminish with increasing altitude. The power law is selected in preference to alternative methods, such as logarithmic profiles, based on its simplicity, widespread application in wind energy contexts, and demonstrated reliability for flat terrain conditions, as confirmed by [36].

Application of the power law requires specification of a wind shear exponent, which depends on surface roughness and atmospheric stability. For this research, a standard exponent value of 0.14 is employed. It should be noted that this represents a commonly used approximation rather than a site-optimized parameter, and may not be optimal for all turbine locations [26].

6.2.3 Power Production Estimation

The final step converts hub-height wind speeds into electrical power output by applying turbine-specific power curves. To associate turbines from the MaStR registry with their corresponding power curves in The Wind Power database, manufacturer and model designations must be linked across the two data sources. This matching process employs fuzzy string matching to accommodate variations in naming conventions, abbreviations, and formatting inconsistencies between the datasets. Of the 4,612 unique turbines in the Brandenburg dataset, 3,341 turbines (72.44%) were successfully matched to specific power curves, while 1,271 turbines (27.56%) could not be matched due to nomenclature discrepancies or database incompleteness.

For turbines lacking manufacturer-specific power curves due to unsuccessful matching, a generic power curve (described in Section 2.1) derived from typical turbine characteristics is applied. This generic curve is scaled proportionally according to the turbine's rated capacity obtained from the MaStR metadata, thereby providing a reasonable approximation in the absence of model-specific data.

Given that power curve data specifies output at discrete 0.5 m/s wind speed increments, the extrapolated hub-height wind speed is rounded to the nearest 0.5 m/s interval to retrieve the corresponding power output value. While this discretization introduces minor quantization error, it maintains computational efficiency.

The power production estimation methodology rests on several simplifying assumptions:

- The turbine's yaw mechanism optimally aligns the rotor perpendicular to the wind direction at all times, thereby maximizing energy capture. While yaw misalignment can reduce production in practice, modern turbines employ active yaw control systems to minimize this effect.
- Wind conditions remain steady and spatially uniform across the rotor swept area during each hourly measurement interval. This assumption neglects turbulence, wind shear variation across the rotor disk, and short-term gusts that may induce instantaneous production fluctuations.
- Wind flow to each turbine is unobstructed, thereby neglecting wake effects from upstream turbines and blockage from other obstacles such as buildings or vegetation. In reality, wake effects can substantially reduce downstream turbine production in wind farm configurations, particularly when turbines are closely spaced.

- The spatially and vertically extrapolated wind speeds accurately represent conditions at the turbine hub, despite inherent uncertainties in the interpolation and extrapolation methodologies.

Under these assumptions, the methodology generates power production estimates for each turbine at hourly resolution. These estimates can be aggregated spatially across multiple turbines and temporally over extended periods to characterize regional wind power generation patterns.

Chapter 7

Results

7.1 Weather Forecasting

Model evaluation is conducted on test data covering the period from April 1, 2025, through October 5, 2025, representing approximately six months of hourly measurements from weather stations. Forecasts are generated at 12-hour intervals for all stations, utilizing 24 hours of historical observations to predict wind speed and direction over the subsequent 12-hour horizon. At each prediction point, the forecasted 12-hour sequence is compared against the corresponding observed values to compute performance metrics including MAE and RMSE. This evaluation framework provides a comprehensive assessment of forecasting accuracy across diverse spatial locations and temporal conditions within the test period.

7.1.1 Persistence Model

The persistence model establishes a baseline reference for evaluating the performance of machine learning approaches. On the test set, this model attained a mean speed error of 1.4313 m/s, an RMSE of 1.8547 m/s, and a mean direction error of 42.79 degrees. Performance metrics across different forecast horizons are presented in Figure 7.1, while representative forecasting examples are illustrated in Figure 7.2. Table 7.1 provides detailed statistics disaggregated by horizon step, demonstrating how prediction errors increase progressively with forecast lead time.

7.1.2 Bidirectional LSTM Model

The Bidirectional LSTM model is configured to process 24 hours of historical observations for generating predictions over the subsequent 12-hour forecast horizon. The network architecture comprises 2 bidirectional LSTM layers with a hidden dimen-

Table 7.1 Statistics by horizon step for the persistence model

Horizon Step	Speed Error		Direction Error	
	Mean (m/s)	Std (m/s)	Mean (°)	Std (°)
1	0.4916	0.4864	20.4125	30.3464
2	0.7677	0.7048	27.0083	34.5686
3	1.0337	0.8735	33.1789	38.1531
4	1.2663	1.0021	37.9813	39.8916
5	1.4370	1.0872	41.7086	41.2293
6	1.5688	1.1434	44.7899	42.6639
7	1.6765	1.1875	46.6071	43.1488
8	1.7468	1.2303	48.7419	44.1532
9	1.7923	1.2636	50.4871	44.4874
10	1.8071	1.2757	51.9901	45.0888
11	1.8031	1.2924	54.4122	46.1606
12	1.7848	1.2906	56.1346	47.0300

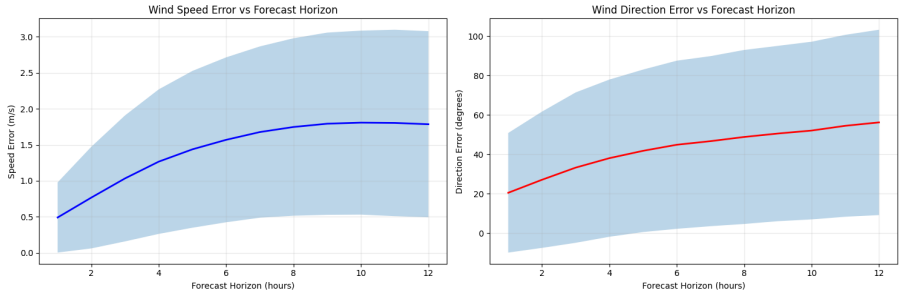

Fig. 7.1 Metrics of the persistence model

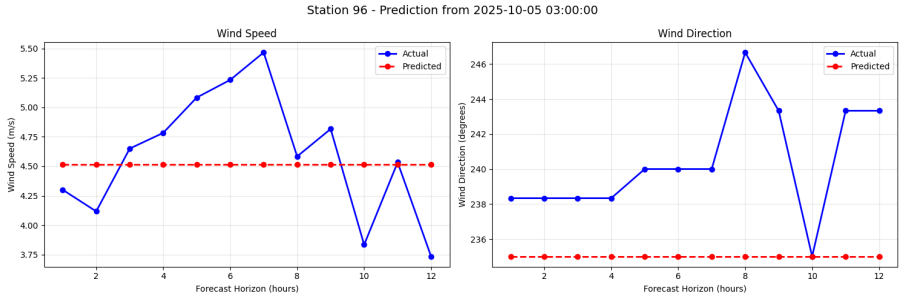

Fig. 7.2 Examples of the persistence model

sion of 128 units, supplemented by station-specific embeddings of dimension 16 to capture spatial heterogeneity across measurement locations. Dropout regularization with a rate of 0.3 is employed to mitigate overfitting during training. Model training is conducted over 100 epochs with a batch size of 256 and a learning rate of 0.0003. Table 7.2 provides the complete hyperparameter configuration.

The Bidirectional LSTM model exhibits substantial performance improvements relative to the persistence baseline, attaining a mean speed error of 0.8817 m/s, an RMSE of 1.1861 m/s, and a mean direction error of 38.73 degrees.

Performance metrics across different forecast horizons are visualized in Figure 7.3, while representative prediction examples are presented in Figure 7.4. Table 7.3 provides horizon-specific error statistics, revealing the model's capacity to maintain relatively stable prediction accuracy throughout the 12-hour forecast period.

Table 7.2 Hyperparameters of the Bidirectional LSTM model

Parameter	**Value**
Model Hyperparameters	
History steps	24
Horizon steps	12
hidden_size	128
num_layers	2
station_embed_dim	16
dropout	0.3
Training Configuration	
batch_size	256
learning_rate	0.0003
num_epochs	100
val_split	0.15
early_stop_patience	15
early_stop_min_delta	0.0001

Table 7.3 Statistics by horizon step for the Bidirectional LSTM model

Horizon Step	**Speed Error**		**Direction Error**	
	Mean (m/s)	**Std (m/s)**	**Mean (°)**	**Std (°)**
1	0.4765	0.4575	20.6423	28.6225
2	0.6338	0.5642	26.2435	32.7185
3	0.7285	0.6192	30.8827	35.7348
4	0.8086	0.6780	34.6031	37.5365
5	0.8649	0.7330	37.3199	38.8564
6	0.9118	0.7742	40.1390	40.9498
7	0.9625	0.8299	41.5833	41.6009
8	0.9946	0.8573	43.5906	42.8673
9	1.0255	0.8832	45.1986	43.2259
10	1.0411	0.8958	46.6903	44.0838
11	1.0612	0.9093	48.2539	44.7179
12	1.0710	0.9086	49.5973	44.9969

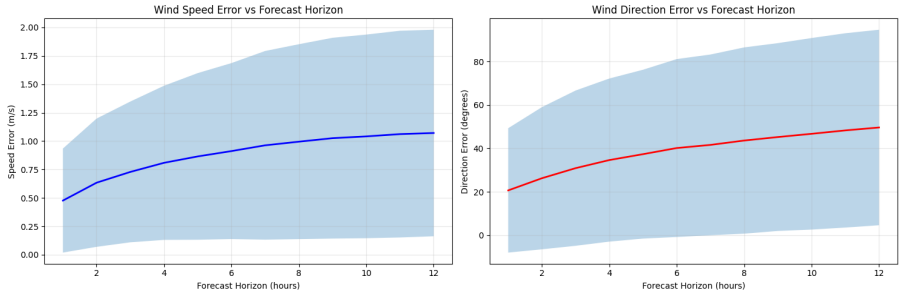

Fig. 7.3 Metrics of the Bidirectional LSTM model

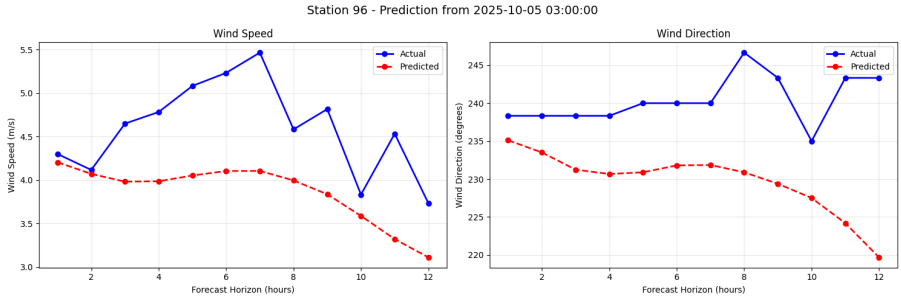

Fig. 7.4 Examples of the Bidirectional LSTM model

7.1.3 PatchTST Model

The PatchTST model employs a transformer-based architecture that partitions the 24-hour input sequence into patches of length 6 with a stride of 3, thereby reducing computational complexity while preserving temporal patterns. The network architecture comprises 2 transformer layers with 4 attention heads and a model dimension of 128. Model training is conducted over 20 epochs with a batch size of 128 and a learning rate of 0.0005. Auxiliary loss terms are incorporated during training, including a differential loss weight of 0.2 and a total variation loss weight of 0.05, to promote smooth predictions. Table 7.4 presents the complete hyperparameter configuration.

The PatchTST model attains a mean speed error of 0.9097 m/s, an RMSE of 1.2187 m/s, and a mean direction error of 39.10 degrees. Performance across different forecast horizons is illustrated in Figure 7.5, while representative prediction examples are provided in Figure 7.6. Table 7.5 presents detailed horizon-wise error statistics, revealing how prediction accuracy evolves throughout the 12-hour forecast window.

7.2 Interpolation

Spatial interpolation of wind measurements from weather station locations to wind turbine sites is performed to generate spatially distributed wind data across the study area. This process yields a total of 128,216,264 interpolated values over the analysis period spanning April 1, 2025, through October 5, 2025.

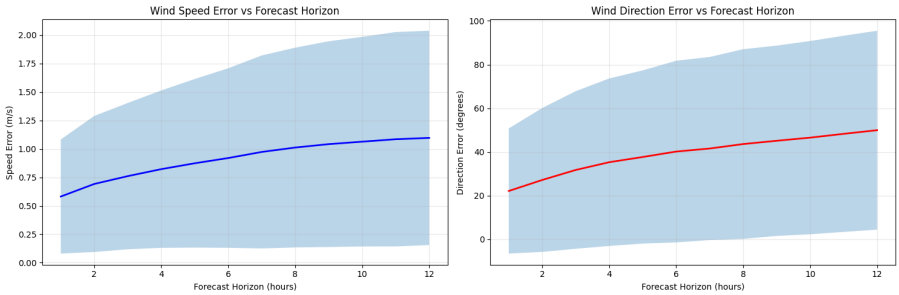

Fig. 7.5 Metrics of the PatchTST model

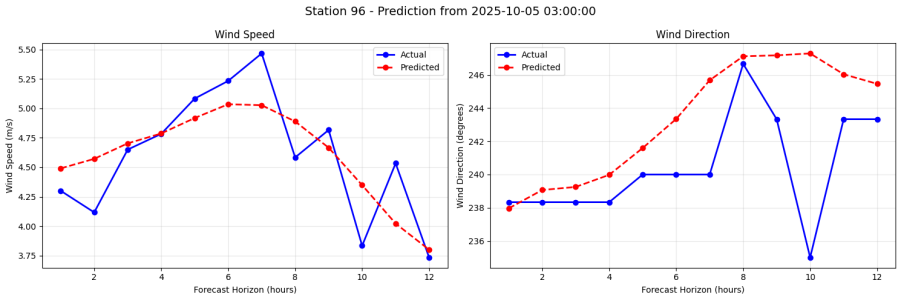

Fig. 7.6 Examples of the PatchTST model

Table 7.4 Hyperparameters of the PatchTST model

Parameter	**Value**
Model Hyperparameters	
History steps	24
Horizon steps	12
d_model	128
nhead	4
num_layers	2
dim_feedforward	256
patch_len	6
stride	3
dropout	0.1
Training Configuration	
batch_size	128
learning_rate	0.0005
num_epochs	20
val_split	0.2
early_stop_patience	4
early_stop_min_delta	0.0001
diff_loss_weight	0.2
tv_loss_weight	0.05

Table 7.5 Statistics by horizon step for the PatchTST model

Horizon Step	Speed Error		Direction Error	
	Mean (m/s)	Std (m/s)	Mean (°)	Std (°)
1	0.5813	0.5008	22.1022	28.6519
2	0.6916	0.5971	27.1560	32.9468
3	0.7604	0.6423	31.7353	36.0871
4	0.8217	0.6909	35.2880	38.3464
5	0.8732	0.7400	37.6746	39.6575
6	0.9193	0.7880	40.1612	41.5957
7	0.9726	0.8471	41.5573	41.9179
8	1.0115	0.8769	43.6023	43.4410
9	1.0413	0.9032	45.1048	43.5736
10	1.0629	0.9201	46.5412	44.2518
11	1.0843	0.9412	48.2774	44.9187
12	1.0962	0.9408	49.9513	45.5436

7.2.1 Wind Speed Evaluation

The interpolated wind speed values span a range from 0.00 m/s to 15.36 m/s, with a mean of 2.82 m/s and a median of 2.58 m/s. All interpolated values remain within the absolute valid range of 0.0 to 30.0 m/s, with 0 values (0.00%) detected outside this range. However, 44,153,904 values (34.44%) fall outside the typical operational range of 2.0 to 15.0 m/s. No NaN or infinite values are present in the wind speed data.

Descriptive statistics for raw measurements, interpolated wind speeds at measurement height, and extrapolated wind speeds at hub height are presented in Table 7.6. Wind speed distributions across these three data processing stages are illustrated in Figure 7.7.

Table 7.6 Wind speed statistics for raw measurements, interpolated data, and hub height extrapolation

Statistic	Raw Measurements	Interpolated (10m)	Hub Height
Mean (m/s)	3.04	2.82	3.94
Standard Deviation (m/s)	1.72	1.56	2.19
Median (m/s)	—	2.58	3.60
Minimum (m/s)	—	0.00	0.00
Maximum (m/s)	—	15.36	22.34

7.2.2 Extreme Values Analysis

A total of 44,153,904 values are identified outside the typical operational range of 2.0 to 15.0 m/s. Among these extreme values, 25 exceed 15.0 m/s, exhibiting a maximum of 15.36 m/s and a mean of 15.19 m/s. The majority of extreme values, comprising 44,153,879 observations, fall below 2.0 m/s, with a minimum of 0.00 m/s and a mean of 1.26 m/s.

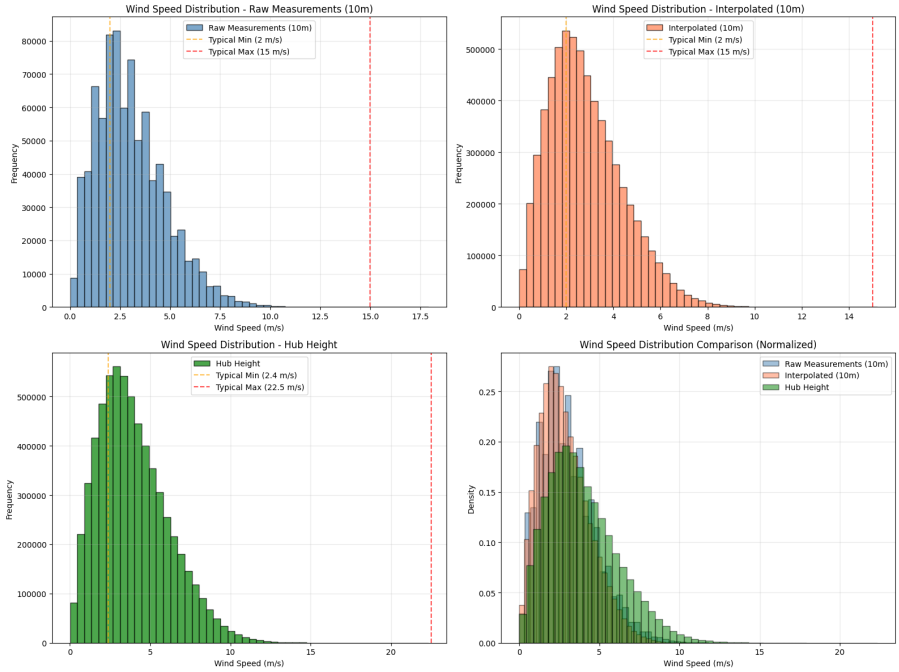

Fig. 7.7 Wind speed distribution comparison across raw measurements (10m), interpolated data (10m), and hub height extrapolation, showing frequency distributions and normalized density comparison with typical operational range boundaries

7.2.3 Wind Direction Evaluation

The interpolated wind direction values span a range from 0.00 degrees to 360.00 degrees, with a mean of 208.35 degrees and a median of 233.79 degrees. All values remain within the valid range of 0.0 to 360.0 degrees, with 0 values (0.00%) detected outside this range. No NaN or infinite values are present in the wind direction data.

Descriptive statistics for both interpolated and raw measurement wind directions are presented in Table 7.7.

Table 7.7 Wind direction statistics for interpolated and raw measurement data

Statistic	**Raw Measurements**	**Interpolated**
Mean (degrees)	209.91	208.34
Standard Deviation (degrees)	93.79	93.34
Median (degrees)	—	233.79
Minimum (degrees)	—	0.00
Maximum (degrees)	—	360.00

7.3 Extrapolation

Wind speed values are vertically extrapolated from measurement height (10m) to turbine hub height using the power law method. A total of 128,216,264 values are processed through this vertical transformation over the analysis period spanning April 1, 2025, through October 5, 2025.

7.3.1 Wind Speed at Hub Height Evaluation

The extrapolated wind speed values at hub height span a range from 0.00 m/s to 22.34 m/s, with a mean of 3.94 m/s and a median of 3.60 m/s. All extrapolated values remain within the absolute valid range of 0.0 to 45.0 m/s, with 0 values (0.00%) detected outside this range. However, 34,486,727 values (26.90%) fall outside the typical operational range of 2.4 to 22.5 m/s. No NaN or infinite values are present in the hub height wind speed data.

Consistency validation confirms that hub height wind speed values are greater than or equal to measurement height wind speed values in all cases.

Descriptive statistics comparing wind speeds at measurement height and hub height are presented in Table 7.6, while the corresponding wind speed distributions are illustrated in Figure 7.7.

7.3.2 Speed Increase Analysis

Vertical extrapolation from measurement height to hub height yields a mean speed increase ratio of 1.3966, corresponding to a mean absolute increase of 1.12 m/s.

Chapter 8

Implementation

8.1 System Architecture

The prototype system implements the methodology described in Chapter 6 through a *client-server architecture* comprising a web-based frontend and a computational backend. This architectural separation allows the frontend to concentrate on data visualization and user interaction, while the backend manages computationally intensive operations including meteorological data retrieval, forecasting model execution, and spatial extrapolation.

The frontend employs Next.js, a React-based web framework that delivers server-side rendering and optimized performance for interactive applications [47]. User interface components and data visualization charts are constructed using shadcn, a component library that ensures consistent styling and responsive design [48]. Interactive geographical visualization of turbine locations is realized through the Google Maps JavaScript SDK, accessed via the @vis.gl/react-google-maps package, which provides React bindings for efficient map rendering and interaction handling [34].

The backend employs FastAPI, a modern Python web framework selected for its asynchronous request handling capabilities and automatic API documentation generation [39]. The backend orchestrates three primary computational pipelines: meteorological data acquisition from DWD, wind forecasting utilizing the trained deep learning models, and power production estimation through spatial extrapolation. The system maintains both historical approximations for the past 12 hours and forecasts for the subsequent 12 hours across all turbines, enabling users to analyze recent production patterns alongside anticipated generation.

8.2 Data Processing Pipelines

The backend executes three sequential data processing pipelines at hourly intervals. This regular execution schedule ensures that the system maintains up-to-date production estimates and forecasts based on the most recently available meteorological observations. Each pipeline stage relies on the output of preceding stages, establishing a coordinated workflow that progresses from raw data acquisition to final production estimates.

8.2.1 Meteorological Data Acquisition

The first pipeline retrieves the most recent meteorological measurements from the DWD open data portal. Specifically, both the *recent* dataset (covering the past 500 days, updated daily) and the *now* dataset (covering the past 24 hours, updated hourly) are fetched.

Retrieved measurements are processed and *upserted* (inserted or updated) into the system database. The upsert operation is necessary because measurements in the DWD *recent* and *now* datasets remain subject to revision as quality control procedures continue. By updating existing records when newer versions become available, the system ensures that subsequent calculations utilize the most accurate measurements, even as DWD applies retrospective corrections.

8.2.2 Wind Condition Forecasting

The second pipeline applies the trained forecasting model to generate predictions of wind speed and direction across all measurement stations. For each station, the pipeline retrieves the most recent 24 hours of meteorological observations from the database to construct the input sequence required by the forecasting models.

The selected forecasting model (either BiLSTM or PatchTST, as described in Chapter 7) generates predictions for wind speed and direction at each station over the subsequent 12-hour horizon at 1 hour resolution. These forecasts are subsequently upserted into the database, where they serve as input for the production estimation pipeline.

8.2.3 Spatial Extrapolation and Power Production Estimation

The third pipeline implements the spatial extrapolation and power estimation methodology detailed in Section 6.2. This pipeline processes both historical measurements (past 12 hours) and forecasted values (future 12 hours) across all measurement

stations, enabling the system to deliver both retrospective production approximations and prospective forecasts.

For each turbine and timestep, the pipeline executes five sequential operations. First, the forecasted or historical wind speed and direction from the models are transformed into u and v wind components to facilitate spatial interpolation. Second, horizontal spatial interpolation transfers these wind components from the 32 stations to the turbine's geographical location through IDW applied to the u and v components. Third, the interpolated u and v components are transformed back to wind speed and direction. Fourth, vertical extrapolation scales the wind speed from the 10-meter measurement height to the turbine's hub height through the power law method. Fifth, the turbine-specific power curve converts the hub-height wind speed into electrical power output. All of these steps are described in detail in Chapter 6.

The resulting production estimates are upserted into the database with metadata flags that distinguish between historical approximations (derived from actual measurements) and forecasts (derived from predicted meteorological conditions). When both forecast-based and measurement-based estimates exist for the same timestamp, the measurement-based estimate takes precedence, ensuring that historical records reflect actual observations rather than predictions once measurements become available.

8.3 Web Dashboard

The web-based dashboard delivers an interactive interface for visualizing wind power production estimates and forecasts across Brandenburg. The interface comprises two primary views: a regional overview presenting aggregated metrics for all turbines, and individual turbine detail pages accessible through interactive map navigation.

The overview page constitutes the entry point to the dashboard, presenting key summary statistics at the top of the interface, as shown in Figure 8.1. These statistics encompass current aggregated production across all turbines, total number of active turbines in the system, and the number of meteorological measurement stations providing data. Below the summary statistics, a time series chart visualizes aggregated production for all turbines, displaying both historical approximations for the past 12 hours and forecasts for the subsequent 12 hours. This dual temporal perspective enables users to assess recent generation patterns alongside anticipated near-term production.

An interactive map occupies the lower portion of the overview page, displaying the geographical distribution of all wind turbines across Brandenburg, as shown in

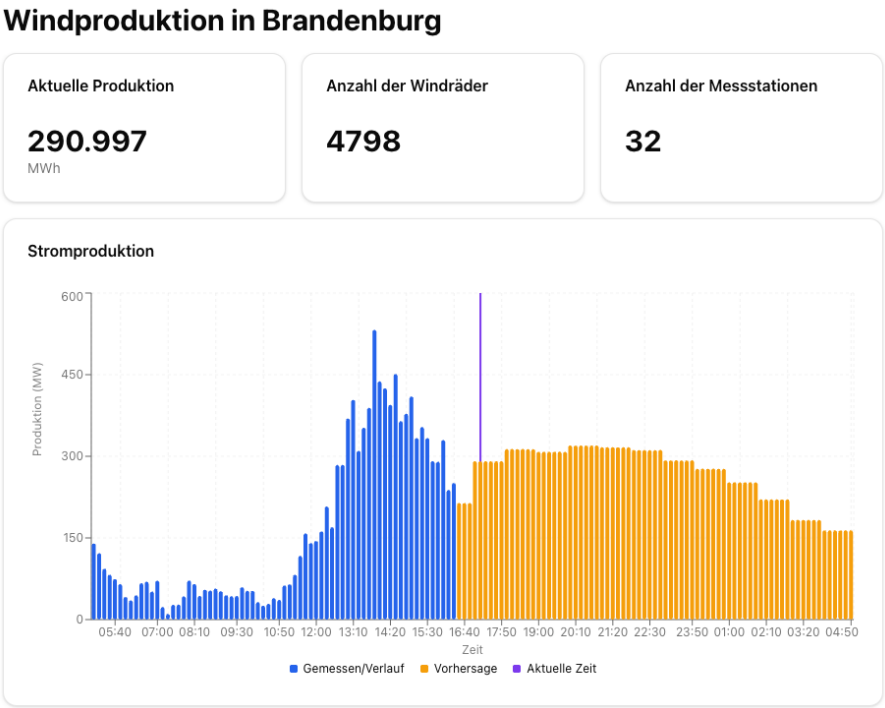

Fig. 8.1 Landing page of the website

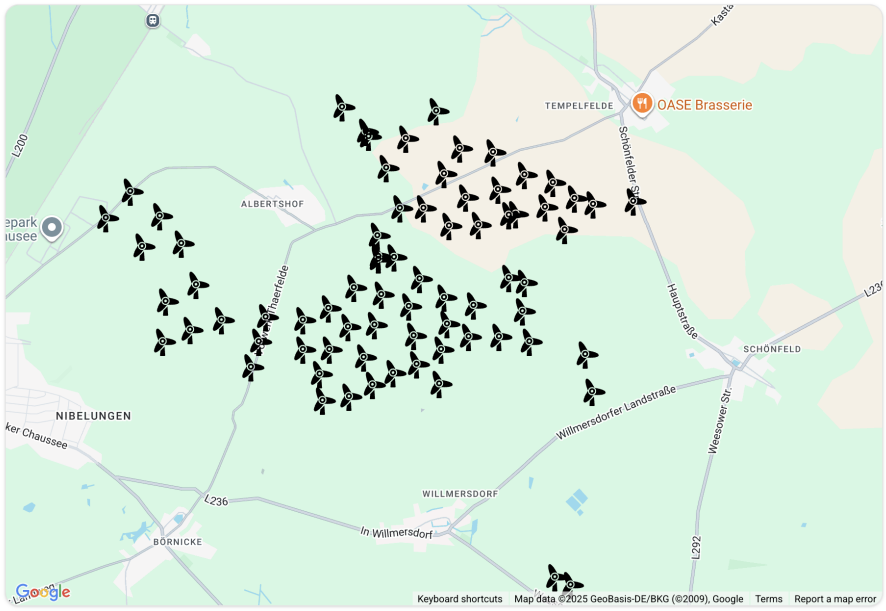

Fig. 8.2 Map view of the website

Figure 8.2. Each turbine is represented as a clickable marker, and selecting a turbine navigates the user to a dedicated detail page for that installation.

The turbine detail page delivers comprehensive information for individual turbines. Summary statistics specific to the selected turbine are displayed at the top of the page, including current production, maximum observed production, and minimum production over the displayed time window. A time series chart analogous to that on the overview page presents production trajectories for the selected turbine, spanning 12 hours of historical data and 12 hours of forecasts. Below the chart, all metadata fields retrieved from the MaStR are displayed, providing technical specifications such as manufacturer, model, hub height, rotor diameter, and rated capacity.

8.4 Data Licensing and Attribution

The implementation adheres to all applicable data licensing requirements for the publicly accessible datasets employed in this research.

Meteorological measurements from DWD are licensed under the Creative Commons Attribution 4.0 (CC BY 4.0) license [11]. This license permits use, modification, and distribution of the data, provided that appropriate credit is attributed to the Deutscher Wetterdienst as the data source. Accordingly, the dashboard interface incorporates attribution to DWD for all meteorological measurements.

Wind turbine metadata from the Marktstammdatenregister is licensed under "Datenlizenz Deutschland - Namensnennung - Version 2.0" (Data License Germany - Attribution - Version 2.0) [29]. This license similarly permits use and redistribution, provided that credit is attributed to the Bundesnetzagentur and that the original data source URL is referenced. The dashboard accordingly attributes turbine metadata to the MaStR and incorporates references to the source portal.

Power curve data from The Wind Power database is subject to commercial licensing terms, as the dataset constitutes a proprietary product available for purchase. The specific licensing conditions governing redistribution and publication of this data are contingent upon the terms negotiated at the time of acquisition.

Chapter 9

Discussion

9.1 Interpretation

The empirical results reveal that machine learning methods significantly surpass the persistence baseline in forecasting accuracy, thereby validating the theoretical framework established in Chapter 6. This section examines how the observed findings relate to fundamental concepts in time series forecasting, including temporal dependencies and architectural considerations in model design.

9.1.1 Persistence Model Performance and Temporal Autocorrelation

Employing a simple forward-propagation strategy that replicates the most recent observation, the persistence model yielded a mean speed error of 1.4313 m/s and an RMSE of 1.8547 m/s over the 12-hour prediction window. As elaborated in Section 2.3, this baseline approach capitalizes on *temporal autocorrelation* inherent in atmospheric phenomena, whereby current meteorological conditions tend to persist into the near future. Table 7.1 documents a systematic decline in forecast quality, with speed error escalating from 0.4916 m/s at the initial forecast step to 1.7848 m/s at the final step, providing empirical evidence for the anticipated deterioration of persistence-based predictions as lead time increases.

The direction error, which increases from 20.41° at the first horizon step to 56.13° at the twelfth step, exhibits even more pronounced degradation than wind speed. This behavior aligns with the meteorological understanding that wind direction exhibits greater short-term variability than wind speed magnitude, making it inherently more difficult to predict through simple extrapolation. The persistence model's inability

to anticipate directional shifts highlights the limitations of *zero-order forecasting* approaches for variables exhibiting high temporal variability.

9.1.2 Machine Learning Model Superiority and Long-Term Dependencies

The Bidirectional LSTM and PatchTST architectures both delivered markedly superior forecasting accuracy relative to the persistence baseline, thereby confirming the theoretical advantages of machine learning systems designed to extract complex temporal relationships. Specifically, the BiLSTM model recorded a mean speed error of 0.8817 m/s (representing a 38.4% improvement over persistence) alongside an RMSE of 1.1861 m/s (36.1% improvement), whereas the PatchTST model attained 0.9097 m/s (36.4% improvement) and 1.2187 m/s (34.3% improvement) for the respective metrics. These considerable gains provide empirical evidence that both architectures successfully identified temporal dependencies extending beyond elementary autocorrelation, thereby capturing intricate patterns embedded within historical observations.

The superior performance of the BiLSTM model can be attributed to its distinctive architectural characteristics. Through its *cell state* mechanism—regulated by forget, input, and output gates — the LSTM architecture facilitates selective preservation of pertinent historical information while eliminating noise and irrelevant fluctuations. This selective memory proves especially beneficial for wind forecasting applications, given that atmospheric dynamics emerge from multiscale temporal interactions spanning turbulent short-term variations, diurnal oscillations, and extended synoptic weather regimes. The bidirectional processing strategy amplifies these capabilities by analyzing sequences in both temporal directions, thereby enabling the model to contextualize each observation using information from both antecedent and subsequent time steps throughout the training phase. Such bidirectional contextualization offers particular advantages when extracting patterns from historical sequences that manifest both forward-looking causal relationships and backward-looking dependencies.

Despite achieving marginally lower accuracy than the BiLSTM, the PatchTST model exhibits competitive forecasting performance through a fundamentally distinct architectural paradigm. The PatchTST approach partitions the input sequence into subseries-level *patches*, thereby compressing the token count from 24 individual hourly observations to approximately 8 patches (utilizing a patch length of 6 with a stride of 3). While this patching strategy enhances computational efficiency by mitigating the quadratic complexity inherent in attention mechanisms, it potentially

compromises fine-grained temporal resolution when compared to the sequential processing characteristic of LSTM architectures. The competitive accuracy achieved by PatchTST indicates that the localized semantic information retained within individual patches adequately captures the predominant patterns governing wind dynamics, although the modest performance differential suggests that the enhanced temporal granularity afforded by LSTM processing confers incremental predictive benefits for this specific forecasting application.

9.1.3 Forecast Horizon Behavior and Error Accumulation

Analysis of horizon-specific error statistics exposes divergent patterns in how forecast uncertainty accumulates across the three modeling approaches. The persistence baseline manifests approximately *linear error progression*, with speed errors escalating from 0.4916 m/s at step 1 to 1.7848 m/s at step 12, representing an average increase of roughly 0.15 m/s per forecast step during the initial six-hour window before stabilizing. This linear trajectory stems directly from the constant-velocity assumption underlying persistence methods, wherein each successive time step compounds forecast uncertainty in the absence of corrective mechanisms.

Conversely, both machine learning architectures exhibit *sub-linear error progression*, with the BiLSTM displaying notably consistent accuracy throughout the prediction horizon. Speed errors for the BiLSTM advance from 0.4765 m/s at the initial step to merely 1.0710 m/s at the twelfth step, with the error trajectory remaining remarkably stable between steps 8 and 12. This pattern indicates that the model has acquired the capacity to generate internally coherent predictions spanning the entire forecast window, rather than producing isolated estimates at individual time steps. The recurrent structure inherent to LSTM architectures naturally generates multi-step forecasts through sequential propagation, whereby each predicted time step depends on preceding predictions, thereby facilitating the learning of smooth, physically realistic trajectories instead of discontinuous point estimates.

The PatchTST architecture manifests comparable sub-linear error progression, albeit with modestly elevated errors at extended forecast horizons (reaching 1.0962 m/s at step 12). Through its self-attention mechanism, the transformer architecture establishes direct relationships between all forecast time steps and all input observations, which likely accounts for its capacity to preserve forecast coherence across the prediction window. Nevertheless, the absence of explicit sequential constraints may permit slightly greater prediction variability at distant forecast steps, thereby contributing to the marginal performance differential relative to the BiLSTM approach.

9.1.4 Wind Direction Forecasting and Vector Component Representation

Directional errors recorded across all forecasting approaches — 42.79° for persistence, 38.73° for BiLSTM, and 39.10° for PatchTST — underscore the substantial challenges inherent in angular prediction tasks. Although the machine learning architectures realize modest gains over the persistence baseline (approximately 9.5% error reduction), the absolute magnitudes of directional errors remain considerable. This constrained improvement derives partially from the circular topology of angular measurements and the inherent difficulty of extracting meaningful patterns when minor numerical variations in degrees can signify substantial physical divergences.

9.1.5 Implications for Wind Power Production Estimation

The forecasting accuracy attained by the machine learning architectures carries significant ramifications for downstream power production estimation processes. As characterized by the power curve, wind turbine power output exhibits a nonlinear relationship with wind speed. The BiLSTM's speed error of approximately 0.88 m/s introduces considerable uncertainty when propagated through the power curve, especially within the *transitional region* between cut-in and rated wind speeds where power output exhibits the steepest gradients. In contrast, at extremely low speeds (below cut-in threshold) or extremely high speeds (above rated power), the identical speed error would yield smaller relative power errors owing to the flatter segments of the power curve. This nonlinear error propagation emphasizes the critical importance of precise wind speed forecasting.

While directional errors remain substantial in absolute magnitude, they may exert limited influence on power estimation for individual turbines under the idealized assumption of perfect yaw alignment. Nevertheless, when applying interpolation to transfer wind measurements from measurement stations to turbine locations, directional errors can affect the spatial coherence of the wind field, potentially introducing systematic biases in regional production estimates.

9.1.6 Spatial Interpolation Reasonability

Applying IDW to transfer wind measurements from weather stations to turbine locations generated 128,216,264 interpolated values throughout the analysis period, as detailed in Section 7.2. Evaluating the reasonability of these interpolated values constitutes an essential step in validating subsequent power production estimations.

Interpolated wind speeds demonstrate a mean of 2.82 m/s relative to 3.04 m/s for raw measurements at station locations, constituting a 7.2% reduction in mean values. This systematic decrease is physically defensible and stems from the spatial distribution of turbines relative to measurement stations. The interpolation procedure effectively aggregates wind conditions across these spatially heterogeneous locations, inherently yielding lower mean values than the deliberately selected, high-exposure measurement sites.

Standard deviation diminished from 1.72 m/s for raw measurements to 1.56 m/s for interpolated values, signifying reduced variability. This *smoothing phenomenon* represents an intrinsic characteristic of spatial interpolation methodologies. Although this process entails some loss of spatial granularity and attenuates local extremes, it remains aligned with the intended objective of estimating regional wind patterns rather than resolving highly localized phenomena. The interpolated range spanning 0.00 to 15.36 m/s resides comfortably within the absolute valid range of 0.0 to 30.0 m/s, with no physically implausible values identified.

Nevertheless, 34.44% of interpolated values fell outside the typical operational range of 2.0 to 15.0 m/s, predominantly comprising low wind speed values below 2.0 m/s (44,153,879 values with a mean of 1.26 m/s). This substantial proportion of low-wind conditions reflects the wind regime characteristics of Brandenburg, particularly during summer months and nighttime periods characterized by weak synoptic forcing. While regional climate data for Brandenburg indicates average hourly wind speeds around 4.0 m/s during summer, with the calmest day exhibiting approximately 3.9 m/s and the windiest day averaging approximately 5.4 m/s [4], both the interpolated values (mean 2.82 m/s) and the underlying raw measurement data (mean 3.04 m/s) exhibit somewhat lower values. This discrepancy likely stems from differences in measurement locations, temporal coverage, and the specific meteorological conditions during the analysis period.

Wind direction statistics reveal minimal divergence between interpolated and raw measurements, with mean values of 208.34° and 209.91° respectively. This close correspondence indicates that the interpolation methodology preserves the predominant wind direction patterns, which proves critical for turbine yaw positioning and wake effect modeling in subsequent production calculations. The absence of any NaN or infinite values substantiates the numerical stability and robustness of the interpolation procedure.

9.1.7 Vertical Extrapolation Reasonability

Applying the power law method to extrapolate wind speeds from measurement height (10m) to turbine hub height transformed 128,216,264 values over the analysis period, as presented in Section 7.3. Assessing the physical reasonability of this extrapolation proves essential, given that hub height wind speed constitutes the primary determinant of power production estimates.

Extrapolated hub height wind speeds demonstrate a mean of 3.94 m/s relative to 2.82 m/s at measurement height, constituting a mean increase ratio of 1.3966 or an absolute increment of 1.12 m/s. This elevation aligns with the theoretical expectation that wind speed intensifies with altitude above ground owing to diminished surface friction effects, as characterized by the *atmospheric boundary layer* wind profile.

The extrapolated wind speed range spanning 0.00 to 22.34 m/s represents a maximum value 45.4% higher than the interpolated maximum of 15.36 m/s. This amplification follows naturally from the multiplicative nature of the power law, wherein higher base wind speeds receive proportionally larger absolute increments. The maximum value of 22.34 m/s remains substantially below the absolute validity threshold of 45.0 m/s and represents realistic extreme wind conditions at hub height. Nevertheless, 26.90% of values fell outside the typical operational range of 2.4 to 22.5 m/s, predominantly at the lower end, reflecting the prevalence of weak wind conditions that persist across all vertical levels.

9.2 Limitations

Multiple constraints limit the scope, validation capacity, and accuracy of the methodology developed in this research. These constraints originate from geographical restrictions, data accessibility challenges, methodological simplifications, and the unavailability of ground-truth production data for validation purposes.

9.2.1 Geographical and Validation Constraints

The methodology has been developed and evaluated exclusively within Brandenburg, thereby restricting the generalizability of findings to other regions characterized by divergent geographical, meteorological, or infrastructural attributes. Validation across diverse settings would prove necessary to evaluate transferability and identify region-specific modifications requisite for broader applicability.

A fundamental constraint stems from the inability to rigorously validate the spatial extrapolation and power production estimation components against actual turbine-level measurements. As discussed in Chapter 1, turbine production data remains

inaccessible to the public, thereby precluding quantitative assessment of extrapolation accuracy and production estimate reliability. Consequently, the evaluation of these components depends on *reasonability assessments* rather than empirical validation metrics. This absence of ground-truth turbine data represents the most significant limitation of this research, as it prevents confirmation of how accurately the production estimates reflect actual turbine output. Without empirical validation against real-world generation data, the practical utility of the methodology for operational applications remains uncertain, despite the theoretical soundness of the underlying approach.

9.2.2 Methodological Simplifications

The spatial interpolation methodology employs IDW, a relatively straightforward approach necessitating several simplifying assumptions. The method presumes uniform ground elevation for all stations and turbines, disregards terrain roughness variations, and exploits Brandenburg's flat topography to rationalize the omission of orographic effects. These simplifications constitute appropriate starting points for proof-of-concept development but may introduce systematic errors in wind speed estimation at individual turbine locations. More sophisticated interpolation techniques incorporating terrain characteristics, local land cover, and atmospheric stability could potentially enhance accuracy, particularly in regions exhibiting complex topography.

Analogously, vertical extrapolation to hub height depends exclusively on the power law method with a fixed shear exponent of 0.14, which furnishes a straightforward empirical relationship between altitude and wind speed. This approach assumes flat terrain and uniform atmospheric conditions, which may not accurately represent local conditions at all turbine sites. In reality, the wind shear exponent varies with surface roughness, atmospheric stability, and time of day, meaning that a single fixed value cannot optimally characterize all locations and meteorological situations. Alternative approaches accounting for atmospheric boundary layer dynamics and site-specific terrain characteristics could yield improved accuracy but at substantially elevated computational cost and data requirements.

The spatial coverage of meteorological measurements is constrained by the availability of merely 32 DWD stations satisfying the requisite criteria (10-minute resolution, active status, wind measurements). Augmenting station density would enhance spatial interpolation accuracy by diminishing the distance between measurement points and turbine locations, thereby minimizing extrapolation uncertainty. A particularly severe constraint manifests along Brandenburg's eastern border, where the absence of measurement stations in neighboring Poland generates a data gap. Turbines situated in eastern Brandenburg depend on more distant stations for interpolation, likely

compromising estimate accuracy in this region. Incorporating Polish meteorological stations, contingent upon establishing data sharing agreements, would substantially enhance coverage.

9.2.3 Power Curve Data Limitations

Not all turbine manufacturer-model combinations registered in the Marktstammdaten-register possess corresponding entries in The Wind Power database. For turbines lacking specific power curves, a generic curve scaled by rated capacity is applied. This introduces approximation error, given that turbine designs exhibit variability in their aerodynamic efficiency, cut-in/cut-out speeds, and performance characteristics. Additionally, the matching process between MaStR turbine records and power curve database entries depends on fuzzy string matching of manufacturer and model names, which lacks complete reliability and may yield occasional mismatches or failed matches.

9.2.4 Idealized Operational Assumptions

The power production estimation methodology incorporates multiple idealized assumptions regarding turbine operation that inadequately reflect real-world conditions. The model presumes that turbines continuously maintain optimal yaw alignment with the wind direction, positioning the rotor perpendicular to the incoming flow to maximize energy capture. Furthermore, the model assumes that each turbine operates in unobstructed conditions, disregarding wake effects from upstream turbines within wind farms. In reality, yaw control systems respond to changing wind directions with finite adjustment velocities, introducing periods of suboptimal alignment that diminish actual production below the power curve predictions. Additionally, yaw error can accumulate owing to sensor inaccuracies or control system limitations, particularly during rapidly shifting wind conditions.

The assumption of steady, uniform wind flow across the rotor disk similarly constitutes an idealization. Real turbines encounter turbulent flow, wind shear variations across the vertical extent of the rotor, and short-term gusts that induce instantaneous fluctuations in production. The 10-minute averaging interval of meteorological measurements attenuates these fluctuations but cannot capture transient dynamics that influence cumulative energy generation.

Furthermore, the model disregards operational constraints and availability factors. Turbines undergo scheduled maintenance, experience unplanned downtime attributable to component failures, and may be curtailed for grid balancing or noise restrictions. The methodology implicitly presumes 100% availability, potentially

overestimating actual production when turbines are offline or operating in reduced-output modes. Incorporating historical availability statistics or real-time operational status data, if accessible, would enhance estimate realism.

9.3 Practical Implications

Despite the limitations outlined above, this research demonstrates several practical contributions relevant to energy transparency and future research directions.

The developed methodology and prototype system establish a foundation for public visualization of wind power production in Brandenburg at turbine-level granularity. Emulating the model pioneered by initiatives such as the E.ON Energy Monitor, the system could amplify public awareness and comprehension of renewable energy generation patterns throughout the region. By empowering citizens to observe both aggregated regional production and individual turbine contributions, the platform could cultivate greater engagement with the energy transition and furnish tangible visibility into the magnitude and variability of wind energy generation.

From a research methodology standpoint, this work constitutes a proof of concept demonstrating that turbine-level production approximation and forecasting can be accomplished utilizing exclusively publicly accessible data sources. The dependence on open DWD meteorological measurements, the public MaStR turbine registry, and commercially available (but non-proprietary) power curve data obviates dependencies on confidential production data or private agreements with turbine operators. This approach diminishes barriers to entry for subsequent research, enabling other investigators to replicate, extend, or adapt the methodology without necessitating access to restricted datasets.

The modular architecture of the methodology, featuring clearly delineated components for data acquisition, forecasting, spatial extrapolation, and power estimation, facilitates incremental enhancements. Individual components can be refined or replaced with more sophisticated alternatives as methods advance or additional data sources emerge, without necessitating comprehensive system redesign. For instance, more advanced spatial interpolation techniques, improved forecasting models, or enhanced power curve databases could be integrated into the existing framework to progressively augment accuracy.

9.4 Future Work

Multiple research trajectories emerge from the constraints and findings of this work, presenting opportunities for validation, methodological refinement, and extension to broader contexts.

9.4.1 Empirical Validation with Ground-Truth Data

The most critical subsequent step involves empirical validation of the spatial extrapolation and power production estimation components against actual turbine measurements. This would necessitate deployment of measurement instrumentation on selected wind turbines or, more feasibly, establishing data-sharing partnerships with turbine operators willing to furnish production data for research purposes. Even a constrained validation dataset encompassing a subset of turbines would facilitate quantitative assessment of extrapolation errors, power curve approximation accuracy, and the ramifications of simplifying assumptions. Such validation would furnish empirical grounding for the current reasonability-based assessment and identify specific error sources amenable to targeted refinement.

9.4.2 Methodological Enhancements

Comparative evaluation of alternative methods for spatial interpolation and weather forecasting could reveal performance improvements beyond the current baseline approaches. For spatial interpolation, terrain-aware models incorporating digital elevation data, surface roughness maps, and land cover characteristics would provide more realistic wind speed estimates at turbine locations. Such approaches could account for topographical channeling effects, sheltering by obstacles, and local acceleration zones that the current IDW method cannot capture. Similarly, vertical extrapolation could be refined by employing site-specific wind shear exponents derived from atmospheric stability indicators or by implementing more sophisticated boundary layer models that account for diurnal variations and terrain-induced flow modifications. Systematic benchmarking against the current IDW, power law method, and BiLSTM/PatchTST baselines would quantify potential gains from these enhancements.

For the forecasting component, the persistence baseline could be enhanced to incorporate spatial context by utilizing wind measurements from neighboring stations as additional input features, rather than relying solely on the temporal history of a single station. This multi-station persistence approach would better capture regional wind patterns and improve baseline forecast accuracy, thereby establishing a more challenging benchmark for evaluating machine learning models. Additionally,

systematic hyperparameter optimization for the deep learning forecasting models, currently tuned through manual experimentation, could improve performance. Techniques such as Bayesian optimization or automated neural architecture search could identify superior configurations, while ensemble methods combining predictions from multiple model variants could enhance robustness.

9.4.3 Geographical Transferability

Evaluating the transferability of the methodology to regions beyond Brandenburg would test its generalizability and identify requisite adaptations. Regions exhibiting varying topographical complexity, climatic regimes, or meteorological station densities would furnish diverse test cases. Establishing quantitative evaluation metrics for cross-regional comparison, potentially grounded in production estimate consistency or physical plausibility checks, would enable systematic assessment. Region-specific data source availability and quality would likewise necessitate evaluation, given that meteorological infrastructure and turbine registries vary internationally.

9.4.4 Extension to Solar Power

The general framework developed for wind power — integrating meteorological forecasting, spatial interpolation, and technology-specific power models — could potentially be adapted for solar photovoltaic generation. Solar irradiance forecasting, spatial interpolation of solar radiation measurements, and application of photovoltaic performance models present analogous challenges. Nevertheless, solar-specific considerations such as panel orientation, shading, temperature effects, and cloud cover dynamics would necessitate substantial methodological modifications. A parallel solar implementation could facilitate integrated renewable energy visualization for regions possessing both wind and solar capacity.

9.4.5 Integration with Grid and Market Data

Future extensions could integrate the turbine-level production estimates with electrical grid data, such as transmission constraints or regional demand patterns, to evaluate renewable energy integration challenges. Linking production forecasts to electricity market price data could facilitate economic analysis of wind generation value. Such integrations would broaden the utility of the system beyond visualization toward decision support for grid operators, energy traders, or policy analysts.

Chapter 10

Conclusion

This thesis tackled the challenge of approximating, visualizing, and forecasting wind power production at the individual turbine level using exclusively publicly accessible data sources. Driven by the imperative for enhanced transparency in renewable energy generation and recognizing the constraints of existing aggregated visualization platforms, this research developed and implemented a proof-of-concept methodology for Brandenburg. The work demonstrates that individual turbine production can be estimated without relying on proprietary production data or establishing private agreements with turbine operators.

The methodological framework synthesizes three publicly available data sources: meteorological measurements from 32 DWD weather stations, turbine metadata from 4,796 active wind installations registered in the MaStR, and manufacturer-specific power curves from The Wind Power database. This approach implements a *multi-stage computational pipeline* that begins with spatial interpolation of station measurements to turbine locations using IDW, proceeds with vertical extrapolation to hub height via the power law method, and concludes with power production estimation through application of turbine-specific power curves. For short-term forecasting capabilities, two deep learning architectures — Bidirectional LSTM and PatchTST — were trained to predict wind speed and direction at measurement stations over a 12-hour horizon, with these forecasts subsequently propagated through the identical spatial interpolation and power production estimation pipeline.

The forecasting component validated the effectiveness of deep learning approaches for short-term wind prediction. Both architectures achieved substantial performance improvements over the persistence baseline, with evaluation on held-out test data providing quantitative validation through MAE and RMSE metrics for wind speed, alongside circular angular error for wind direction. Although further hyperparameter optimization and architectural refinements could yield additional performance gains,

the results confirm that both models successfully capture meaningful temporal patterns in meteorological conditions suitable for downstream production forecasting.

A fundamental limitation acknowledged throughout this research concerns the inability to validate spatial interpolation and power production estimates against actual turbine measurements, as such data remains inaccessible from private operators. Consequently, validation of these components relied on *reasonability assessments* — examining whether estimates respect physical constraints such as rated capacity limits and exhibit plausible spatiotemporal patterns — rather than empirical error quantification. This absence of ground-truth validation represents the primary weakness of the methodology, preventing confirmation of how accurately the production estimates reflect real-world turbine output. Furthermore, the interpolation and vertical extrapolation methods, while constituting appropriate starting points for proof-of-concept development, employ simplified assumptions regarding terrain uniformity and atmospheric conditions that may not hold universally across all turbine locations. These constraints, though inherent given current data availability, emphasize the necessity of future validation efforts involving partnerships with turbine operators or sensor deployment on selected installations.

Notwithstanding methodological simplifications and validation constraints, the research successfully delivered a functional web-based prototype system that visualizes both historical production approximations and 12-hour forecasts for all wind turbines across Brandenburg. The dashboard offers both regional aggregated views and individual turbine detail pages, providing stakeholders and citizens with an accessible interface to explore wind power generation patterns. This prototype establishes that turbine-level transparency is achievable within the constraints of public data, thereby creating a foundation for enhanced public engagement with renewable energy systems.

The *modular architecture* of the methodology enables incremental improvement. Individual components — spatial interpolation, vertical extrapolation, forecasting models, or power curve databases — can be refined or substituted as more sophisticated methods or enhanced data sources become available, without necessitating complete system redesign. As elaborated in Chapter 9, numerous opportunities exist for methodological enhancement, including adoption of terrain-aware interpolation models that incorporate topographical and land cover data, refinement of vertical extrapolation through site-specific wind shear parameters, enhancement of the persistence baseline through multi-station spatial features, ensemble forecasting approaches, systematic hyperparameter optimization, and empirical validation against ground-truth measurements. Extension of the framework to additional geo-

graphical regions or adaptation for solar photovoltaic generation represent further promising directions for future research.

In conclusion, this thesis contributes a demonstrated *proof of concept* for public-data-driven wind power transparency at turbine-level resolution, encompassing both current production approximation and short-term forecasting. While acknowledging the limitations inherent in relying on spatial interpolation and idealized operational assumptions, the work establishes that meaningful insight into regional wind generation can be achieved without access to proprietary production data. The methodology's emphasis on simplicity, modularity, and reproducibility positions it as an accessible foundation for subsequent research, policy applications, or public engagement initiatives seeking to advance transparency in the renewable energy transition.

References

[1] Décio Alves, Fábio Mendonça, Sheikh Shanawaz Mostafa, and Fernando Morgado-Dias. The Potential of Machine Learning for Wind Speed and Direction Short-Term Forecasting: A Systematic Review. *Computers*, 12(10):206, October 2023. ISSN 2073-431X. doi: 10.3390/computers12100206.

[2] Cambridge University Press and Assessment. Extrapolation Definition. https://dictionary.cambridge.org/dictionary/english/extrapolation, October 2025.

[3] Cambridge University Press and Assessment. Interpolation Definition. https://dictionary.cambridge.org/dictionary/english/interpolation, October 2025.

[4] Cedar Lake Ventures Inc. Brandenburg an der Havel Summer Weather, Average Temperature (Brandenburg, Germany) - Weather Spark. https://weatherspark.com/s/73885/1/Average-Summer-Weather-in-Brandenburg-an-der-Havel-Brandenburg-Germany.

[5] Stephanie Cole. Wind Turbine Power Curve - TheRoundup, August 2021.

[6] DataCamp. Mean Absolute Error Explained: Measuring Model Accuracy. https://www.datacamp.com/tutorial/mean-absolute-error, .

[7] DataCamp. RMSE Explained: A Guide to Regression Prediction Accuracy. https://www.datacamp.com/tutorial/rmse, .

[8] Deutscher Wetterdienst. Deutscher Wetterdienst - Der DWD. https://www.dwd.de/DE/derdwd/derdwd_node.html, .

[9] Deutscher Wetterdienst. Deutscher Wetterdienst - Leistungen - Open Data. https://www.dwd.de/DE/leistungen/opendata/opendata.html, .

[10] Deutscher Wetterdienst. Deutscher Wetterdienst - Leistungen - Berechnung des Rasters der Windgeschwindigkeit. https://www.dwd.de/DE/leistungen/uhi/info_methodic/05_wind.html, .

[11] Deutscher Wetterdienst. Deutscher Wetterdienst - Rechtliche Hinweise. https://www.dwd.de/DE/service/rechtliche_hinweise/rechtliche_hinweise_node.html, .

[12] Deutscher Wetterdienst. CDC-OpenData area, August 2020.

[13] Deutscher Wetterdienst. Datensatzbeschreibung - 10-minütige Stationsmessungen des Windes für Deutschland, March 2024.

[14] Dictionary.com LLC. Approximation. https://www.dictionary.com/browse/approximation, February 2025.

[15] Die Gas- und Wasserstoffwirtschaft e.V. Klimaneutrales Stromsystem: Herausforderungen und Lösungen. https://gas-h2.de/transformation-energiesystem/strom-aus-gas/herausforderung-stromsystem/.

[16] Con Doolan. Taller, faster, better, stronger. Wind towers are only getting bigger. https://www.unsw.edu.au/newsroom/news/2019/07/taller–faster–better–stronger–wind-towers-are-only-getting-b, July 2019.

[17] Dmitry Duplyakin, Sagi Zisman, Caleb Phillips, and Heidi Tinnesand. Bias Characterization, Vertical Interpolation, and Horizontal Interpolation for Distributed Wind Siting Using Mesoscale Wind Resource Estimates. Technical Report NREL/TP–2C00-78412, 1760659, MainId:32329, January 2021.

[18] ECMWF. ERA5: How to calculate wind speed and wind direction from u and v components of the wind? - Copernicus Knowledge Base - ECMWF Confluence Wiki. https://confluence.ecmwf.int/pages/viewpage.action?pageId=133262398, March 2024.

[19] E.ON SE. EnergieMonitor: Deutsche Technologie als Transparenztreiber. https://www.eon.com/de/innovation/zukunft-der-energie/intelligente-netze/energiemonitor-deutsche-technologie-als-treiber-fuer-mehr-energietransparenz.html, January 2021.

[20] Katja Evers. Wenn sich Windräder den Wind klauen: So viel Einfluss hat der Wake-Effekt | MDR.DE. https://www.mdr.de/wissen/umwelt-klima/wenn-sich-windraeder-den-wind-klauen-einfluss-wake-effekt100.html, September 2024.

[21] GeeksforGeeks. Bidirectional LSTM in NLP. https://www.geeksforgeeks.org/nlp/bidirectional-lstm-in-nlp/, October 2025.

[22] Stuart Grange. *Technical Note: Averaging Wind Speeds and Directions*. June 2014. doi: 10.13140/RG.2.1.3349.2006.

[23] K Hartmann, J Krois, and A. Rudolph. Statistics and Geodata Analysis using R (SOGA-R). https://www.geo.fu-berlin.de/en/v/soga-r/Advances-statistics/Geostatistics/Inverse-Distance-Weighting-IDW/index.html, 2023.

[24] IBM. Was ist Machine Learning (ML)? | IBM. https://www.ibm.com/de-de/think/topics/machine-learning, September 2021.

[25] Iguazio. What is Baseline Models. https://www.iguazio.com/glossary/baseline-models/.

[26] Christopher Jung and Dirk Schindler. The role of the power law exponent in wind energy assessment: A global analysis. *International Journal of Energy Research*, 45(6):8484–8496, May 2021. ISSN 0363-907X, 1099-114X. doi: 10.1002/er.6382.

[27] Mathis Arne Kopp. Interview regarding E.ON Monitor, March 2025.

[28] Vijay Kotu and Bala Deshpande. Chapter 12 - Time Series Forecasting. In Vijay Kotu and Bala Deshpande, editors, *Data Science (Second Edition)*, pages 395–445. Morgan Kaufmann, January 2019. ISBN 978-0-12-814761-0. doi: 10.1016/B978-0-12-814761-0.00012-5.

[29] Marktstammdatenregister. Marktstammdatenregister - Datendownload. https://www.marktstammdatenregister.de/MaStR/Datendownload, .

[30] Marktstammdatenregister. Marktstammdatenregister - Zielsetzung des Marktstammdatenregister. https://www.marktstammdatenregister.de/MaStRHilfe/subpages/GrundlagenZielsetzung.html, .

[31] Christian Matthaei. Wind Turbines and Measurement Stations Map.

[32] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers, March 2023.

[33] Christopher Olah. Understanding LSTM Networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/, August 2015.

[34] OpenJS Foundation. React Google Maps. https://visgl.github.io/react-google-maps/.

[35] Mike Optis, Nicola Bodini, Mithu Debnath, and Paula Doubrawa. New methods to improve the vertical extrapolation of near-surface offshore wind speeds. *Wind Energy Science*, 6(3):935–948, June 2021. ISSN 2366-7443. doi: 10.5194/wes-6-935-2021.

[36] A. Pintor, C. Pinto, J. Mendonça, Rosa Maria Pilão, and P. Pinto. Insights on the use of wind speed vertical extrapolation methods. 2022.

[37] Daniel Vázquez Pombo, Peder Bacher, Charalampos Ziras, Henrik W. Bindner, Sergiu V. Spataru, and Poul E. Sørensen. Benchmarking physics-informed machine learning-based short term PV-power forecasting tools. *Energy Reports*, 8:6512–6520, November 2022. ISSN 2352-4847. doi: 10.1016/j.egyr.2022.05.006.

[38] QGIS. 11. Spatial Analysis (Interpolation) — QGIS Documentation documentation. https://docs.qgis.org/3.40/en/docs/gentle_gis_introduction/spatial_analysis_interpolation.html, October 2025.

[39] Sebastián Ramírez. FastAPI. https://fastapi.tiangolo.com/.

[40] Siva Sankari Subbiah, Senthil Kumar Paramasivan, Karmel Arockiasamy, Saminathan Senthivel, and Muthamilselvan Thangavel. Deep Learning for Wind Speed Forecasting Using Bi-LSTM with Selected Features. *Intelligent Automation & Soft Computing*, 35(3):3829–3844, 2023. ISSN 1079-8587. doi: 10.32604/iasc.2023.030480.

[41] Suisse Eole. Weibull Calculator. https://wind-data.ch/tools/weibull.php?lng=en.

[42] The Wind Power. The Wind Power - Wind energy Market Intelligence - About. https://www.thewindpower.net/about_en.php.

[43] topographic-map.com. Brandenburg topographic map, elevation, terrain.

[44] Umweltbundesamt. Umweltbundesamt - Erneuerbare Energien in Zahlen. https://www.umweltbundesamt.de/themen/klima-energie/erneuerbare-energien/erneuerbare-energien-in-zahlen#uberblick, June 2025.

[45] U.S. Department of Energy. How Do Wind Turbines Work? https://www.energy.gov/eere/wind/how-do-wind-turbines-work.

[46] U.S. Department of Energy. What Is Wind Power? https://windexchange.energy.gov/what-is-wind, 2025.

[47] Vercel Inc. Next.js by Vercel - The React Framework. https://nextjs.org/, .

[48] Vercel Inc. Shadcn/ui. https://ui.shadcn.com/, .

[49] wind-turbine.com GmbH. Who owns wind turbines or wind farms in Germany? https://en.wind-turbine.com/magazine/news-from-the-market/231223/who-owns-wind-turbines-or-wind-farms-in-germany.html, December 2023.

[50] Zhou Wu, Gan Luo, Zhile Yang, Yuanjun Guo, Kang Li, and Yusheng Xue. A comprehensive review on deep learning approaches in wind forecasting applications. *CAAI Transactions on Intelligence Technology*, 7(2):129–143, 2022. ISSN 2468-2322. doi: 10.1049/cit2.12076.

[51] Chang Xu, Chenyan Hao, Linmin Li, Xingxing Han, Feifei Xue, Mingwei Sun, and Wenzhong Shen. Evaluation of the Power-Law Wind-Speed Extrapolation Method with Atmospheric Stability Classification Methods for Flows over Different Terrain Types. *Applied Sciences*, 8(9):1429, September 2018. ISSN 2076-3417. doi: 10.3390/app8091429.

[52] Shiwei Xu, Yongjun Wang, Xinglei Xu, Guang Shi, He Huang, and Yingya Zheng. A PatchTST-GRU based heterogeneous seq2seq model with numerical weather prediction refinement for multi-step wind power forecasting. *Scientific Reports*, 15(1):16547, June 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-00741-9.

[53] XU Exponential University of Applied Sciences GmbH. RITS Homepage. https://resiliente-infrastruktur.de/en/, 2025.

Appendix A

Code

The complete source code is publicly available on GitHub at https://github.com/Chrissydrx/bachelor-project.git.

The repository encompasses the web-based prototype described in Chapter 8, as well as Jupyter notebooks that facilitate interaction with the system's services and enable comprehensive evaluation of the forecasting results. While the project does not provide exhaustive documentation, the fundamental concepts and implementation details are explained in the README.md file and within the notebooks themselves.

Appendix B

AI-Assisted Writing Tool

Cursor (https://cursor.com/), an AI-assisted code and text editor, was employed to enhance the textual quality of this thesis. The tool was utilized in a two-stage refinement process. In the first stage, preliminary bullet-point outlines were transformed into coherent prose. In the second stage, the generated prose was further refined to improve clarity, conciseness, and academic tone. Following each stage, the AI-generated text was reviewed and adapted by the author as necessary to ensure accuracy and appropriateness. It is important to note that the tool was not employed for conceptual development or ideation; rather, it served exclusively to enhance the linguistic presentation of author-generated content. The AI assistant operated according to the following predefined rules:

Rules

No Code Agent

Role

You are a scientific writing assistant, and your goal is to help me draft a scientific paper.

Instructions

Code Agent

Don't behave like a code agent. We don't need to write any code. We only want to write a scientific paper in LaTeX.

Latex Instructions

Role

You are a scientific writing assistant, and your goal is to help me draft a scientific paper.

Instructions

Writing styles

When editing, please focus on improving clarity, conciseness, and coherence while maintaining the technical accuracy and integrity of the content. Ensure that the tone remains formal but accessible for the general public. Please diversify the use of words and avoid overusing certain words or phrases, especially in the same sentence. Write in full sentences and avoid using bullet points.

Format

The document is written in LaTeX, so when you make changes, follow the LaTeX syntax.

You should start a new line for each sentence.

Latex Prose

Role

You are a scientific writing assistant, and your goal is to help me draft a scientific paper.

Instructions

Prosa

I will provide you with bullet points that I want to convert to prose. It is really important that you keep the same meaning and facts from the bullet points. The bullet points may include literature references. When converting to prose, please preserve these LaTeX citations. Note that you can consolidate multiple citations into a single reference if they point to the same source. The references are provided with each bullet point to indicate the origin of the information and to ensure you maintain this attribution when transforming the text into prose. Keep the correct citation in mind. To improve the text quality, please also have a look at the other chaperts before writing the prose. All chapters are in folders that start with a number (1 - 10). Use a tone that is suitable for a scientific paper.

Example Bullet Points to Prose

Input:

- The expansion of renewable energy sources is accelerating, making them an increasingly significant component of the energy mix. In Germany, the share of renewable energies in the electricity sector increased to 54.4% in 2024. This is an increase of 3% compared to 2023. [44]

- In 2024 renewable energy sources produced 284 Billion kWh [44]

- Winde energy production covers 138.9 Billion kWh (48.9%) of the renewable energy production and is therefore the largest renewable energy. [44]

- In 2024 new wind turbines with a total capacity of 3,337 MW were installed. [44]

- In 2024, new wind turbines with a total capacity of 15,000 MW were approved. So, the expansion will increase in the future drastically. [44]

Prompt:

"Please rephrase the following bullet points to prose. Keep the same meaning and facts from the bullet points. Don't add any new information."

Output:

The expansion of renewable energy sources is accelerating, making them an increasingly significant component of the energy mix. In Germany, the share of renewable energies in the electricity sector increased to 54.4% in 2024, representing a 3% increase compared to 2023. In 2024, renewable energy sources produced 284 billion kWh. Wind energy production covered 138.9 billion kWh (48.9%) of the renewable energy production, making it the largest renewable energy source. In 2024, new wind turbines with a total capacity of 3,337 MW were installed. Furthermore, new wind turbines with a total capacity of 15,000 MW were approved in 2024, indicating that the expansion will increase drastically in the future [44].

Example Prose to Latex

Input:

Taken from the previous example.

Prompt:

"Rephrase this, keep the citations in mind"

Output:

Renewable energy sources are experiencing accelerated expansion and have become an increasingly important element of the energy mix. In Germany, renewable energies accounted for 54.4% of electricity generation in 2024, marking a 3% rise

from the previous year. Total renewable energy production reached 284 billion kWh in 2024. Among renewable sources, wind energy dominated with 138.9 billion kWh, constituting 48.9% of total renewable energy production. During 2024, wind turbines with a combined capacity of 3,337 MW were newly installed. Additionally, approvals were granted for wind turbines totaling 15,000 MW in capacity, suggesting a substantial acceleration of expansion in the coming years [44].