

Weather Stability and Renewable Energy Forecasting Model Performance

A Comparative Robustness Analysis of Statistical and Deep Learning
Models
across Stable and Unstable Weather Regimes

Bachelor Thesis

Author:

Abdelrahman Elnagar

Supervisors:

Prof. Dr. Stephan Schlüter
Abhinav Das, M.Sc.

Date:

November 6, 2025

Abstract

Contents

1	Introduction	4
1.1	Motivation	4
1.2	Research Questions	5
1.3	Objectives	5
1.4	Thesis Structure	6
2	Literature Review	8
2.1	Introduction	8
2.2	Background Theory	8
2.2.1	Weather Regime Classification Fundamentals	8
2.2.2	Statistical Methods for Weather Classification	9
2.3	Renewable Energy Forecasting Models	9
2.3.1	Statistical Models	9
2.3.2	Machine Learning and Deep Learning Models	10
2.3.3	Comparative Performance Analysis	12
2.4	Weather Classification in Renewable Energy Forecasting	12
2.4.1	Incorporating Weather Information	12
2.4.2	Weather-Driven Model Selection	13
2.4.3	Performance Across Weather Regimes	13
2.5	Model Robustness and Evaluation	13
2.5.1	Statistical Validation Methods	13
2.5.2	Robustness Metrics	14
2.5.3	Experimental Design for Comparative Analysis	14
2.6	Gap Analysis	14
2.6.1	Identified Research Gaps	14
2.6.2	Research Questions Alignment	15
2.7	Summary	15
3	Methodology	17
3.1	Research Design	17
3.1.1	Pipeline Architecture	17
3.1.2	Design Rationale	19
3.2	Data Collection	19

3.2.1	Weather Data	20
3.2.2	Renewable Energy Data	21
3.3	Data Preprocessing	22
3.3.1	Completed Weather Data Preprocessing	22
3.3.2	Future Preprocessing Steps for Integration	24
3.4	Weather Stability Index Development	26
3.4.1	Feature Engineering	26
3.4.2	WSI Computation	29
3.4.3	Classification Methods	29
3.4.4	Validation Criteria	30
3.5	Renewable Energy Prediction Models	31
3.6	Performance Evaluation	31
3.7	Robustness Analysis	32
3.8	Validation Approach	32
3.9	Ethical Considerations	32
4	Data and Experimental Setup	33
4.1	Data Description	33
4.2	Data Sources	33
4.3	Experimental Setup	33
4.4	Software and Tools	33
5	Results	34
5.1	Introduction	34
5.2	Weather Stability Results (2024)	34
5.2.1	WSI Characteristics and Regimes	34
5.2.2	Monthly Stability Patterns	35
5.3	Main Findings	35
5.4	Analysis	36
5.5	Summary	36
6	Discussion	37
6.1	Interpretation of Results	37
6.2	Comparison with Literature	37
6.3	Limitations	37
6.4	Implications	37
7	Conclusion	38
7.1	Summary of Contributions	38
7.2	Research Questions Revisited	38
7.3	Future Work	38
7.4	Final Remarks	38
A	Appendix A	39

B Appendix B **40**

References **45**

1 Introduction

1.1 Motivation

Renewable energy sources, particularly solar photovoltaics (PV) and wind power, have become integral components of Germany’s energy transition towards a carbon-neutral future. As the share of renewable energy in the electricity grid continues to grow, accurate forecasting of renewable energy production has become critical for grid stability, energy market operations, and optimal resource allocation (1; 2). However, the inherent variability and uncertainty associated with weather-dependent renewable energy generation pose significant challenges for prediction models.

Weather conditions exhibit substantial temporal variability, transitioning between relatively stable periods characterized by consistent meteorological patterns and unstable periods marked by rapid changes in atmospheric conditions. These weather regimes directly influence renewable energy production: stable weather typically results in more predictable generation patterns, while unstable weather introduces greater uncertainty and variability. Yet, the extent to which weather stability affects the performance of forecasting models (both statistical and deep learning) remains insufficiently explored in the existing literature.

Grid operators and energy market participants require reliable forecasts under all weather conditions to ensure system stability, optimize dispatch decisions, and manage grid congestion effectively (2; 3). Current operational forecasting systems often employ a single model or ensemble approach without considering the prevailing weather regime (28). This one-size-fits-all strategy may be suboptimal, as different prediction models may exhibit varying levels of robustness to weather instability. A model that performs excellently under stable conditions may degrade significantly during unstable periods, potentially compromising operational reliability.

Despite the recognized importance of weather variability in renewable energy forecasting, there is a notable gap in the literature regarding systematic evaluation of model robustness across different weather stability regimes. Most studies evaluate model performance on aggregate datasets without explicitly accounting for weather regime characteristics (37; 43). This research addresses this gap by developing a comprehensive framework to quantify weather stability and assess how both statistical and deep learning prediction models respond to stable versus unstable weather conditions. The framework builds upon established weather regime classification methods (5; 6; 7) and applies them to renewable energy forecasting contexts.

The significance of this research extends beyond academic interest. By identifying which models

are most robust to weather instability and providing operational guidance on model selection based on weather conditions, this work contributes to improving the reliability and efficiency of renewable energy integration. The findings can inform grid operators, energy traders, and forecasting system designers about optimal model selection strategies that adapt to changing weather conditions, ultimately supporting more reliable and cost-effective renewable energy management.

1.2 Research Questions

To address the research gap identified in Section 1.1, this thesis investigates the following research questions:

1. **RQ1:** Does renewable energy prediction accuracy differ significantly between weather-stable and weather-unstable periods?

This question seeks to establish whether weather stability is a meaningful factor influencing prediction model performance. A positive answer would confirm that weather regime characteristics affect forecasting accuracy, justifying the development of weather-adaptive model selection strategies.

2. **RQ2:** Which forecasting models (statistical and deep learning) are most robust to weather instability (i.e., show smallest performance degradation during unstable periods)?

This question aims to identify which models maintain their predictive performance when weather conditions become unstable. Understanding model-specific robustness characteristics enables informed selection of appropriate forecasting models based on expected weather conditions.

3. **RQ3:** Can weather stability information improve model selection strategies for operational renewable energy forecasting?

This question evaluates the practical utility of incorporating weather stability information into operational forecasting systems. It assesses whether knowledge of weather regime characteristics can be leveraged to select optimal models dynamically, potentially improving overall forecasting accuracy and reliability.

These research questions guide the development of the dual-pipeline methodology described in Chapter 3, where weather stability classification and energy prediction models are evaluated independently before being combined for comparative analysis.

1.3 Objectives

To address the research questions outlined in Section 1.2, this research aims to achieve the following objectives:

1. **Develop a Weather Stability Index (WSI)** using 11 weather attributes to classify periods as stable or unstable. This index will be computed from meteorological data

including temperature, cloudiness, wind, precipitation, pressure, and other relevant atmospheric variables. The WSI will provide a quantitative measure of weather variability, enabling systematic classification of weather regimes throughout the analysis period.

2. **Implement and evaluate statistical and deep learning models** to predict renewable energy production for 2024. Multiple established forecasting models will be implemented, including statistical models (persistence models, ARIMA/SARIMA, Prophet, exponential smoothing) and deep learning models (LSTM, CNN-LSTM, TCN, and ensemble approaches). All models will be trained and evaluated using historical renewable energy production data for Germany.
3. **Compare model performance under stable vs unstable weather conditions** by stratifying prediction errors according to the weather stability classification. This comparison will quantify performance differences and identify which models are most affected by weather variability.
4. **Identify which models are most robust to weather instability** by analyzing performance degradation patterns. Models that maintain relatively stable accuracy across different weather regimes will be considered more robust and potentially more suitable for operational deployment.
5. **Provide operational recommendations on model selection** based on weather conditions. The findings will be synthesized into actionable guidance for grid operators and forecasting system designers, suggesting optimal model selection strategies that adapt to prevailing weather stability characteristics.

These objectives collectively support the overarching goal of improving renewable energy forecasting reliability through weather-aware model selection, contributing to more effective renewable energy integration and grid management.

1.4 Thesis Structure

This thesis is organized into seven main chapters, following a logical progression from problem identification through methodology development to analysis and conclusions.

Chapter 1 (Introduction) establishes the research context, motivation, and objectives. This chapter presents the research questions and outlines the overall structure of the thesis.

Chapter 2 (Literature Review) reviews existing research on renewable energy forecasting, weather regime classification, and model evaluation methodologies. It identifies gaps in current knowledge and situates this research within the broader academic discourse.

Chapter 3 (Methodology) presents the research design and methodological framework. The chapter describes the dual-pipeline architecture, which consists of (1) a Weather Stability Classification pipeline that processes meteorological data to identify stable and unstable periods, (2) a Renewable Energy Prediction Models pipeline that implements and evaluates both statistical and deep learning forecasting models, and (3) a Comparative Analysis pipeline that merges

stability classifications with model performance metrics to assess robustness. This chapter also details data collection procedures, preprocessing steps, and analysis methods.

Chapter 4 (Data and Experimental Setup) provides comprehensive documentation of the datasets used in this research, including weather data from the German Meteorological Service (DWD) and renewable energy production data. The chapter describes data sources, collection procedures, quality assurance measures, and experimental configuration.

Chapter 5 (Results) presents the empirical findings, including Weather Stability Index characteristics, model performance metrics stratified by weather regime, statistical test results, and model robustness rankings. Results are presented through descriptive statistics, tables, and visualizations.

Chapter 6 (Discussion) interprets the results in the context of existing literature, discusses implications for operational forecasting systems, acknowledges limitations of the study, and explores the practical significance of the findings.

Chapter 7 (Conclusion) summarizes the main contributions of this research, revisits the research questions, discusses their resolution, and suggests directions for future work.

The thesis follows a structured approach where the dual-pipeline methodology enables systematic evaluation of model robustness to weather instability, ultimately converging on comparative analysis that addresses all three research questions.

2 Literature Review

2.1 Introduction

This chapter provides a comprehensive review of existing literature relevant to weather stability analysis in renewable energy forecasting. The review synthesizes research from three interconnected domains: (1) weather regime classification and meteorological stability analysis, (2) statistical and machine learning models for renewable energy forecasting, and (3) the integration of weather classification into forecasting model selection and evaluation.

The chapter is organized to establish the theoretical foundations, examine comparative model performance, explore weather-classified forecasting approaches, and identify gaps in current knowledge that this research addresses. This review situates the current work within the broader academic discourse and provides the context necessary to understand the significance and contributions of the proposed methodology.

2.2 Background Theory

2.2.1 Weather Regime Classification Fundamentals

Weather stability represents a fundamental concept in meteorology, describing the persistence of atmospheric conditions over time. Stable weather regimes are characterized by consistent meteorological patterns with minimal variability, while unstable regimes exhibit rapid transitions, high variability, and non-stationary behavior (5; 6; 7). The distinction between stable and unstable weather conditions is crucial for understanding atmospheric dynamics and has significant implications for renewable energy generation patterns.

Weather systems operate across multiple temporal scales, each with characteristic time horizons. Mesoscale phenomena, such as convective systems and local wind patterns, operate over 2–6 hour periods, while synoptic-scale systems, including frontal passages and pressure changes, occur over 6–12 hour windows. Planetary-scale patterns, encompassing large-scale atmospheric circulation, operate over days to weeks. These temporal scales inform appropriate window sizes for weather stability classification and align with operational forecasting horizons for renewable energy systems.

The concept of weather regimes has been extensively studied in climatology and meteorology. Huth et al. (5) provide a comprehensive review of atmospheric circulation pattern classifications,

highlighting the importance of regime-based approaches for understanding climate variability. Michelangeli et al. (6) demonstrate that weather regimes exhibit quasi-stationarity and recurrence, making them suitable for classification and prediction. Vautard (7) establishes the theoretical foundation for multiple weather regime detection over the North Atlantic, emphasizing the importance of identifying regime transitions and persistence.

2.2.2 Statistical Methods for Weather Classification

Weather stability classification requires methods capable of identifying distinct atmospheric states from multivariate meteorological data. Several statistical approaches have been developed for this purpose, each with distinct advantages and limitations.

Threshold-based methods provide simple, interpretable classifications but suffer from arbitrary cutoff selection and limited sensitivity to regime transitions. K-means clustering offers an unsupervised approach but assumes spherical clusters and is sensitive to initialization, potentially missing non-linear patterns in meteorological data.

Principal Component Analysis (PCA) provides linear dimensionality reduction but may fail to capture non-linear relationships inherent in weather systems. Gaussian Mixture Models (GMM) offer a probabilistic framework for regime detection, providing uncertainty quantification and handling non-spherical cluster shapes (8). GMMs model the distribution of weather features as a mixture of Gaussian components, with each component representing a distinct weather regime.

Hidden Markov Models (HMM) explicitly account for temporal dependence in weather stability, recognizing that atmospheric states exhibit persistence and that transitions between regimes follow probabilistic patterns (9). HMMs model weather stability as a hidden state sequence, where observations (weather features) are emitted probabilistically from underlying stable or unstable states. The Viterbi algorithm enables estimation of the most likely state sequence, providing robust temporal smoothing of classifications.

Robust normalization methods are essential for weather stability analysis, as meteorological data frequently contains outliers and extreme events. Median and interquartile range (IQR) based normalization provides resistance to outliers while maintaining sensitivity to legitimate extreme weather events.

2.3 Renewable Energy Forecasting Models

2.3.1 Statistical Models

Statistical models have been the foundation of renewable energy forecasting for decades, providing transparent, interpretable predictions with modest computational requirements. These models are particularly effective for short-term forecasting in stable conditions with linear or near-linear relationships.

ARIMA and SARIMA Models

Autoregressive Integrated Moving Average (ARIMA) models and their seasonal variants (SARIMA) represent classical time series forecasting approaches. These models capture temporal dependencies through autoregressive terms, account for non-stationarity through differencing, and model residual structure through moving average components. For renewable energy forecasting, SARIMA models are particularly relevant as they explicitly model seasonal patterns inherent in solar and wind generation (1; 2).

However, ARIMA/SARIMA models face limitations when dealing with non-linear relationships, high-dimensional input spaces, and long-term forecasting horizons. Their performance degrades rapidly with increased data complexity or when external factors such as weather variability introduce non-stationarity (15; 16; 17).

Exponential Smoothing Methods

Exponential smoothing methods, including Holt-Winters seasonal decomposition, provide robust forecasting for time series with trends and seasonality. These methods apply exponentially decreasing weights to historical observations, giving greater importance to recent data. Triple exponential smoothing (Holt-Winters) explicitly models trend and seasonal components, making it suitable for renewable energy forecasting where both diurnal and annual cycles are present.

Prophet Model

The Prophet model, developed by Facebook, provides an additive seasonality framework designed for business forecasting with automatic handling of holidays, trends, and seasonality. Prophet has been successfully applied to renewable energy forecasting, particularly for solar power, where it handles multiple seasonal patterns and trend changes effectively.

Strengths and Limitations of Statistical Models

Statistical models offer several advantages: (1) interpretability and transparency, enabling understanding of model behavior and regulatory compliance; (2) computational efficiency, requiring minimal resources; (3) effectiveness in stable, linear scenarios with sufficient historical data; and (4) established theoretical foundations with well-understood properties (54; 55; 53).

However, statistical models exhibit significant limitations: (1) limited ability to capture non-linear relationships inherent in complex weather-energy interactions; (2) performance degradation with increasing forecast horizons; (3) difficulty handling high-dimensional input spaces with multiple meteorological variables; and (4) reduced accuracy during volatile or extreme weather conditions (15; 37; 46).

2.3.2 Machine Learning and Deep Learning Models

Machine learning and deep learning models have emerged as powerful alternatives to statistical approaches, particularly for complex, non-linear forecasting tasks with large datasets. These

models excel at capturing intricate patterns, temporal dependencies, and high-dimensional relationships that challenge traditional statistical methods.

Long Short-Term Memory (LSTM) Networks

LSTM networks represent a class of recurrent neural networks designed to capture long-term temporal dependencies while avoiding the vanishing gradient problem of traditional RNNs. For renewable energy forecasting, LSTMs effectively model sequences of weather and generation data, learning complex temporal patterns that influence production (18; 19; 20).

Multiple studies demonstrate that LSTM models consistently outperform statistical approaches in renewable energy forecasting, particularly for solar and wind power prediction (15; 16; 22; 23). LSTMs show particular strength in handling non-stationary patterns and adapting to changing weather conditions.

Convolutional Neural Networks (CNN) and Hybrid Architectures

CNNs, originally developed for image processing, have been adapted for time series forecasting through one-dimensional convolutions that capture local patterns in temporal sequences. Hybrid CNN-LSTM architectures combine the pattern recognition capabilities of CNNs with the temporal modeling of LSTMs, achieving superior performance in renewable energy forecasting (25; 26; 33).

Studies consistently report that hybrid CNN-LSTM models outperform standalone deep learning models and significantly exceed statistical model performance, with improvements of up to 37% in prediction accuracy (36; 38; 27).

Temporal Convolutional Networks (TCN)

TCN represents a modern architecture combining dilated convolutions with residual connections to achieve long-range temporal modeling without recurrent structures. TCNs provide parallel processing advantages over RNNs while maintaining comparable or superior forecasting performance (47; 48).

Ensemble and Meta-Learning Approaches

Ensemble methods combine predictions from multiple models, leveraging diversity to improve accuracy and robustness. Meta-learning frameworks automate model selection based on weather conditions, dynamically blending forecasts from different architectures to optimize performance for specific meteorological regimes (35; 32).

Strengths and Limitations of Deep Learning Models

Deep learning models offer significant advantages: (1) superior ability to capture non-linear relationships and complex patterns; (2) adaptation to non-stationary data and changing conditions; (3) handling of high-dimensional input spaces with multiple meteorological variables; and (4) continued performance improvement with increasing data volume and complexity (18; 19; 20).

However, deep learning models face challenges: (1) substantial data requirements, with performance dependent on large, high-quality datasets; (2) significant computational resources for training and inference; (3) limited interpretability, creating “black box” concerns for operational deployment; (4) risk of overfitting and sensitivity to hyperparameter selection; and (5) reduced effectiveness in data-scarce scenarios (18; 19; 36; 39).

2.3.3 Comparative Performance Analysis

The literature reveals a clear consensus: deep learning models consistently outperform traditional statistical models in most energy forecasting scenarios, especially when dealing with complex, non-linear, and high-dimensional data (15; 16; 17; 18; 19; 20). For example, Cabello-López et al. (16) report that deep learning models reduced mean absolute error by up to 47% compared to official statistical forecasts in national-level solar energy prediction.

The performance gap between statistical and deep learning models widens with increasing data complexity, forecast horizon, and weather variability. Statistical models perform competitively in simple, linear, or data-scarce scenarios, but their performance plateaus with increasing complexity, while deep learning models continue to improve with richer data and more complex feature sets (15; 39; 40; 42).

Hybrid models that combine statistical and deep learning approaches leverage the strengths of both paradigms, achieving superior accuracy and robustness. These hybrid approaches often outperform standalone models by combining statistical stability with deep learning flexibility (17; 25; 23; 27; 24).

2.4 Weather Classification in Renewable Energy Forecasting

2.4.1 Incorporating Weather Information

The integration of weather classification into renewable energy forecasting has become increasingly important, as weather conditions are the primary driver of renewable energy variability. Weather classification methods segment meteorological data into homogeneous regimes or clusters, each associated with distinct characteristics that influence generation patterns.

Weather classification is typically achieved through clustering algorithms (K-means, DBSCAN), regime identification methods (Lamb Weather Types), or feature selection based on meteorological variables such as cloud cover, wind speed, temperature, and atmospheric pressure (28; 29; 30). These methods enable targeted model application, where different forecasting models can be selected or weighted based on prevailing weather conditions.

Studies consistently report that incorporating weather classification improves model performance for both statistical and deep learning approaches, with gains more pronounced for deep learning models (37; 38; 40; 41; 42; 43). Weather regime identification enhances feature relevance and model performance by focusing learning on meteorologically coherent data segments.

2.4.2 Weather-Driven Model Selection

Recent research explores adaptive frameworks where model selection is dynamically informed by prevailing weather regimes (28; 29; 30; 31; 32). Rather than applying a single forecasting model universally, these approaches select or weight models based on weather type, optimizing accuracy by leveraging model strengths under specific conditions.

Wang et al. (28) demonstrate that optimal model type varies with weather regime: simpler statistical or regression models often suffice under stable meteorological conditions, while deep learning models (e.g., LSTM, CNN-LSTM hybrids) outperform others during periods of high variability or extreme weather. This finding suggests that weather-aware model selection can significantly improve forecasting accuracy.

Cervantes et al. (29) apply weather classification to solar radiation forecasting across Köppen climate zones, showing that clustering by climate type and data dispersion leads to optimal model selection for each region. Multivariate linear models performed best in tropical climates, while polynomial models excelled in arid climates, demonstrating the importance of region-specific, weather-informed model selection.

Meta-learning frameworks automate this selection process, using machine learning to learn which base models perform best under specific weather conditions (35; 32). These systems blend forecasts from multiple models, assigning weights or selecting optimal models dynamically based on weather features.

2.4.3 Performance Across Weather Regimes

The literature demonstrates that model performance is highly sensitive to weather type, with distinct models excelling under different conditions. Under stable meteorological conditions, statistical models remain competitive and often provide sufficient accuracy with lower computational requirements (29; 28; 31; 32).

However, during unstable or extreme weather periods, deep learning models consistently outperform statistical approaches. Deep learning models' ability to capture non-linearities and adapt to complex weather patterns makes them particularly effective during variable conditions (28; 33; 34). Hybrid and ensemble approaches show superior performance across all weather regimes, combining the stability of statistical models with the adaptability of deep learning approaches (44; 45).

2.5 Model Robustness and Evaluation

2.5.1 Statistical Validation Methods

Validating the relationship between weather stability and model performance requires rigorous statistical methods capable of handling non-normal error distributions and temporal dependencies. The Mann-Whitney U test provides a non-parametric alternative to t-tests, robust to non-normal distributions and appropriate for skewed error distributions common in forecasting (10).

Welch's t-test offers a parametric alternative that accounts for unequal variances between stable and unstable groups.

Effect size measures, including Cohen's d and percentage increase in error, quantify the practical significance of performance differences beyond statistical significance (11). Multiple testing corrections, such as the Bonferroni method and False Discovery Rate control (12), prevent inflation of Type I error when comparing multiple models across different weather regimes.

Linear mixed-effects models provide a comprehensive framework for comparing model performance while accounting for temporal clustering, repeated measures, and confounding variables. These models enable simultaneous evaluation of weather stability effects, model-specific differences, and interaction effects between weather conditions and model types.

2.5.2 Robustness Metrics

Model robustness quantifies the ability of forecasting models to maintain performance across different conditions. Relative Performance Degradation (RPD) measures performance change from stable to unstable conditions relative to baseline performance, normalizing for model-specific accuracy levels. RPD provides a fair comparison across models with different baseline errors, identifying which models maintain accuracy most effectively during unstable weather.

Absolute Performance Degradation (APD) captures the operational impact of performance changes, useful when baseline accuracy matters for grid operations. Composite robustness indices combine accuracy and robustness metrics, balancing “good on average” performance with “consistent across conditions” behavior.

2.5.3 Experimental Design for Comparative Analysis

Rigorous comparative analysis requires careful experimental design. Data segmentation strategies divide datasets into stable and unstable weather periods using weather classification or clustering methods (28; 47; 48). Models are then trained and evaluated on each segment, enabling systematic comparison of performance across weather regimes.

Evaluation metrics including MAE, RMSE, MAPE, and skill scores provide comprehensive performance assessment. The experimental approach recommended in the literature involves training both statistical and machine learning models on each weather segment, using multiple metrics to compare performance, and interpreting results to determine model sufficiency across conditions (47; 48; 49; 50).

2.6 Gap Analysis

2.6.1 Identified Research Gaps

Despite significant progress in renewable energy forecasting, several critical gaps remain in the literature regarding weather stability impact on model performance:

Limited Systematic Evaluation: Most studies evaluate model performance on aggregate datasets without explicitly accounting for weather regime characteristics. There is insufficient research systematically comparing model robustness across stable and unstable weather conditions using standardized methodologies.

Insufficient Weather Stability Focus: While weather classification is increasingly incorporated into forecasting models, there is limited research specifically examining how weather stability (as opposed to weather type) affects model performance. The distinction between stable and unstable conditions within similar weather types (e.g., sunny days with varying cloud cover variability) remains underexplored.

Gaps in Long-Term Forecasting: Research on weather-classified forecasting predominantly focuses on short-term horizons. Long-term forecasting with weather classification and robustness analysis across weather regimes represents a significant gap.

Limited Interpretability Research: Deep learning models' superior performance comes with reduced interpretability, creating barriers to operational adoption. Research on interpretable deep learning for weather-classified forecasting remains limited.

Generalizability Challenges: Most studies focus on specific geographic regions or energy types. The generalizability of weather-driven model selection frameworks across diverse climates and energy sources requires further investigation.

2.6.2 Research Questions Alignment

This research addresses these gaps by: (1) developing a systematic framework for weather stability classification using quantitative indices and validated statistical methods; (2) conducting comprehensive comparative analysis of statistical and machine learning models across stable and unstable weather regimes; (3) providing rigorous statistical validation of error-stability relationships with appropriate effect size measures; and (4) establishing robustness metrics that enable operational model selection guidance.

The dual-pipeline methodology enables independent evaluation of weather stability classification and model performance, preventing data leakage and ensuring methodological rigor. The comparative robustness analysis provides novel insights into model behavior across weather conditions, addressing a critical gap in current literature.

2.7 Summary

This literature review has established the theoretical foundations and current state of research in weather stability analysis for renewable energy forecasting. Key findings include:

- Weather regime classification provides a valuable framework for understanding atmospheric variability and its impact on renewable energy generation patterns.
- Statistical models remain effective for simple, linear, or data-scarce scenarios but face limitations with increasing complexity, non-linearity, and weather variability.

- Deep learning models consistently outperform statistical approaches in complex forecasting scenarios, particularly when weather classification is incorporated, but face challenges related to data requirements, computational costs, and interpretability.
- Weather-driven model selection represents an emerging approach that dynamically selects models based on prevailing weather conditions, consistently improving forecasting accuracy.
- Significant gaps remain in systematic evaluation of model robustness across weather stability regimes, creating an opportunity for methodological contributions.

The following chapter presents the methodology developed to address these gaps, providing a comprehensive framework for weather stability classification and comparative model robustness analysis. The dual-pipeline approach ensures rigorous evaluation while enabling practical insights for operational forecasting systems.

3 Methodology

3.1 Research Design

The research design follows a dual-pipeline architecture that enables systematic evaluation of model performance under different weather stability regimes. This design addresses the research questions (Section 1.2) by independently processing weather classification and energy prediction before combining them for comparative analysis.

3.1.1 Pipeline Architecture

The methodology consists of three interconnected pipelines that operate in parallel and converge for final analysis:

Pipeline 1: Weather Stability Classification

This pipeline processes meteorological data to identify stable and unstable weather periods:

- **Input:** Eleven weather attributes collected for 2024 at hourly resolution, covering Germany-wide and 16 Bundesländer (federal states). The attributes include:
 - Temperature (mean, minimum, maximum)
 - Cloudiness (cloud cover percentage)
 - Dew point (atmospheric moisture indicator)
 - Extreme wind (peak wind measurements)
 - Moisture (relative humidity)
 - Precipitation (rainfall in mm)
 - Pressure (atmospheric pressure in hPa)
 - Soil temperature
 - Sun (sunshine duration)
 - Visibility (horizontal visibility)
 - Weather phenomena (categorical weather events)
 - Wind and wind_synop (wind speed and direction)

- **Process:** The pipeline applies feature engineering to extract variability, trend, and extreme event indicators from raw weather data. These features are used to compute a Weather Stability Index (WSI) that quantifies atmospheric variability. The WSI is then classified into stable and unstable regimes using Gaussian Mixture Models (GMM) (8) with temporal smoothing via Hidden Markov Models (HMM) (9) to account for regime persistence.
- **Output:** A timeline with weather stability labels (stable/unstable) for each hour in the analysis period, enabling stratification of subsequent analyses by weather regime.

Pipeline 2: Renewable Energy Prediction Models

This pipeline implements and evaluates both statistical and deep learning forecasting models:

- **Input:** Renewable energy production data (solar PV and wind power) for 2024, along with relevant weather features. For solar prediction, key inputs include solar radiation (if available), cloudiness, temperature, and temporal features (hour of day, day of year, day of week). For wind prediction, inputs include wind speed, wind direction, atmospheric pressure, temperature, and temporal features.
- **Process:** Multiple forecasting models are implemented and evaluated, including:
 - **Statistical models:**
 - * Persistence model (baseline: next-hour equals current-hour)
 - * Seasonal persistence (same hour from previous week)
 - * ARIMA/SARIMA (autoregressive integrated moving average with seasonal components)
 - * Prophet (additive seasonality model by Facebook)
 - * Exponential smoothing (Holt-Winters with trend and seasonality)
 - **Deep learning models:**
 - * LSTM (Long Short-Term Memory networks)
 - * CNN-LSTM (hybrid convolutional and LSTM architectures)
 - * TCN (Temporal Convolutional Networks)
 - * Ensemble approaches combining multiple models

Models are trained using walk-forward validation, where each model is trained on a rolling window of historical data and evaluated on subsequent periods, mimicking operational forecasting conditions.

- **Output:** Hourly predictions for renewable energy production along with performance metrics (MAE, RMSE, MAPE) computed at each time step for all models.

Pipeline 3: Comparative Analysis

This pipeline combines outputs from the previous two pipelines:

- **Input:** Weather stability timeline from Pipeline 1 and model performance timeline from Pipeline 2.
- **Process:** The stability labels are merged with model performance metrics, creating a unified dataset where each observation includes timestamp, weather regime (stable/unstable), model predictions, actual production, and error metrics. Statistical tests are then applied to compare performance between stable and unstable periods for each model. This includes:
 - Mann-Whitney U tests (10) (non-parametric comparison)
 - Welch's t-tests (parametric alternative)
 - Linear mixed-effects models (accounting for temporal clustering)
 - Effect size calculations (11) (Cohen's d, percentage increase)

Multiple testing correction procedures (12) are applied to control the false discovery rate across model comparisons. Models are ranked by robustness metrics, particularly Relative Performance Degradation (RPD), which measures performance change from stable to unstable conditions relative to baseline performance.

- **Output:** Model rankings, statistical test results, robustness assessments, and operational recommendations on model selection based on weather conditions.

3.1.2 Design Rationale

This dual-pipeline design addresses the research questions as follows:

- **RQ1** is addressed by Pipeline 3's statistical comparisons between stable and unstable periods.
- **RQ2** is answered through robustness rankings computed in Pipeline 3, identifying which models show smallest performance degradation.
- **RQ3** is evaluated by synthesizing findings into operational recommendations that leverage weather stability information for model selection.

The design ensures that weather classification and model evaluation are performed independently, preventing data leakage and maintaining methodological rigor. The convergence point in Pipeline 3 enables systematic comparison while preserving the integrity of each component pipeline.

3.2 Data Collection

This section describes the data sources and collection procedures for both weather and renewable energy datasets used in this research.

3.2.1 Weather Data

Data Provider

Weather data is sourced from the **Deutscher Wetterdienst (DWD)**, the German Meteorological Service, which operates a comprehensive network of weather stations throughout Germany. DWD provides open-access climate data through its climate data center, making it suitable for academic research.

Data Characteristics

The collected weather data has the following specifications:

- **Source:** Hourly weather observations from DWD climate data center
- **Period:** 2024 (full calendar year, January 1 to December 31)
- **Spatial Coverage:**
 - Germany-wide aggregated data
 - 16 Bundesländer (federal states) individually
 - 636+ individual weather stations mapped to regions
- **Station Network:** 636+ active weather stations distributed across Germany
- **Data License:** Open data license, freely available for research use
- **Temporal Resolution:** Hourly observations (8,760 hours per year)

Weather Attributes Collected

Eleven weather attributes have been collected, each contributing to the Weather Stability Index computation:

1. **Temperature** - hourly mean, minimum, and maximum values
2. **Cloudiness** - cloud cover percentage (0-100%)
3. **Dew point** - atmospheric moisture indicator
4. **Extreme wind** - peak wind measurements
5. **Moisture** - relative humidity percentage
6. **Precipitation** - rainfall amount in millimeters
7. **Pressure** - atmospheric pressure in hectopascals (hPa)
8. **Soil temperature** - ground temperature measurements
9. **Sun** - sunshine duration in hours
10. **Visibility** - horizontal visibility in meters

11. **Weather phenomena** - categorical weather events
12. **Wind and wind_synop** - wind speed (m/s) and wind direction (degrees)

These attributes capture multiple dimensions of atmospheric conditions, enabling comprehensive quantification of weather variability and stability.

3.2.2 Renewable Energy Data

Data Status and Requirements

Renewable energy production data is required for model training and evaluation. The following specifications are needed:

- **Period:** 2024 (full year) to match weather data temporal coverage
- **Resolution:** Hourly production data aligned with weather observations
- **Geographic Scope:**
 - Germany-wide total production (primary requirement)
 - Bundesland-level production (desirable for regional analysis)
- **Energy Types:**
 - Solar PV production (MW or GWh)
 - Wind power production (MW or GWh)
- **Data Quality:** Complete temporal coverage (all 8,760 hours), validated against installed capacity statistics

Potential Data Sources

Several potential data sources have been identified for renewable energy production data:

1. ENTSO-E Transparency Platform (transparency.entsoe.eu)

- European transmission system operators' data platform
- Provides hourly generation data by technology type
- Free access with API key registration
- Historical data available for Germany

2. Fraunhofer ISE Energy Charts (energy-charts.info)

- Public dashboard for German renewable energy statistics
- Provides downloadable CSV files
- Historical data available
- Well-documented and widely used in research

3. SMARD (Strommarktdaten) (smard.de)

- Federal Network Agency (Bundesnetzagentur) platform
- Official market data including renewable generation
- Free registration required
- Authoritative source for German energy market data

Final data source selection will be based on data availability, completeness, accessibility, and alignment with research requirements. Data validation procedures will ensure consistency with known installed capacity statistics and detect any anomalies or missing values.

3.3 Data Preprocessing

This section documents the data preprocessing steps that have been completed for weather data and outlines additional steps required for dataset integration and analysis. The preprocessing pipeline ensures data quality, consistency, and compatibility across different data sources.

3.3.1 Completed Weather Data Preprocessing

The following preprocessing steps have been completed for the 2024 Germany weather dataset:

Step 1: Download and Initial Processing

- Downloaded raw weather data files from DWD servers for all 11 weather attributes
- Converted data format from semicolon-delimited TXT files to CSV format for easier processing
- Standardized timestamp format in the `MESS_DATUM` column to ISO 8601 format (YYYY-MM-DD HH:MM:SS)
- Removed metadata and description files that are automatically included in DWD data packages
- Cleaned up file structure and organized files by attribute type
- Filtered data to include only 2024 observations (removed any historical data from multi-year files)
- Logged all download and processing operations in the `logs/` directory for traceability

Step 2: Bundesland Aggregation

- Mapped 636 weather stations to 16 German Bundesländer using a reference file (`regions.csv`) that contains station ID to Bundesland assignments
- Aggregated station-level data into Bundesland-level files by combining all stations within each federal state

- Created 16 CSV files per weather attribute (one file per Bundesland), preserving station IDs for traceability
- Output structure: Data/*_by_bundesland/*.csv (e.g., Data/temperature_by_bundesland/Bayern.csv)
- Maintained data integrity by preserving all original columns while adding aggregation metadata

Step 3: Germany-Wide Aggregation

- Combined all Bundesland data into single Germany-wide aggregated files
- Aggregated across all 16 federal states to create country-level time series
- Output structure: Data/*_germany_aggregated/Germany_total.csv
- Enables analysis at both regional (Bundesland) and national (Germany-wide) scales

Step 4: Data Quality Assurance

- Removed empty files that contained no data for the 2024 period
- Validated data completeness across all 13 weather attributes
- Documented missing data patterns and temporal gaps
- Generated data inventory summaries listing available stations, coverage periods, and data quality indicators
- Identified and flagged potential outliers using domain knowledge (e.g., unrealistic temperature values)

Step 5: Comprehensive Preprocessing with Imputation

A comprehensive preprocessing pipeline has been implemented for all 13 weather attributes using statistical imputation instead of row removal. This approach preserves the temporal continuity of the time series while ensuring data quality. The preprocessing steps are as follows:

1. **Missing Value Identification:** Missing value markers (-999, -999.0, "-999", "-999.0") are replaced with NaN to standardize missing data representation across all attributes.
2. **Outlier Detection:** Three complementary methods are used to identify outliers:
 - **Domain-specific thresholds:** Values outside physically plausible ranges (e.g., temperature < -50°C or > 50°C, wind speed > 200 km/h) are flagged as outliers
 - **Z-score method:** Values exceeding 3 standard deviations from the mean are flagged
 - **Interquartile Range (IQR) method:** Values outside $1.5 \times \text{IQR}$ from Q1/Q3 are flagged

Outliers are identified using the union (OR operation) of these three methods.

3. **Outlier Removal:** Detected outlier values are replaced with NaN (rather than removing entire rows) to preserve temporal continuity.

4. **Statistical Imputation:** Missing values (including removed outliers) are imputed using distribution-aware methods:

- **Mean imputation:** Applied when data distribution is approximately normal. Normality is tested using the Shapiro-Wilk test ($p > 0.05$) and skewness assessment ($|skewness| < 1$).
- **Median imputation:** Applied when data distribution is skewed or non-normal, providing robustness against extreme values.

The distribution assessment is performed dynamically for each attribute and each Bundesland/Germany dataset to account for regional variations in data characteristics.

5. **Data Preservation:** Unlike row deletion methods, this approach maintains all 8,760 hourly observations per location, ensuring complete temporal coverage for time series analysis.

Output Structure: Preprocessed data is saved to a separate directory structure (`Preprocessed_Data/`) to preserve original data integrity. The output maintains the same hierarchical structure as the original data:

- `Preprocessed_Data/Bundesland_aggregation/` - Contains preprocessed data for all 16 Bundesländer
- `Preprocessed_Data/Germany_aggregation/` - Contains preprocessed Germany-wide aggregated data

Attributes Processed: The preprocessing pipeline is applied to all 13 weather attributes: temperature, cloudiness, wind, wind_synop, precipitation, pressure, dew_point, moisture, extreme_wind, soil_temperature, sun, and visibility. Each attribute uses its specific domain thresholds and statistical properties for appropriate outlier detection and imputation.

Quality Metrics: Comprehensive preprocessing reports are generated documenting:

- Number of missing value markers replaced per file
- Number of outliers detected and imputed per file
- Distribution of mean vs. median imputations (indicating data distribution characteristics)
- Validation errors and data quality statistics

This preprocessing methodology ensures data quality while maximizing data retention, which is critical for time series analysis and machine learning applications.

3.3.2 Future Preprocessing Steps for Integration

The following preprocessing steps will be required when integrating weather and renewable energy datasets:

Timestamp Harmonization

- Standardize all timestamps to a consistent timezone (UTC recommended to avoid daylight saving time complications)
- Ensure ISO 8601 format throughout: YYYY-MM-DD HH:MM:SS
- Create a master time index covering all 8,760 hours of 2024 (accounting for daylight saving time transitions)
- Verify temporal alignment between weather and energy datasets

Missing Data Strategy

Note: This section describes planned future steps. The current preprocessing (Step 5 above) already handles missing values using statistical imputation (mean/median based on distribution).

For future integration, a tiered approach could be applied to handle missing values in merged datasets:

- **Tier 1 (2 hours missing):** Linear interpolation between adjacent valid observations
- **Tier 2 (3-6 hours missing):** Forward-fill using the last valid observation, with a flag indicating imputation
- **Tier 3 (>6 hours missing):** Statistical imputation (mean/median) based on distribution characteristics
- All imputations are logged in preprocessing reports documenting imputation type, count, and affected variables
- Data completeness is tracked throughout the preprocessing pipeline

Dataset Merging

A unified analysis dataset will be created containing:

- Timestamp (hourly, aligned across all variables)
- Location identifier (Germany-wide or Bundesland)
- All 11 weather attributes (normalized and feature-engineered)
- Renewable energy production values (solar and wind)
- Weather Stability Index (WSI) values
- Stability classification labels (stable/unstable)
- Model predictions (for each implemented model)
- Performance metrics (MAE, RMSE, MAPE per time step)
- Data quality flags (missing data indicators, imputation flags, exclusion markers)

This unified dataset enables efficient analysis and ensures consistent temporal alignment across all variables.

Quality Checks

Final quality assurance procedures will verify:

- No duplicate timestamps in the merged dataset
- No temporal gaps in the time series
- Reasonable value ranges (e.g., temperature between -40°C and 50°C for Germany)
- Renewable energy production values consistent with installed capacity limits
- Summary statistics computed for all variables to identify potential data issues
- Cross-validation of aggregated values against source data

All preprocessing steps are documented with scripts stored in the `Scripts/` directory, ensuring reproducibility and transparency of the data preparation process.

3.4 Weather Stability Index Development

The Weather Stability Index (WSI) is developed using established methods for weather regime detection (5; 6; 7). This section describes the development methodology, including feature engineering approaches, WSI computation using robust normalization, Gaussian Mixture Model (GMM) classification (8) for stable/unstable regime detection, and validation criteria. The implementation follows a multi-level hierarchical classification framework that combines instantaneous stability metrics with temporal context to produce robust classifications.

3.4.1 Feature Engineering

Feature engineering transforms raw weather observations into meaningful stability indicators. The selected features are based on meteorological principles where stability is characterized by low variability, consistent trends, and absence of extreme events. This multi-scale approach aligns with the hierarchical classification framework recommended in the methodology.

Variability Features

Atmospheric variability is a key indicator of weather instability. Variability features are computed using a 24-hour rolling window, which captures diurnal cycles while identifying periods of abnormal variability. This temporal scale aligns with synoptic-scale meteorology where weather systems operate on characteristic timescales (4).

The following variability features are computed:

- **Temperature standard deviation (σ_T):** Computed over a 24-hour rolling window as $\sigma_T = \sqrt{\frac{1}{n-1} \sum_{i=t-23}^t (T_i - \bar{T})^2}$ where $n = 24$. High temperature variability indicates atmospheric instability and transition periods between air masses (4).

- **Temperature range:** Maximum minus minimum temperature over 24-hour window, capturing extreme temperature swings within a day.
- **Pressure change:** Absolute change in pressure over 24 hours, $|\Delta P| = |P_t - P_{t-24}|$. Rapid pressure changes ($> 5 \text{ hPa}/24\text{h}$) indicate storm development or frontal passages (4).
- **Wind coefficient of variation:** $CV_{wind} = \sigma_{wind}/\mu_{wind}$ over 24-hour window. This normalized variability measure accounts for different baseline wind speeds, making it comparable across seasons. High CV indicates gusty, unstable conditions associated with convective activity.
- **Precipitation intensity:** Rolling sum over 3-hour window, $P_{int}(t) = \sum_{i=t-2}^t P_i$. The 3-hour window captures convective precipitation events and rapid changes in precipitation intensity (?).
- **Humidity standard deviation:** Dew point standard deviation over 24-hour window. High variability indicates air mass changes, convective mixing, or frontal passages.

Trend Features

Linear trends capture directional changes in atmospheric conditions that signal transitions between weather regimes. Trend analysis is well-established in climatology for detecting regime changes (?). Trend features are computed using linear regression (degree 1) over a 24-hour window:

- **Temperature trend:** Slope of linear regression $T_i = \beta_0 + \beta_1 \cdot i + \epsilon_i$ over 24-hour window. Steep temperature trends indicate rapid atmospheric changes and regime transitions.
- **Pressure trend:** Barometric pressure trend slope. Pressure trends predict weather system movement and are fundamental to synoptic meteorology (4). Falling pressure indicates approaching storms; rising pressure indicates clearing conditions.
- **Wind trend:** Wind speed trend slope. Wind speed changes reflect pressure gradient changes and are indicators of approaching weather systems.

Extreme Event Flags

Extreme weather events are clear indicators of atmospheric instability. Binary flags provide interpretable, domain-expert validated indicators that complement continuous variability measures (?). The following flags are computed:

- **High wind flag:** Binary indicator when wind speed exceeds the 90th percentile threshold. High winds indicate strong pressure gradients and unstable conditions.
- **Heavy precipitation flag:** Binary indicator when precipitation exceeds 5 mm threshold. This threshold is meteorologically significant and represents moderate to heavy precipitation indicating active weather systems (?).
- **Rapid temperature change flag:** Binary indicator when $|\Delta T| > 5C$ in 3 hours. This threshold is meteorologically significant and indicates rapid atmospheric changes such as

frontal passages (?).

- **Storm flag:** Combined condition flag when wind exceeds 90th percentile AND pressure drop exceeds 5 hPa in 6 hours. Multi-variate extreme events are stronger indicators than single variables, reducing false positives.

Feature Selection and Normalization

Feature selection reduces multicollinearity, improves model interpretability, and prevents overfitting. Highly correlated features ($|r| > 0.9$) are identified using Pearson correlation coefficient and redundant features are removed, targeting a final feature set of 6–10 features (?).

Robust normalization is applied using median and interquartile range (IQR):

$$x_{normalized} = \frac{x - \text{median}(x)}{\text{IQR}(x)} \quad (3.1)$$

where $\text{IQR} = Q_3 - Q_1$ (interquartile range). Robust scaling is preferred over z-score normalization for weather data because it is resistant to outliers and extreme events (13?). This is critical for weather data which contains legitimate extreme events that should not dominate normalization. All features are oriented so higher values indicate more instability.

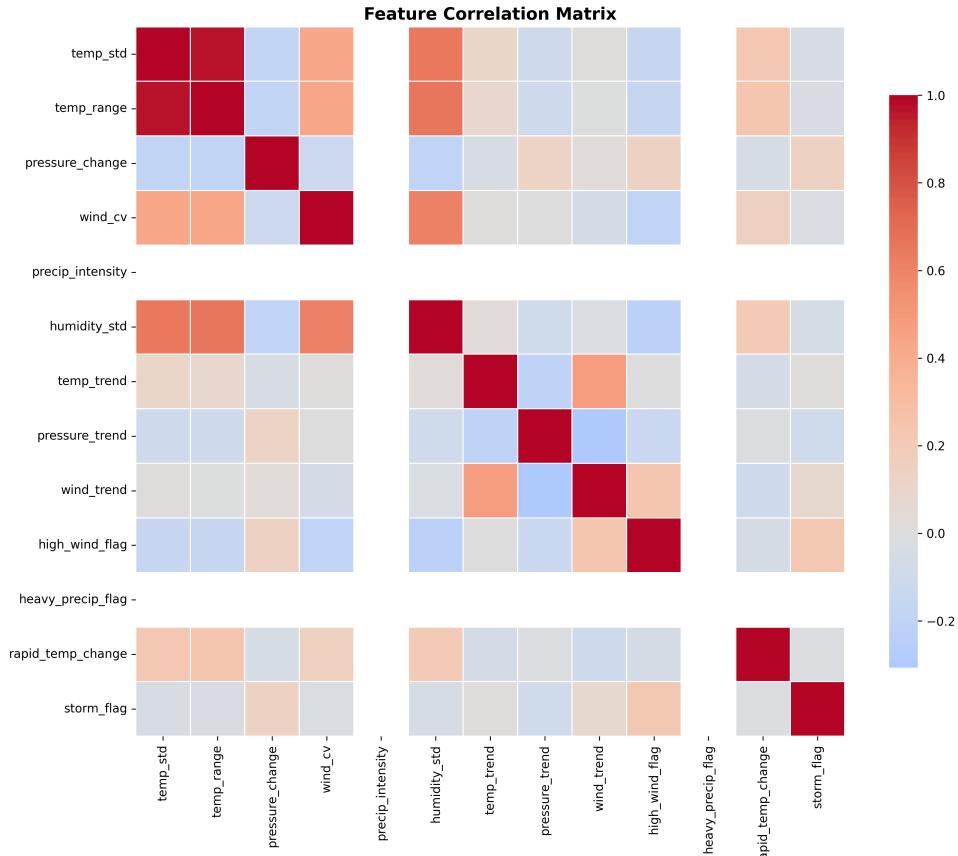


Figure 3.1: Feature correlation matrix used for removal of highly correlated features ($|r| > 0.9$).

3.4.2 WSI Computation

The instantaneous WSI is computed using equal-weight averaging of normalized features:

$$\text{WSI}_t = \frac{1}{p} \sum_{i=1}^p \frac{x_{i,t} - \text{median}(x_i)}{\text{IQR}(x_i)} \quad (3.2)$$

where p is the number of selected features. Since features are already normalized (robust scaling), this simplifies to $\text{WSI}_t = \frac{1}{p} \sum_{i=1}^p x_{i,t}^{norm}$. Equal-weight averaging is the baseline approach, providing interpretability and avoiding overfitting to specific features.

Rolling Window Statistics

The 6-hour window is justified by synoptic-scale meteorology where weather systems operate on characteristic timescales of 6–12 hours (4). This temporal resolution captures synoptic-scale changes while maintaining sensitivity to mesoscale phenomena. For each time point t , a centered window $[t - 2, t - 1, t, t + 1, t + 2, t + 3]$ is used:

$$\text{WSI}_{\text{window},t} = \frac{1}{6} \sum_{i \in W_t} \text{WSI}_i \quad (3.3)$$

$$\text{WSI}_{\text{std},t} = \sqrt{\frac{1}{5} \sum_{i \in W_t} (\text{WSI}_i - \text{WSI}_{\text{window},t})^2} \quad (3.4)$$

$$\text{WSI}_{\text{trend},t} = \text{slope}(\text{linear_regression}(\text{WSI} \sim \text{time})) \text{ over 6h window} \quad (3.5)$$

where W_t represents the 6-hour window centered at time t .

Temporal Smoothing

Temporal smoothing is applied using a median filter with kernel size 3:

$$\text{WSI}_{\text{smoothed},t} = \text{median}(\text{WSI}_{\text{window},t-1}, \text{WSI}_{\text{window},t}, \text{WSI}_{\text{window},t+1}) \quad (3.6)$$

Median filtering is robust to outliers and prevents isolated misclassifications. The kernel size of 3 (3-hour smoothing) provides additional noise reduction without excessive lag, preserving sharp transitions while removing spurious fluctuations (?).

3.4.3 Classification Methods

Gaussian Mixture Model Classification

Gaussian Mixture Model (GMM) classification is the primary method for identifying stable and unstable weather regimes. GMM provides probabilistic classification with uncertainty quantification and handles non-spherical clusters better than k-means (8). The distribution of WSI values is modeled as:

$$P(\text{WSI}) = \sum_{k=1}^K \pi_k \mathcal{N}(\text{WSI} | \mu_k, \sigma_k^2) \quad (3.7)$$

where K is the number of regimes (determined by Bayesian Information Criterion, BIC), π_k are mixing proportions, and μ_k, σ_k^2 are regime-specific means and variances.

Model selection uses BIC to prevent overfitting:

$$\text{BIC} = -2 \ln(L) + k \ln(n) \quad (3.8)$$

where L is the likelihood, k is the number of parameters, and n is the sample size. The model with the lowest BIC is selected from candidate models with 1–3 components.

Soft classification probabilities are computed as:

$$P(\text{regime}_k | \text{WSI}_t) = \frac{\pi_k \mathcal{N}(\text{WSI}_t | \mu_k, \sigma_k^2)}{\sum_{j=1}^K \pi_j \mathcal{N}(\text{WSI}_t | \mu_j, \sigma_j^2)} \quad (3.9)$$

Hard classification assigns the most likely regime:

$$\text{regime}_t = \arg \max_k P(\text{regime}_k | \text{WSI}_t) \quad (3.10)$$

Binary classification uses a threshold $\tau = 0.5$:

$$\text{unstable}_t = \begin{cases} 1 & \text{if } P(\text{unstable} | \text{WSI}_t) > \tau \\ 0 & \text{otherwise} \end{cases} \quad (3.11)$$

The unstable regime is identified as the cluster with higher mean WSI value.

Alternative Classification Methods

For comparison and validation, two alternative classification methods are implemented:

- **K-Means clustering (k=2):** Applied directly to normalized features. Provides baseline comparison but assumes spherical clusters and provides no uncertainty quantification.
- **Percentile threshold:** Classifies as unstable if $\text{WSI} \geq 75\text{th percentile}$. Simple, interpretable method for comparison.

3.4.4 Validation Criteria

Multiple validation metrics ensure classification quality:

- **Silhouette score:** $s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$ where a_i is the average distance to points in the same

cluster, b_i is the average distance to points in the nearest other cluster. Acceptable threshold: $s > 0.4$ (?).

- **Davies-Bouldin index:** $DB = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \frac{\sigma_i + \sigma_j}{d(c_i, c_j)}$ where σ_i is cluster dispersion and $d(c_i, c_j)$ is cluster distance. Lower values indicate better clustering.
- **Calinski-Harabasz score:** $CH = \frac{SSB/(K-1)}{SSW/(N-K)}$ where SSB is between-cluster sum of squares and SSW is within-cluster sum of squares. Higher values indicate better clustering.
- **Cluster balance:** Target 40/60 to 60/40 split to avoid extreme imbalances (90/10) which indicate poor classification or insufficient discriminative power.

3.5 Renewable Energy Prediction Models

This section describes the implementation of both statistical and deep learning prediction models for renewable energy forecasting. The model selection rationale is based on established methods in the literature (1; 2; 15; 18; 19), encompassing statistical models (persistence models, Seasonal Persistence, ARIMA/SARIMA, Prophet, and Exponential Smoothing) and deep learning models (LSTM, CNN-LSTM, TCN, and ensemble approaches).

Feature selection for solar PV prediction includes radiation, cloudiness, temperature, and temporal features, while wind power prediction utilizes wind speed, wind direction, pressure, temperature, and temporal features. For deep learning models, additional feature engineering may include lagged values, rolling statistics, and encoded temporal patterns. The walk-forward validation strategy employs rolling training windows with appropriate forecast horizons and update frequencies, mimicking operational forecasting conditions. All models are implemented and trained using the collected data, ensuring consistent evaluation across model types.

3.6 Performance Evaluation

Performance evaluation methods include metrics calculation (MAE, RMSE, MAPE, Bias, Forecast Skill, error percentiles) following established practices in renewable energy forecasting (1; 2). Stratified performance analysis by weather stability regime (stable vs unstable) enables systematic comparison across different weather conditions. Statistical tests for comparing performance include the Mann-Whitney U test (10), Welch's t-test, and linear mixed-effects models. Effect size calculations (11) (Cohen's d, percentage increase, rank-biserial correlation) quantify the practical significance of performance differences. Multiple testing correction procedures (12) (Bonferroni, False Discovery Rate) control the false discovery rate, and bootstrap confidence intervals provide uncertainty quantification for performance metrics.

3.7 Robustness Analysis

3.8 Validation Approach

3.9 Ethical Considerations

This research uses publicly available weather data from the German Meteorological Service (DWD) under open data license and renewable energy production data from publicly accessible sources. All data collection and processing procedures comply with data usage terms and conditions specified by data providers. No personal or sensitive information is involved in this research. The study focuses on aggregate meteorological and energy production data that cannot be traced to individuals.

All code and methodology are documented to ensure reproducibility, and results are presented transparently with appropriate statistical reporting. The research aims to contribute to public knowledge about renewable energy forecasting without any commercial or political motivations that could bias the analysis or conclusions.

4 Data and Experimental Setup

4.1 Data Description

4.2 Data Sources

4.3 Experimental Setup

4.4 Software and Tools

5 Results

5.1 Introduction

5.2 Weather Stability Results (2024)

This section summarizes the empirical characteristics of the Weather Stability Index (WSI) and stability classifications for 2024.

5.2.1 WSI Characteristics and Regimes

The smoothed WSI spans from -0.366 to 0.663 with mean 0.051 and standard deviation 0.176 . Gaussian Mixture Model (GMM) model selection by BIC chose three components. Cluster means were approximately -0.125 , 0.078 , and 0.304 ; the highest-mean component was designated *unstable*. Cluster proportions were 40.4% , 21.9% (unstable), and 37.7% .

Cluster quality metrics indicate distinct regimes: Silhouette score 0.549 , Davies–Bouldin index 0.541 , and Calinski–Harabasz score 22375.45 . Alternative methods showed high agreement with the GMM baseline (Kappa: GMM vs Percentile 0.913 , GMM vs K-Means 0.551).

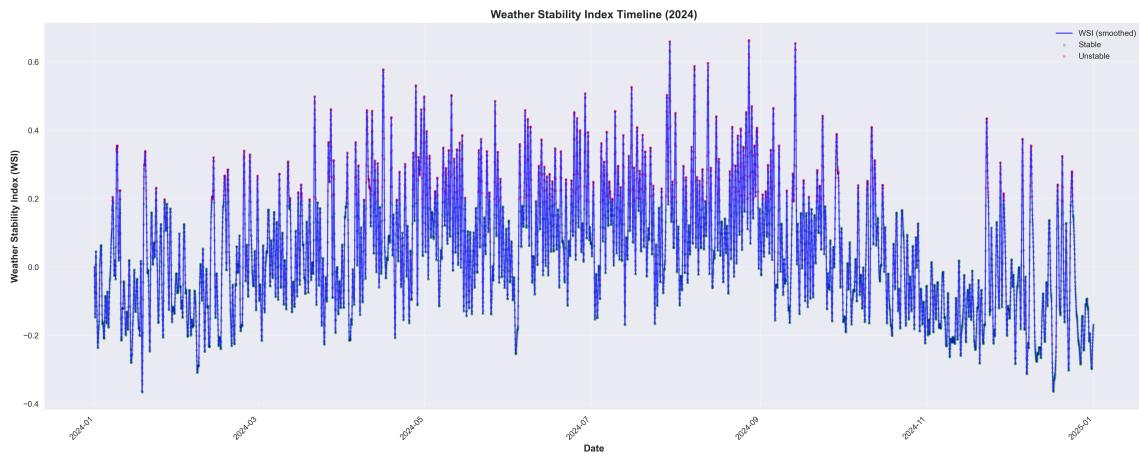


Figure 5.1: Weather Stability Index timeline (2024) with stable (green) and unstable (red) periods.

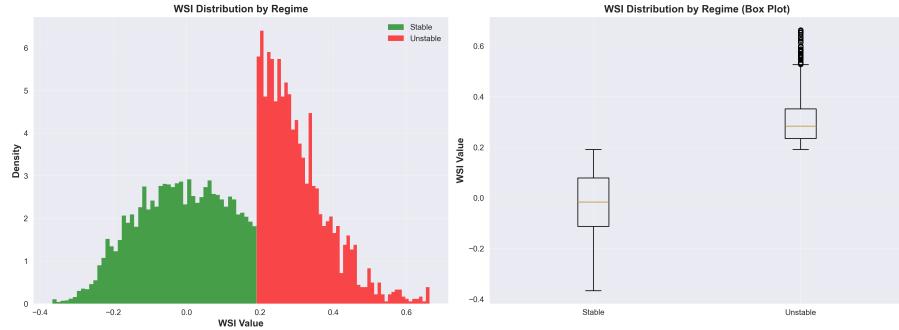


Figure 5.2: WSI distribution by regime (histogram and box plot).

5.2.2 Monthly Stability Patterns

Monthly stability patterns show seasonal variation in the fraction of unstable hours. Figure 5.3 (left) shows stacked stable vs unstable hours; Figure 5.3 (right) shows percentage unstable per month.

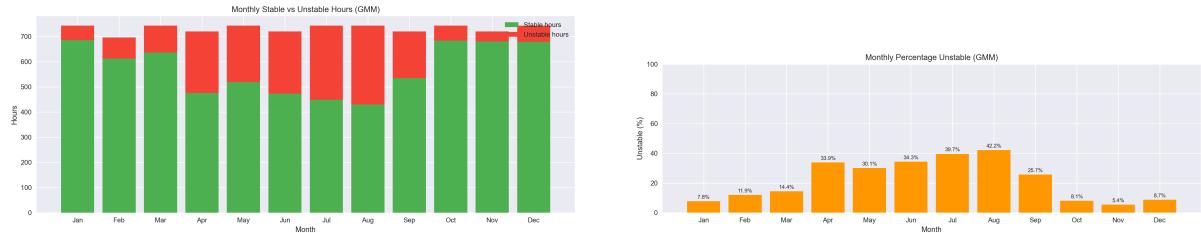


Figure 5.3: Monthly stability: hours (left) and percentage unstable (right).

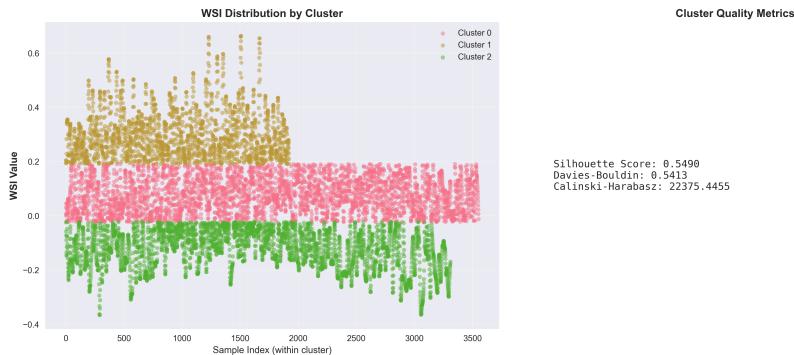


Figure 5.4: Cluster validation summary with WSI distribution per cluster and quality metrics (Silhouette, Davies-Bouldin, Calinski-Harabasz).

5.3 Main Findings

The WSI-based classification yields meteorologically coherent regimes with strong separation by cluster metrics. Across 2024, 78.1% of hours were stable and 21.9% unstable. Monthly plots indicate clear seasonal modulation of instability, which will be used to stratify forecasting errors in subsequent analyses.

5.4 Analysis

5.5 Summary

6 Discussion

6.1 Interpretation of Results

6.2 Comparison with Literature

6.3 Limitations

6.4 Implications

7 Conclusion

7.1 Summary of Contributions

7.2 Research Questions Revisited

7.3 Future Work

7.4 Final Remarks

A Appendix A

B Appendix B

Bibliography

- [1] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, “The state of the art in short-term prediction of wind power: A literature overview,” ANEMOS.plus, Tech. Rep., 2011, pp. 1–100.
- [2] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martínez-de-Pisón, and A. Sanz-García, “Review of photovoltaic power forecasting,” *Solar Energy*, vol. 136, pp. 78–111, 2016.
- [3] Y. Zhang, J. Wang, and X. Wang, “Short-term wind speed prediction based on spatial correlation and artificial neural networks,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 186, pp. 17–25, 2019.
- [4] J. R. Holton, *An Introduction to Dynamic Meteorology*, 4th ed. Elsevier Academic Press, 2004.
- [5] R. Huth, C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, and O. E. Tveito, “Classifications of atmospheric circulation patterns: recent advances and applications,” *Annals of the New York Academy of Sciences*, vol. 1146, no. 1, pp. 105–152, 2008.
- [6] P.-A. Michelangeli, R. Vautard, and B. Legras, “Weather regimes: Recurrence and quasi stationarity,” *Journal of the Atmospheric Sciences*, vol. 52, no. 8, pp. 1237–1256, 1995.
- [7] R. Vautard, “Multiple weather regimes over the North Atlantic: Analysis of precursors and successors,” *Monthly Weather Review*, vol. 118, no. 10, pp. 2056–2081, 1990.
- [8] G. J. McLachlan and D. Peel, *Finite mixture models*. New York, NY, USA: John Wiley & Sons, 2000.
- [9] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [10] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
- [11] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ, USA: Routledge, 1988.

- [12] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [13] P. J. Huber, *Robust Statistics*. John Wiley Sons, 1981.
- [14] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
- [15] A. Sedai, R. Dhakal, S. Gautam, A. Dhamala, A. Bilbao, Q. Wang, A. Wigington, and S. Pol, "Performance Analysis of Statistical, Machine Learning and Deep Learning Models in Long-Term Forecasting of Solar Power Production," *Forecasting*, vol. 5, no. 1, 2023.
- [16] T. Cabello-López, M. Carranza-García, J. Riquelme, and J. García-Gutierrez, "Forecasting solar energy production in Spain: A comparison of univariate and multivariate models at the national level," *Applied Energy*, vol. 341, p. 121645, 2023.
- [17] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," *Applied Computing and Informatics*, 2020.
- [18] J. Devaraj, R. Elavarasan, G. Shafiullah, T. Jamal, and I. Khan, "A holistic review on energy forecasting using big data and deep learning models," *International Journal of Energy Research*, vol. 45, no. 9, pp. 13489–13530, 2021.
- [19] G. Alkhayat and R. Mehmood, "A review and taxonomy of wind and solar energy forecasting methods based on deep learning," *Energy and AI*, vol. 5, p. 100060, 2021.
- [20] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019.
- [21] S. Haider, M. Sajid, H. Sajid, E. Uddin, and Y. Ayaz, "Deep learning and statistical methods for short- and long-term solar irradiance forecasting for Islamabad," *Renewable Energy*, 2022.
- [22] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, p. 120240, 2021.
- [23] M. Husein, E. Gago, B. Hasan, and M. Pegalajar, "Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies," *Heliyon*, vol. 10, 2024.
- [24] A. Mirza, Z. Shu, M. Usman, M. Mansoor, and Q. Ling, "Quantile-transformed multi-attention residual framework (QT-MARF) for medium-term PV and wind power prediction," *Renewable Energy*, 2023.
- [25] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Applied Energy*, vol. 259, p. 114216, 2020.
- [26] I. Jamil, H. Lucheng, S. Iqbal, M. Aurangzalb, R. Jamil, H. Kotb, A. Alkuhayli, and K. AboRas, "Predictive evaluation of solar energy variables for a large-scale solar power plant based on triple deep learning forecast models," *Alexandria Engineering Journal*, 2023.

- [27] D. Venkateswaran and Y. Cho, "Efficient solar power generation forecasting for greenhouses: A hybrid deep learning approach," *Alexandria Engineering Journal*, 2024.
- [28] X. Wang, Y. Sun, D. Luo, and J. Peng, "Comparative study of machine learning approaches for predicting short-term photovoltaic power output based on weather type classification," *Energy*, vol. 231, p. 122733, 2021.
- [29] I. Cervantes, C. Cervantes-Ortiz, D. Vázquez-Santana, and A. Arguelles, "Heuristic-machine learning models for solar radiation forecasting in Köppen climate zones," *Applied Soft Computing*, vol. 171, p. 112807, 2025.
- [30] F. Serras, K. Vandelanotte, R. Borgers, B. Van Schaeybroeck, P. Termonia, M. Demuzere, and N. Van Lipzig, "Optimizing climate model selection in regional studies using an adaptive weather type based framework: a case study for extreme heat in Belgium," *Climate Dynamics*, 2024.
- [31] S. Jain, "A novel seasonal segmentation approach for day-ahead load forecasting," *Energy*, vol. 258, p. 124752, 2022.
- [32] K. Blazakis, N. Schetakis, P. Bonfini, K. Stavrakakis, E. Karapidakis, and Y. Katsigiannis, "Towards Automated Model Selection for Wind Speed and Solar Irradiance Forecasting," *Sensors*, vol. 24, no. 15, p. 5035, 2024.
- [33] S. Lim, J. Huh, S. Hong, C. Park, and J. Kim, "Solar Power Forecasting Using CNN-LSTM Hybrid Model," *Energies*, vol. 15, no. 21, p. 8233, 2022.
- [34] A. Unlu and M. Peña, "Comparative Analysis of Hybrid Deep Learning Models for Electricity Load Forecasting During Extreme Weather," *Energies*, vol. 18, no. 12, p. 3068, 2025.
- [35] E. Sarmas, E. Spiliotis, E. Stamatopoulos, V. Marinakis, and H. Doukas, "Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent Long Short-Term Memory models," *Renewable Energy*, 2023.
- [36] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," *Journal of Cleaner Production*, vol. 318, p. 128566, 2021.
- [37] R. Ahmed, V. Sreeram, Y. Mishra, and M. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable & Sustainable Energy Reviews*, vol. 124, p. 109792, 2020.
- [38] R. Rajagukguk, R. Ramadhan, and H. Lee, "A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power," *Energies*, vol. 13, no. 24, p. 6623, 2020.
- [39] A. Assaf, H. Haron, H. Hamed, F. Ghaleb, S. Qasem, and A. Albarak, "A Review on Neural Network Based Models for Short Term Solar Irradiance Forecasting," *Applied Sciences*, vol. 13, no. 14, p. 8332, 2023.
- [40] A. Gupta and R. Singh, "A review of the state of the art in solar photovoltaic output power forecasting using data-driven models," *Electrical Engineering*, 2024.

- [41] Q. Zhang, B. Lei, J. Zhou, and Y. Liu, "A Review of Photovoltaic Power Generation Forecasting Techniques and Deep Learning Models," in *2024 The 9th International Conference on Power and Renewable Energy (ICPRE)*, 2024, pp. 1342–1347.
- [42] P. Gupta and R. Singh, "PV power forecasting based on data-driven models: a review," *International Journal of Sustainable Engineering*, vol. 14, no. 6, pp. 1733–1755, 2021.
- [43] D. Pombo, P. Bacher, C. Ziras, H. Bindner, S. Spataru, and P. Sorensen, "Benchmarking physics-informed machine learning-based short term PV-power forecasting tools," *Energy Reports*, 2022.
- [44] K. Ferkous, M. Guermoui, S. Menakh, A. Bellaour, and T. Boulmaz, "A novel learning approach for short-term photovoltaic power forecasting - A review and case studies," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108502, 2024.
- [45] M. Lipu, M. Miah, M. Hannan, A. Hussain, M. Sarker, A. Ayob, M. Saad, and M. Mahmud, "Artificial Intelligence Based Hybrid Forecasting Approaches for Wind Power Generation: Progress, Challenges and Prospects," *IEEE Access*, vol. 9, pp. 102460–102489, 2021.
- [46] Y. Dou, S. Tan, and D. Xie, "Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review," *Frontiers in Energy Research*, 2023.
- [47] P. Hewage, M. Trovati, E. Pereira, and A. Behera, "Deep learning-based effective fine-grained weather forecasting model," *Pattern Analysis and Applications*, vol. 24, no. 1, pp. 343–366, 2020.
- [48] C. Hachimi, S. Belaqziz, S. Khabba, B. Hssaine, M. Kharrou, and A. Chehbouni, "Advancements in weather forecasting for precision agriculture: From statistical modeling to transformer-based architectures," *Stochastic Environmental Research and Risk Assessment*, 2024.
- [49] L. Xu, N. Chen, X. Zhang, and Z. Chen, "A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale," *Climate Dynamics*, vol. 54, no. 7, pp. 3355–3374, 2020.
- [50] M. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. Leufen, A. Mozaffari, and S. Stadtler, "Can deep learning beat numerical weather prediction?" *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, 2021.
- [51] C. Vennila, A. Titus, T. Sudha, U. Sreenivasulu, N. Pandu, R. Reddy, K. Jamal, D. Lakshmaiah, P. Jagadeesh, and A. Belay, "Forecasting Solar Energy Production Using Machine Learning," *International Journal of Photoenergy*, 2022.
- [52] M. Shahhosseini, G. Hu, and S. Archontoulis, "Forecasting Corn Yield With Machine Learning Ensembles," *Frontiers in Plant Science*, vol. 11, p. 1120, 2020.
- [53] S. Fatima and A. Rahimi, "A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems," *Machines*, vol. 12, no. 6, p. 380, 2024.

- [54] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and Machine Learning forecasting methods: Concerns and ways forward,” *PLoS ONE*, vol. 13, no. 3, p. e0194889, 2018.
- [55] K. Szostek, D. Mazur, G. Drałus, and J. Kusznier, “Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production,” *Energies*, vol. 17, no. 19, p. 4803, 2024.