NOVA IMS
INFORMATION MANAGEMENT SCHOOL

Master's Degree Program in
Data-Driven Marketing

Specialization in
Digital Marketing & Analytics

Forecasting Hybrid Power Plant Generation.

The case of Joint and Isolated Solar and Wind Forecasting in Hybrid Power Plants in the Iberian Peninsula.

Leonor Mergulhão Alves Baptista de Almeida

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade NOVA de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

FORECASTING HYBRID POWER PLANT GENERATION

The case of Joint and Isolated Solar and Wind Forecasting in Hybrid Power Plants in the Iberian Peninsula.

Research Question

To what extent does joint forecasting of solar and wind generation in hybrid power plants at the point of interconnection reduce predictive uncertainty compared to isolated forecasting in the Iberian Peninsula?

by
Leonor Mergulhão Alves Baptista de Almeida

Master Thesis presented as partial requirement for obtaining the Master's degree in Data-Driven Marketing, with a specialization in Digital Marketing & Analytics

Supervised by

Professor Flávio Luís Portas Pinheiro, NOVA Information Management School

February, 2026

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

[Lisbon, 2026]

Leonor Mergulhão Alves Baptista de Almeida

DEDICATION

This dissertation is dedicated to EDP.

To the organization whose values of responsibility, innovation and excellent are shaping not only my professional path, but also my perspective on the impact that energy and technology can have on society.

The energy sector, defined the direction I wish to pursue and the purpose around I aspire to build my career, and to the remarkable people I have met along the way, whose trust, guidance and example profoundly influenced my growth and played a meaningful role in the achievement of this milestone.

FORWARD

“This work addresses, with methodological rigor and clear practical relevance, one of the central challenges of power systems with high renewable penetration: managing the uncertainty associated with generation forecasting. By adopting a system-level perspective and focusing the analysis on the point of interconnection of hybrid solar-wind plants in the Iberian Peninsula, the methodology goes beyond the traditional technology-specific forecasting approach, proposing an innovative joint forecasting framework. I followed the development of this work with particular interest, especially in a context where the hybridization of generation assets is gaining increasing relevance as a menas of maximizing the utilization of existing infrastructure, both at the transmission and distributions network levels.”

Berto Martins, Director and Engineer of Global Energy Management, EDP

“This thesis presents a well-structured analysis of joint solar-wind forecasting in hybrid power plants understanding actual behavior in renewable markets.
Leonor demonstrates a high knowledge of technologies and, based on the real operation data, a significant reduction on aggregated uncertainty through joint forecasting, with clear practical implications.
This is a high-quality analysis developed with analytical rigor.”

Jose Angel Rodriguez, Engineer at EDP

“This approach, based on hybridization concepts which, in essence, aim to utilize the unused capacity of the grid connection point allocated to a given installation by combining renewable energy conversion technologies, sought to focus the analysis on forecasting error in relation to the actual observed generation.Indeed, electricity generation facilities operating in a market context, without installed energy storage capacity, must concentrate their efforts on reducing imbalances associated with the variability of renewable energy resources (excluding hydroelectric plants with storage capacity).Accordingly, both the combination of conversion technologies selected to form the hybrid system and the forecasting tool itself are fundamental to mitigating deviations (whether positive or negative), as well as the resulting penalties, which invariability have a negative impact on business performance. At a time when producers are transitioning from feed-in tariff schemes to market-based remuneration, and facing increased competition driven by the strong and rapid growth of solar energy in the Iberian Peninsula, the topic addressed in the work proves to be highly timely.
Congratulations on your work.”

Adelino Barbosa, Engineer at EDP

ACKNOWLEDGEMENTS

This dissertations represents an important milestone in my academic journey at NOVA Information Management School. It concludes two challenging and transformative years marked by growth, discipline, and resilience. Completing this work is an achievement and a reflection those who supported me throughout this path. I am grateful to NOVA IMS for the foundation and environment that made this journey possible.

I would like to express my sincere appreciation to my supervisor, Professor Flávio Luís Portas Pinheiro, for his guidance and constructive feedback which were essential to the development of this research.

To my team at EDP, for the trust and flexibility in balancing professional and academic responsibilities. To my manager, Abdul Karim, for his mentorship and belief in my potential.

To Berto Martins, Director Engineer, for his involvement, supervision during this process and for investing in my professional growth. To Jose Angel Rodriguez, Daniel Cabezon and Adelino Barbosa, Engineers, for providing the resources that strengthened this work.

To my mother, for installing in me the determination and permanent intellectual curiosity, to my father, for fostering my analytical thinking, and to both, for their unconditional support and constant encouragement in yet another significant chapter of my life.

To my grandparents and my great-aunt, who are no longer here but remain deeply present in my life. Their values continue to guide me, and I know they would be proud.

To by boyfriend, for being my safe harbour during this stage and for sharing yet another chapter of our lives together.

To my family and friends, for standing by me and always supporting my decisions and ambitions.

To my closest friends, for their patience, their belief in me and for being a steady source of strength during every stage of this process. You truly made the difference.

A big thank you to everyone who has been part of this rewarding journey.

ABSTRACT

The rapid expansion of wind and solar generation in the Iberian Peninsula has reshaped power systems operations, elevating short-term forecasting to a core risk management tool. In hybrid solar–wind plants connected at a single interconnection point, uncertainty arises not only from individual resource variability, but from technological interaction under shared grid constraints. Yet solar and wind generation are typically forecasted independently and aggregated only a posteriori. This dissertation examines whether jointly forecasting reduces predictive uncertainty relative to isolated approaches. Using real operational data from hybrid plants operated by EDP in Portugal and Spain, the study adopts a system-level perspective centered on net export at the interconnection point. A comparative framework evaluates isolated and joint formulations under identical data, training, and evaluation conditions. Statistical baselines and structured deep learning architectures are tested with emphasis on NHITS due to its multi-scale temporal representation. Results show that joint forecasting reduces Mean Absolute Error (MAE) by approximately 20.54% compared to isolated formulations. This improvement does not increase the intrinsic predictability of solar or wind generation but emerges from partial error compensation across technologies with distinct temporal dynamics. By mitigation extreme deviations, the joint formulation reduces exposure to tail events that drive operational and market risk in renewable dominated systems. These findings demonstrate that forecasting performance in hybrid plants depends not only on model sophistication but on how the forecasting problem is structured and evaluated, positioning hybrid forecasting as a system-level uncertainty management challenge under realistic industrial constraints.

KEYWORDS

Hybrid Renewable Power Plants; Solar-Wind Complementarity; Joint Forecasting; Net Export Forecasting; NHITS; Renewable Energy Uncertainty; ReCycle;

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ARIMA	Autoregressive Integrated Moving Average
BESS	Battery Energy Storage System
CNN	Convolutional Neural Network
CRPS	Continuous Ranked Probability Score
CSV	Comma-Separated Values
EDP	Energias de Portugal
ES	Exponential Smoothing
GBM	Gradient Boosting Machine
GBR	Gradient Boosting Regressor
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MIBEL	Iberian Electricity Market
ML	Machine Learning
NHITS	Neural Hierarchical Interpolation for Time Series
NWP	Numerical Weather Prediction
OMIE	Operador do Mercado Ibérico de Energia
PV	Photovoltaic
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
UTC	Coordinated Universal Time
VAR	Vector Autoregressive

1. INTRODUCTION

In organized electricity markets, such as the Iberian Electricity Market (MIBEL), deviations between forecasted generation and actual generation have direct economic and operational consequences. These deviations result in imbalance costs, production curtailment decisions and greater exposure to operational risk by market participants (Frew, et al., 2021; OMIE, 2024; Pinson, 2013). As wind and solar photovoltaic generation gains greater relevance in the electricity mix, the investigation of projections ceases to be merely a technical concern to become an essential practical requirement in managing short-term uncertainty.

This requirement is particularly critical in Portugal and Spain, given the strong predominance of wind and solar technologies, resulting from the sustained growth of renewable capacity combined with beneficial climatic conditions (Jornal de Negócios, 2023). Although this growth is fundamental for decarbonization objectives, it simultaneously increases the system's sensitivity to meteorological patterns, further reinforcing the importance of short-term production forecasting in operational decision-making (Xie et al., 2023). Thus, this forecasting capacity becomes a central pillar in the operation of electrical systems with high penetration of renewable sources, requiring the development of new management models that accompany this evolution.

Among the alternatives that seek to respond to this growing operational complexity, hybrid solar-wind plants, a model in which EDP was a pioneer in the Iberian Peninsula, stand out as a strategic solution. By co-locating complementary technologies, these installations respond pragmatically to grid constraints, land availability and licensing limitations, allowing more efficient use of connection capacity.
However, this physical integration model exposes a critical methodological gap: solar and wind generation, although operationally coupled through shared constraints and joint injection limits, continue to be forecasted in a segregated manner, with independent models whose joint behavior is only aggregated a posteriori. This disconnect raises a fundamental operational and methodological question: To what extent does joint forecasting of solar and wind generation in hybrid power plants at the point of interconnection reduce predictive uncertainty compared to isolated forecasting in the Iberian Peninsula?

This study on short-term forecasting is based on real production data and focuses on a set of large-scale hybrid solar-wind plants operated by EDP in the Iberian Peninsula. Although these plants operate under different climatic and resource conditions, they encompass a common operational structure and are proven from a system-level perspective, centered on net export at the interconnection point. The analysis empirically demonstrates that joint forecasting of aggregated power in these hybrid plants can significantly reduce predictive uncertainty when compared to the conventional practice of forecasting each technology in isolation.

Complementarily, this investigation seeks to identify which forecasting models best exploit this advantage under realistic operational conditions. To this end, several representative approaches are evaluated, from classical and standard statistical models to widely used recurrent neural networks and more recent deep learning architectures, such as NHITS (Neural Hierarchical Interpolation for Time Series Forecasting), frequently referenced in forecasting literature (Benidis et al., 2022; Hyndman & Athanasopoulos, 2021; Kong et al., 2019). The set of models tested demonstrates that the NHITS model presents the most reliable and robust performance across different horizons and technologies, under real operational conditions.

The results indicate that the main value of the joint forecasting formulation does not arise from greater individual predictability of solar or wind generation, but from how forecast errors interact and partially compensate at the system level, leading to an observed reduction of approximately 20% in the mean absolute error of net export. This result reinforces recent conclusions that forecasting performance in systems dominated by renewable sources depends as much on problem formulation and temporal structure as on the complexity of the model itself (Cerqueira, et al., 2020; Maragkos & Refanidis, 2025).

The remaining document is structured as follows: Chapter 2 reviews the relevant literature on renewable energy forecasting, hybrid plants and uncertainty characterization, highlighting the limitations of technology-specific approaches in operational contexts; Chapter 3 presents the Iberian case study of hybrid plants, describing data sources, preprocessing steps, forecasting formulations and the evaluation framework adopted in the analysis; Chapter 4 presents and discusses the empirical results, combining a descriptive analysis of solar and wind generation behavior with an assessment, at the modeling level, of interference performance and uncertainty reconfiguration at the interconnection point; Finally, Chapter 5 summarizes the main conclusions, discusses the study's limitations, and points to directions for future investigation in forecasting of hybrid systems and uncertainty-sensitive operational analysis.

2. RELATED WORK

Renewable energy forecasting has received considerable research attention over the past decades, as the volume of wind and solar power has increased in power systems worldwide (Hong et al., 2020). Early researched framed forecasting primarily as a resource prediction problem, focusing on estimating wind speed of solar irradiance and then converting these estimates into power output. In the traditional view, forecast performance was largely assessed using average error metrics and a substantial effort was devoted to improving point forecasts for each technology individually (Pinson, 2006; Pinson, 2013; Soman et al., 2010).

As renewable penetration deepened, the limitations of this formulations became increasingly evident, particularly when forecasts were used for operational and market decision-making in systems that have strong temporal dependence and a mixture of sources of variability (Lund, et al., 2017). This structural shift has progressively redirected in the literature from accuracy driven model comparisons to a more thorough investigation of temporal structure, long, horizon behavior and robustness under non-stationary conditions, particular in the context of utility-scale hybrid power plants (Cerqueira, et al., 2020; Das et al., 2025).

The growing hybridization of co-located photovoltaic and wind assets further reinforces the need for system-level forecasting approaches, as aggregate production profiles and shared grid constraints reshape the structure and aggregation of uncertainty (Chinaris et al., 2024; Hassan et al., 2023).

2.1 TEMPORAL PATTERNS IN RENEWABLE GENERATION

Renewable energy generation time series are not arbitrary or nor purely stochastic. Both solar and wind power exhibit pronounced temporal structure driven by physical processes that govern resource availability, including diurnal and seasonal cycles as well as regime-dependent behavior associated with large-scale atmospheric dynamics (Hong et al., 2020; Pinson, 2013; Yang et al., 2012). The existence of recurring temporal patterns provides the fundamental justification for forecasting renewable generation and motivates the widespread use of time-series methods that explicitly exploit seasonality, persistence and temporal dependence (Hyndman & Athanasopoulos, 2021).

Solar power generation is dominated by strong deterministic components linked to the Earth–Sun geometry, which induce highly regular daily and annual cycles in irradiance and PV output (Yang et al., 2012). Wind generation is typically more irregular at short time scales, but it still displays persistent temporal structure due to synoptic-scale weather systems and seasonal circulation patterns, leading to serial dependence and clustering of high- and low-production periods (Hong et al., 2020; Pinson, 2006). These properties imply that renewable generation is best understood as temporally structured variability rather than independent random fluctuations.

Renewable generation time series often exhibit non-stationary characteristics, including abrupt transitions between regimes, prolonged periods of anomalously high or low production and asymmetric variability across time scales (Hong et al., 2020; Pinson, 2013;). As a result, forecast errors may also present temporal dependence and clustering, which limits the adequacy of modelling and evaluation assumptions that implicitly treat errors as independent disturbances across time (Hong et al., 2020; Pinson, 2006). This temporal organization and regime dependence form the empirical basis for the forecasting approaches reviewed in the following sections, from early statistical baselines to modern learning-based and structured deep learning models (Benidis et al., 2022; Hyndman & Athanasopoulos, 2021)

2.2 EARLY STATISTICAL FORECASTING APPROACHES

The use of statistical methods for renewable energy forecasting long predates the recent expansion of machine learning based approaches. As early as the mid-2000s, forecasting problems in wind and solar energy were formulated as time-series tasks, motivated by the presence of recurring temporal patterns and the need for short-term operational predictability (Yang et al., 2007; Deshmukh & Deshmukh, 2008).

Early studies predominantly relied on simple statistical and empirical models, in which forecasting was treated as the extrapolation of historical behavior rather than as a complex learning problem. Reviews from this period document the widespread use of persistence-based approaches and basic time-series formulations, for short forecasting horizons where temporal dependence is strongest (Soman et al., 2010). These methods assumed that recent observations or values from analogous periods contain sufficient information to characterize near-future behavior.

Such statistical baselines were adopted not only for their simplicity, but also because they offered transparent and robust reference points for model evaluation. Persistence and seasonal repetition implicitly encoded the dominant temporal structure of renewable generation data, providing competitive performance under stable conditions (Dufo-López et al., 2009). As a result, these approaches became standard benchmarks against which subsequent methodological developments were assessed.

The reliance on fixed temporal patterns limited the adaptability of early statistical models under changing regimes or prolonged deviations from typical behavior. These limitations motivated the development of more structured statistical formulations capable of explicitly representing temporal dependence and seasonality, which are discussed in the following section.

2.3 STANDARD FORECASTING MODELS

In contrast to early empirical baselines, standard or classical forecasting models introduced explicit parametric representations of temporal dependence, grounded in statistical time-series theory. Following the early use of persistence-based and seasonal statistical baselines, the literature progressively adopted more structured time-series models to explicitly represent temporal dependence and seasonality in renewable generation data. These classical forecasting models constitute the first generation of formal predictive tools applied to wind and solar power forecasting beyond simple extrapolation.

Autoregressive and moving-average formulations, together with their seasonal extensions, became widely used due to their ability to model serial correlation and periodic patterns within a transparent and mathematically grounded framework. By explicitly parameterizing temporal dependence, these models enabled multi-step forecasting under well-defined assumptions, while maintaining interpretability and relatively low data requirements. As forecasting problems expanded to include multiple variables, locations, or technologies, multivariate linear models were introduced to capture interdependencies across time series, allowing joint modelling of correlated processes and shared temporal dynamics, particularly in hybrid wind-solar configurations (Huang et al., 2015). Empirical evaluations show that such standard or classical models often achieve competitive performance under stable conditions and moderate variability, particularly for short- to medium-term forecasting horizons (Bergmeir, & Benítez, 2012; Cerqueira, et al., 2020; Shih et al., 2019).

This forecasting models rely on assumptions of linearity, stationarity, and relatively stable temporal structure. Comparative studies indicate that their performance deteriorates under conditions characterized by strong non-stationarity, abrupt regime shifts, or highly asymmetric variability, which are common in renewable generation time series (Maragkos & Refanidis, 2025; Taieb et al., 2011). Despite these limitations, standard or classical models remain a central reference in the literature, providing interpretable benchmarks against which more complex forecasting approaches are evaluated. Their shortcomings, rather than their successes, motivated the transition toward learning-based models capable of representing non-linear temporal dependencies, discussed in the following section.

2.4 LEARNING-BASED MODELS AND THE BLACK-BOX TRADE-OFF

The availability of larger datasets and increased computational capacity led to the adoption of learning-based models in renewable energy forecasting, particularly neural network architectures capable of learning non-linear temporal dependencies directly from data. Neural networks have been widely explored in wind energy systems for modelling and forecasting applications (Marugán et al., 2018). Beyond single-series modelling, recurrent neural networks have also been extended to learn shared representations across groups of related time series, improving robustness (Bandara et al., 2020). Among this approaches, recurrent neural networks and Long Short-Term Memory (LSTM) models gained predominance due to their ability to represent delayed effects and complex temporal interactions (Kong et al., 2019).

The principal advantage of learning-based models lies in their high representational capacity. By relying on flexible function approximators, these models can capture complex patterns that are difficult to specify in classical statistical formulations, often leading to improvements in average forecasting accuracy under highly variable conditions (Benidis et al., 2022; Devaraj et al., 2021).

However this increased flexibility comes at the cost of interpretability. The internal mechanisms through which learning-based models generate forecasts are inherent typically to interpret , making it difficult to explain their behavior in terms of physical drivers or temporal structure. Unlike standard or classical models, their parameters cannot be directly associated with specific lags, cycles or regimes, which has led to their characterization as black-box models in the forecasting literature (Benidis et al., 2022).

In operational contexts, this lack of explanatory capacity limits the ability to assess forecast reliability, diagnose failures or anticipate model behavior under rare events or regime shifts. Benchmarking evidence further suggests that gains in average point accuracy do not necessarily translate into robust performance across horizons or operating conditions, emphasizing a fundamental trade-off between predictive power and interpretability (Gneiting, & Raftery, 2007; Hong et al., 2020; Pinson, 2013).

2.5 NHITS: STRUCTURED LONG-HORIZON FORECASTING

NHITS (Neural Hierarchical Interpolation for Time Series Forecasting) is a structured deep learning architecture designed to address long-horizon forecasting problems, where relevant temporal dynamics extend far beyond short-term dependencies. In contrast to many learning-based models that prioritize short-horizon accuracy, NHITS explicitly represents temporal structure across multiple time scales, allowing complex and long-range patterns to be modelled in a coherent manner (Challu, et al., 2022; Stefenon et al., 2025).

The core strength of NHITS lies in its hierarchical and multi-scale formulation (Challu, et al., 2022). By decomposing the forecasting task into specialized components operating at different temporal resolutions, the model avoids short-sighted behavior and reduces the tendency to extrapolate distant horizons solely from short-term patterns (Benidis et al., 2022; Taieb et al., 2011). This design enables NHITS to maintain temporal coherence and stability even when forecasting over extended horizons, potentially spanning several years, which is a known limitation of many conventional neural forecasting models (Hong et al., 2020; Wang, et al., 2021). Recent empirical applications of NHITS in energy systems further confirm its robustness in multi-horizon settings (Stefenon et al., 2025).

From a modelling perspective, NHITS expands explanatory capacity relative to simpler formulations. Rather than imposing a single global relationship between past and future values, the architecture allows different temporal components to be learned independently, capturing both fast variations and slow-moving dynamics within the same framework. This contrasts sharply with linear regression models, which assume fixed linear relationships and are inherently limited in their ability to represent multi-scale and non-stationary behavior. As such, NHITS occupies an intermediate position between restrictive linear models and unconstrained black-box neural networks, combining flexibility with explicit temporal structure (Benidis et al., 2022). The hierarchical organization of the model architecture proposed by (Challu, et al. 2022) is illustrated in Figure 1.

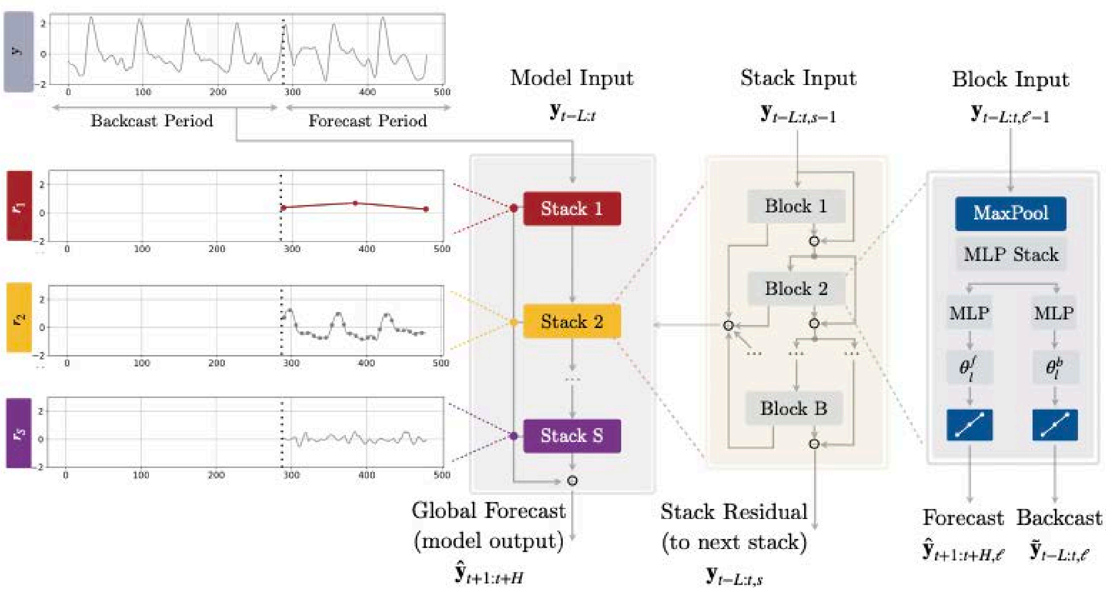

Figure 1 - NHITS Architecture (proposed by Challu, et al. 2022)

The NHITS architecture is organized into multiple hierarchical stacks operating at different temporal resolutions. Each stack receives the residual signal from the previous layer and produces both a Backcast component, which refines the representation of past inputs, and a forecast component, which contributes to the final prediction. Through this iterative residual refinement mechanism, the model decomposes the forecasting task across scales, progressively capturing coarse and fine-grained temporal dynamics before aggregating them into the final long-horizon forecast.

2.6 ReCycle: MODELLING DEVIATIONS FROM AVERAGE BEHAVIOR

ReCycle is a structured forecasting approach designed to improve long-horizon robustness by explicitly separating dominant cyclic behavior from residual dynamics before model learning (Weyrauch et al., 2024). Rather than modelling the full time series directly, ReCycle reformulates the forecasting problem by first extracting an average or representative cycle that captures the regular, recurring structure of the data.

This average behavior is obtained through a segmentation and compression process that identifies the dominant cyclic component of the time series. The resulting baseline represents typical behavior across comparable temporal segments, such as daily or seasonal cycles. Forecasting is then performed only on the residual signal, defined as the deviation from this average cycle. By construction, the learning model focuses on irregular dynamics, regime-dependent behavior and atypical deviations, rather than repeatedly relearning well-established cyclic patterns (Weyrauch et al., 2024).

When combined with structured forecasting architectures such as NHITS, this formulation further enhances long-horizon stability. NHITS is applied only to the residual component, allowing its hierarchical and multi-scale structure to model complex deviations without being dominated by strong deterministic cycles. This separation reduces model complexity, limits error accumulation and improves robustness under non-stationary conditions, where deviations from typical behavior are operationally most relevant.

From a modelling perspective, ReCycle does not introduce additional predictive power through increased model flexibility, but through problem reformulation. By isolating average behavior from residual dynamics, it reshapes the forecasting task in a way that aligns model capacity with the structure of uncertainty in the data. This makes ReCycle particularly suitable for forecasting applications characterized by strong cyclic patterns coexisting with episodic regime shifts, complementing the structured long-horizon capabilities of NHITS (Benidis et al., 2022; Weyrauch et al., 2024).

Figure 2 reproduces the ReCycle architecture proposed by Weyrauch et al. (2024). The model first compresses the time series according to its primary cycles, computes residual signals relative to historical profiles and subsequently learns these residuals using an encoder–decoder framework.

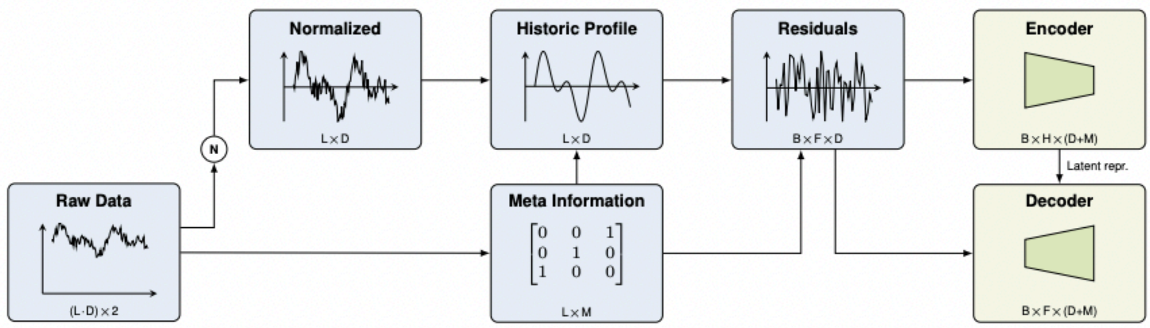

Figure 2 - ReCycle architecture (adapted from Weyrauch et al., 2024).

To contextualize the system-level results presented in the following chapters and to ensure that any observed gains are not driven by underperforming component models, the proposed forecasts are benchmarked against literature-based reference performance ranges. Table 1 provides a structured overview of the main forecasting model families and their methodological characteristics, while Table 2 and Table 3 summarize typical MAE ranges reported in the literature for solar and wind generation, respectively. These benchmarks are compiled from comparative studies reporting representative error magnitudes across different model classes and technologies. This positioning allows the empirical results to be interpreted within established performance intervals instead of in isolation, reinforcing the validity of subsequent system-level comparisons.

Table 1 - Forecasting Approaches in Renewable Energy

Forecasting approach	Representative models	Technology	Typical focus in the literature	Strengths	Limitations	Key references
Simple statistical baselines	Persistence, Seasonal Naïve	Solar, Wind	Short-term point forecasting	Transparent, robust benchmarks, strong short-horizon performance	No10daptability to regime changes; ignore interaction effects	Hyndman & Athanasopoulos (2021); Maragkos & Refanidis (2025)
Classical time-series models	ARIMA, SARIMA	Solar, Wind	Modelling temporal and seasonal structure	Interpretable; explicit seasonality modelling	Assume linearity and stationarity; limited under high variability	Hyndman & Athanasopoulos (2021)
Multivariate linear models	VAR	Solar, Wind	Joint modelling of multiple variables or technologies	Structured representation of linear interdependencies	Limited ability to capture non-linear interactions and shared constraints	Cerqueira, et al. (2020)
Recurrent neural networks	LSTM	Solar, Wind, Hybrid	Non-linear temporal dependency learning	Capture memory effects and delayed responses	Limited long-horizon coherence; training instability	Devaraj et al. (2021);Kong et al. (2019)
Structured deep learning models	NHITS, ReCycle	Solar, Wind, Hybrid	Multi-scale and long-horizon forecasting	Improved temporal coherence and robustness	Higher complexity; limited evidence in operational hybrid plants	Challu, et al. (2022); Weyrauch et al. (2024)
Hybrid system forecasting perspective	Joint / system-level formulations	Hybrid	Net export and system-level uncertainty	Explicitly address interaction and shared constraints	Still underexplored in short-term operational contexts	Couto & Estanqueiro (2021); Polo et al. (2025); Yang et al. (2012)

Table 2 - PV forecasting MAE reference ranges

MODEL	OUTPUT	MAE (0-1)	MAE (MW)	REFERENCES
ARIMA	PV (MW)	0.60 – 0.80	(8 - 25)	(Hyndman & Athanasopoulos, 2021; Soman et al., 2010; Yang et al., 2012)
SNaive	PV (MW)	0.75 – 0.90	(10 - 30)	(Hyndman, & Koehler, 2006; Yang et al., 2012)
SARIMA	PV (MW)	0.50 – 0.70	(6 - 20)	(Soman et al., 2010; Yang et al., 2012)
VAR	PV (MW)	0.45 – 0.65	(5 -18)	(Cerqueira, et al., 2020; Couto & Estanqueiro, 2021)
Prophet	PV (MW)	0.45 - 0.65	(5 -18)	(Devaraj et al., 2021; Wang, et al., 2021; Yang et al., 2012)
Random Forest	PV (MW)	0.40 - 0.60	(4 -15)	(Devaraj et al., 2021; Wang, et al., 2021; Xie et al., 2023)
XGBoost	PV (MW)	0.35 – 0.55	(3.5 - 14)	(Devaraj et al., 2021; Taieb & Hyndman, 2014; Xie et al., 2023)
LSTM	PV (MW)	0.30 – 0.50	(3 – 12)	(Devaraj et al., 2021; Silva-Rodriguez et al. 2024)
Wavelet + NN	PV (MW)	0.25 – 0.45	(2.5 – 10)	(Devaraj et al., 2021; Zhang et al., 2014)
AutoPV	PV (MW)	0.23 - 0.40	(2 - 8)	(Devaraj et al., 2021; Xie et al., 2023)
Transformer	PV (MW)	0.25 - 0.40	(2 - 9)	(Benidis et al., 2022)
NHITS	PV (MW)	0.18 - 0.30	(1.8-7)	(Benidis et al., 2022; Challu, et al., 2022; Stefenon et al., 2025)

Table 3 - Wind forecasting MAE reference ranges

MODEL	OUTPUT	MAE (0-1)	MAE (MW)	REFERENCES
ARIMA	WIND (MW)	0.60 – 0.80	(12 - 30)	(Hong et al., 2020; Pinson, 2013; Soman et al., 2010)
SNaive	WIND (MW)	0.70 – 0.90	(15 - 35)	(Hong et al., 2020; Hyndman, & Koehler, 2006; Soman et al., 2010)
SARIMA	WIND (MW)	0.50 – 0.70	(10 - 25)	(Hong et al., 2020; Soman et al., 2010; Yang et al., 2012)
VAR	WIND (MW)	0.45 – 0.65	(8 - 22)	(Cerqueira, et al., 2020; Couto & Estanqueiro, 2021)
Prophet	WIND (MW)	0.45 0.65	(8 – 22)	(Devaraj et al., 2021; Wang, et al., 2021)
Random Forest	WIND (MW)	0.40 - 0.60	(7 - 20)	(Devaraj et al., 2021; Wang, et al., 2021; Xie et al., 2023)
XGBoost	WIND (MW)	0.35 – 0.55	(6 - 18)	(Devaraj et al., 2021; Taieb & Hyndman, 2014; Wang, et al., 2021)
LSTM	WIND (MW)	0.30 - 0.50	(5 - 15)	(Kong et al., 2019; Silva-Rodriguez et al. 2024; Wang, et al., 2021)
Wavelet + NN	WIND (MW)	0.25 - 0.45	(4 - 13)	(Wang, et al., 2021; Zhang et al., 2014)
Transformer	WIND (MW)	0.25 - 0.40	(4 - 12)	(Benidis et al., 2022)
NHITS	WIND (MW)	0.18 - 0.30	(2.5 - 9)	(Benidis et al., 2022; Challu, et al., 2022; Stefenon et al., 2025)

3. DATA AND METHODS

3.1 DATA SOURCES AND AVAILABLE INFORMATION

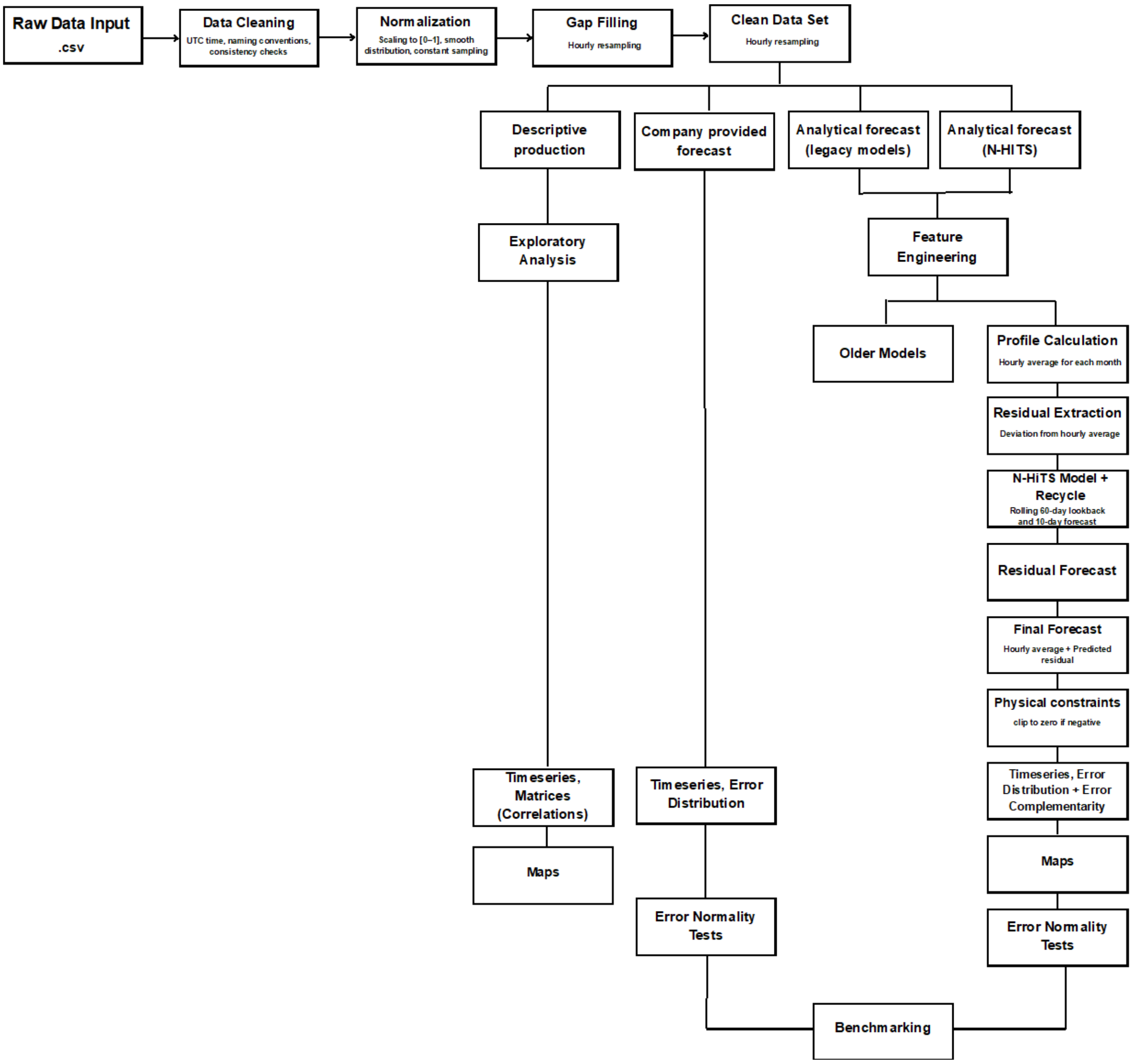

Figure 3 - Diagram representation with each step of the data processing detailed on the methodology section

The empirical analysis relies on hourly time series of solar and wind energy production and corresponding operational forecasts. Focusing on production data rather than meteorological variables reflects a deliberate modelling choice. Prior work shows that autoregressive formulations can capture relevant short-term temporal dependencies when the objective is operational predictability rather than physical resource estimation (Taieb & Hyndman, 2014; Hyndman & Athanasopoulos, 2021). In market-oriented forecasting applications, the variable of interest is not the underlying wind speed or solar irradiance, but the energy effectively injected into the grid, which directly conditions system operation and market participation outcomes (OMIE, 2024; Pinson, 2013).

Although meteorological variables are not included as model inputs in this work, the operational forecasts used in the analysis are derived from numerical weather prediction models. All forecast data were generated by an external specialised forecasting provider responsible for supplying renewable generation forecasts to EDP. These forecasts incorporate detailed meteorological information, including wind and solar resource modelling and are produced within an operational forecasting framework aligned with market and system operation requirements.

All datasets used in this study were provided by EDP and correspond to real operational conditions of renewable generation assets located in the Iberian Peninsula. The data include information for solar and wind power plants operating in both Portugal and Spain, spanning diverse climatic regimes, geographical contexts, and grid connection characteristics. The data are hourly, consistent with most short-term forecasting studies in power systems and aligned with the granularity required for day-ahead market participation and dispatch-related processes.

Across the Iberian Peninsula, EDP has progressively implemented a portfolio of hybrid renewable power plants connected at a single point of interconnection, reflecting a strategic commitment to the hybridization of renewable generation assets. As of the most recent developments, this portfolio comprises ten hybrid plants, including eight in solar-wind hybrid systems and two hybrid solar-hydro configurations. Of the eight solar-wind hybrid plants, five are located in Spain and three in Portugal (EDP, 2024). While full access to operational forecast production data is available for all Spanish hybrid plants, in Portugal complete datasets are available for a single hybrid site. This constraint explains directly the empirical design of the present study and motivates the selection of the Hybrid Plant 1, which combines Plant 1 Solar with Plant 1 Wind, as the core reference case.

Regarding forecast data, hourly forecasts are available for the following assets:

- six solar plants, five located in Spain and one in Portugal;
- six wind plants, similarly distributed across five locations in Spain and one in Portugal.

Regarding actual production, hourly observations are available for the following assets:

- six wind plants, five in Spain and one in Portugal;
- one solar plant, Plant 1 Solar located in Portugal.

The hybrid configuration selected as the core case study throughout the dissertation is the Hybrid Plant 1, that represents one of the earliest hybrid projects currently operating in the Iberian Peninsula. Beyond data availability constraints, this configuration provides a coherent system-level framework. The presence of aligned hourly forecast and observed production data enables direct cross-technology comparison and consistent evaluation of joint forecasting performance.

The common period between observed production data and forecast data used for model training and evaluation spans from 1 October 2023 to 1 November 2025. Restricting the analysis to this overlapping interval ensures temporal consistency across data sources and avoids biases arising from non-overlapping periods or structural breaks associated with data availability. All subsequent pre-processing, modelling, and evaluation steps are conducted exclusively within this common time window. Table 4 describes the variables available in the dataset, including their type, temporal resolution, units, and role in the forecasting process.

The dataset does not include detailed operational capacity information for the hybrid plants, such as availability levels, curtailment events, maintenance periods, or temporary reductions in effective capacity. Consequently, it is not possible to determine whether observed production corresponds to full installed capacity or reflects partial operational constraints. Forecast deviations may therefore capture, in part, unobserved operational effects in addition to meteorological variability or model inaccuracies. Although this does not affect the internal consistency of the comparative evaluation, as all models are assessed under identical informational conditions, it limits the ability to fully disentangle capacity-driven effects from intrinsic forecasting performance.

Table 4 - Description of available variables

VARIABLES	UNITS	DESCRIPTION
Plant Name	(String)	Identifier of the renewable energy plant or park from which the data were collected
Coordinates	(Latitude, longitude)	Geographical location of the plant, expressed in latitude and longitude coordinates
Energy Type	(Wind, Solar)	Technology associated with the plant, indicating whether the generation source is wind or solar
Time	Hours	Temporal resolution of the observations, expressed in hourly intervals
Original Energy Produced	kWh	Measured electrical energy output of the plant before any normalization or transformation
Start Date	day	Beginning date of the time interval associated with the observation or forecast
End Date	day	Ending date of the time interval associated with the observation or forecast
Month	month	Calendar month corresponding to each observation, used to capture seasonal patterns
Normalized Energy Produced	kWh	Energy production adjusted through normalization to ensure comparability across plants or periods
Average Normalized Energy Produced	kWh	Mean value of normalized energy production over a defined reference period
Normalized Energy Residual	kWh	Deviation of normalized energy production from its average value, capturing short-term variability
Forecasted Energy	kWh	Energy output predicted by the forecasting model for a given time horizon
Forecast Error (Delta)	kWh	Difference between forecasted energy and observed energy production
Mean Absolute Error (MAE)	kWh	Average of the absolute forecast errors, measuring typical deviation magnitude
Root Mean Squared Error (RMSE)	kWh	Square root of the mean of squared forecast errors, emphasizing larger deviations

3.2 DATA QUALITY, INTEGRITY AND TEMPORAL CONSISTENCY

Before model development, the dataset was reviewed to ensure consistency and avoid technical errors during training. The analysis uses hourly data covering the same period of all assets. All-time series share the same start and end date corresponding to the common analysis period defined in Section 3.1, ensuring a common and fully aligned time window. No resampling or aggregation was required, as the data were already provided at hourly resolution.

Variable names were standardised and kept consistent across all datasets to avoid implementation errors and ensure clarity during modelling. Basic validation checks were performed to identify missing timestamps, duplicated records or inconsistent entries. Missing values (NaN) were removed to prevent computational issues during model training. No interpolation or artificial data filling was applied. The data did not require major corrections or structural adjustments. Following these checks, the final datasets consisted of clean, aligned hourly time series ready for modelling.

3.3 DATA TRANSFORMATION AND FORECASTING PROBLEM FORMULATION

After validation and integrity checks, the processing pipeline proceeds from raw inputs provided in comma-separated values (.csv) files to transformations required for learning-based forecasting. These files contain time-stamped production and forecast series for solar, wind and hybrid power plants. All timestamps are standardised to Coordinated Universal Time (UTC), and variable names are harmonised across datasets to ensure structural consistency.

Following temporal alignment, the series are organized into a structured hourly format suitable for supervised forecasting to ensure coherence across plants and technologies. No external manual rescaling is applied to the production or forecast variables. Model training is conducted using the NeuralForecast (Nixtla, 2015) framework, which internally handles the scaling procedures required for stable optimisation of neural architectures. Scaling parameters are estimated exclusively from the training data and consistently applied during validation and testing, preventing information leakage across temporal splits. This approach preserves the physical interpretability of the original energy values (kWh) while ensuring numerical stability during optimisation. Periods of zero production are retained throughout the transformation process and treated as valid operational observations, maintaining the empirical distribution and temporal dynamics of renewable generation.

3.4 MODELS, TRAINING STRATEGY AND EVALUATION FRAMEWORK

After data preparation and transformation, the forecasting problem is formulated within a supervised learning framework designed to reflect realistic operational forecasting conditions. Temporal dependencies are encoded through sliding windows of lagged observations, allowing models to learn short- and medium-term temporal structure directly from historical production data (Hyndman & Athanasopoulos, 2021). In addition to lagged production values, simple calendar-based covariates, namely the month of the year, are included to provide seasonal context, while no external explanatory variables are introduced. This modelling choice ensures that all forecasting performance arises from temporal structure and problem formulation rather than from exogenous information.

Three forecasting formulations are considered:

1. isolated solar forecasting,
2. isolated wind forecasting,
3. joint solar–wind forecasting.

In the joint formulation, a single model simultaneously produces forecasts for both technologies using a shared temporal representation. This approach enables the

model to implicitly capture interaction effects between solar and wind generation through a multi-task learning structure, without explicitly enforcing aggregation constraints (Caruana, 1997; Ruder, 2017). Aggregated net export is not forecasted directly, avoiding artificial effects associated with summing correlated forecast errors.
Model selection follows the literature reviewed in Chapter 2 and is intentionally limited to a representative subset of forecasting paradigms. The study does not aim to implement all model classes exhaustively. Instead, a structured review of comparative forecasting studies was conducted to compile representative performance ranges reported for solar and wind generation. These literature-based performance intervals are used as external reference benchmarks, allowing the empirical results to be evaluated against established ranges rather than solely against the models implemented in this work.

NHITS is selected as the primary deep learning architecture due to its structured multi-scale formulation, which is aligned with the long horizon forecasting objectives of this study. Given the strong temporal dependence and multi-scale variability observed in hybrid renewable generation, NHITS provides an appropriate framework to assess whether structured temporal representations yield system-level performance gains under shared operational constraints. All forecasting tasks isolated solar, isolated wind, and joint solar–wind are implemented within a common modelling framework using the NeuralForecast Python library (Nixtla, 2015).

Model training follows a sliding-window strategy in which a fixed number of past hourly observations (historical input window) is used to predict future values over predefined forecasting horizons. In practice, the model receives a sequence of past production values as input and learns to map this historical context to the subsequent forecast horizon. This design mirrors real operational forecasting conditions and avoids temporal leakage. A strict chronological separation between training, validation, and test sets is enforced by partitioning the time series into contiguous temporal blocks. The earliest observations are used for model training, the subsequent block is reserved for validation and hyperparameter selection, and the most recent segment is held out as an independent test set. No future information is used during model estimation.

ReCycle is implemented as a deterministic preprocessing step. For each plant, an average production profile is computed as a function of month and hour of day, and residuals are defined as deviations from this profile. NHITS is trained exclusively on the residual series. The final forecast is obtained by adding the predicted residual component to the corresponding average cyclic profile.

Forecasts are generated using a rolling-origin evaluation framework, where each input window is constructed exclusively from past observations relative to each prediction timestamp. At each forecasting step, the model only has access to information available up to that moment, replicating real operational day-ahead conditions. No random shuffling or cross-sectional sampling is applied, , preserving temporal order and preventing information leakage (Bergmeir, & Benítez, 2012; Hyndman & Athanasopoulos, 2021).

Hyperparameter tuning is performed using validation MAE over a restricted set of architectural parameters. For NHITS, the tuned parameters include historical input window, number of stacks, number of blocks per stack, hidden layer size, and maximum training steps. The forecasting horizon is fixed at 240 hours (10 days). No exhaustive grid search is performed, parameters are adjusted iteratively to ensure stable convergence and comparability across isolated and joint forecasting formulations.

Forecasting performance is evaluated using Mean Absolute Error (MAE) Equation (1) and Root Mean Squared Error (RMSE) Equation (2), both computed in physical units (kWh), ensuring direct operational interpretability. The MAE is defined as the average of the absolute deviations between forecasts and observations, corresponding to one divided by NNN times the sum, over all time steps, of the absolute value MAE.

$$MAE \ = \frac{1}{n}\sum_{i=1}^{n}|y_i - \ (y)_i| \tag{1}$$

The RMSE is defined as the square root of the average of the squared forecast errors, corresponding to the square root of one divided by N times the sum, over all time steps. By squaring individual deviations before averaging, RMSE places greater weight on large errors and is therefore more sensitive to infrequent but severe forecast deviations. This property makes RMSE particularly relevant in operational and market-oriented contexts, where large forecast errors may lead to disproportionate imbalance costs or violations of system constraints (Hyndman, & Koehler, 2006; Pinson, 2013).

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

Reporting both MAE and RMSE provides complementary information. While MAE reports average forecasting performance, RMSE highlights the presence and impact of large deviations. Beyond these point metrics, the distribution of forecasting errors is also analyzed, since operational reliability depends not only on expected accuracy but also on error variability and the occurrence of extreme deviations, which are known to exhibit temporal structure in renewable generation forecasting problems (Pinson, 2013).

All comparisons between isolated and joint forecasting formulations are conducted under strictly identical conditions. Models are trained on the same temporal samples, evaluated over the same test periods, and assessed using the same metrics. Under these conditions, observed performance differences can be attributed to the formulation of the forecasting problem rather than to discrepancies in data handling, training procedures, or evaluation criteria (Hyndman & Athanasopoulos, 2021). All data processing, modelling and evaluation steps are implemented in Python using a reproducible, notebook-based workflow.

Both NHITS and ReCycle were implemented using a unified training framework to ensure comparability across isolated and joint forecasting tasks. Hyperparameters were selected based on validation performance under identical data splits and forecasting horizons. Training procedures, optimization criteria and data splits were kept identical across models to isolate the effect of forecasting formulation from architectural differences.

To ensure transparency and reproducibility, the full modelling pipeline, including data preprocessing, model implementation and evaluation procedures, is publicly available in a dedicated GitHub repository (Almeida, 2026). Due to confidentiality agreements with the data provider, raw operational data cannot be publicly shared. The repository includes the complete codebase and detailed documentation required to replicate the experimental framework under equivalent data conditions.

3.5 Statical analysis

In hybrid renewable power plants, system-level uncertainty depends not only on average forecast accuracy but also the dependence structure between forecast errors across technologies, which directly influences net export variability at the point of interconnection. The statistical analysis focuses on error dependence and distributional properties, complementing aggregate metrics such as MAE and RMSE.

Linear dependence between error series is quantified using the Pearson correlation coefficient, which provides a direct measure of the extent to which forecast deviations co-occur and affect net export aggregation.

$$r = \frac{\left(n\sum xy - (\sum x)(\sum y)\right)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}} \tag{3}$$

A positive correlation ($r > 0$) indicates that forecast errors tend to occur in the same direction. In a solar–wind hybrid plant, this corresponds to situations where both technologies are either overestimated or underestimated. For example, a regional weather system producing unexpectedly low irradiance and low wind speeds may generate simultaneous negative forecast errors, increasing total system imbalance.

A negative correlation ($r < 0$) suggests partial compensation effects between technologies. In hybrid systems, this may occur when wind production is underestimated while solar production is overestimated, or vice versa. Such anti-correlated deviations partially offset aggregate forecast errors, thereby reducing net export variability and enhancing system-level stability (Bett, & Thornton, 2016; Couto & Estanqueiro, 2021).

Correlation values close to zero indicate the absence of linear dependence between error series, implying that aggregated error variance is primarily driven by individual component uncertainty rather than interaction effects. All correlation estimates are computed on out-of-sample error series to ensure operational relevance and avoid in-sample bias (Cerqueira, et al., 2020; Bergmeir, & Benítez, 2012).

4. RESULTS AND DISCUSSION

This chapter presents and discusses the empirical results in direct continuity with the conceptual framework and research questions developed earlier. The analysis examines the empirical structure of the observed production and forecast series. This approach follows a central premise of the dissertation: in hybrid solar-wind power plants, uncertainty arises not only from modelling decisions, but from system-level interactions between technologies with distinct physical drivers, temporal dynamics, and spatial characteristics. The section characterises the system under study. Persistence, seasonality, heterogeneity, temporal structure, and cross-technology dependence are examined to establish the empirical context in which forecasting results and error patterns should be interpreted. This chapter has three parts: a descriptive analysis of solar and wind generation behaviour, an examination of forecast errors under real operational conditions and a subsequent, model-based assessment of forecasting performance. This distinction allows system behaviour to be characterised independently, before evaluating how modelling choices reshape uncertainty under identical data conditions.

4.1 DESCRIPTIVE ANALYSIS OF SOLAR AND WIND GENERATION

The analysis begins with solar generation, as photovoltaic production is typically the most structured component of hybrid systems and often sets expectations regarding predictability. A direct inspection of the raw production time series provides the initial empirical reference point.

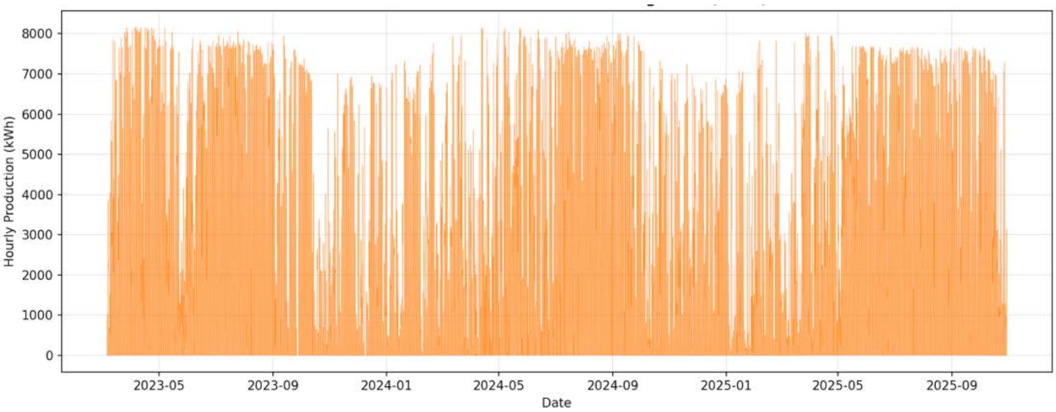

Figure 4 - Hourly Solar Power Production (kWh) for Plant 1 Solar

Figure 4 presents the hourly solar production time series for the Plant 1 Solar over the common analysis period (October 2023- November 2025). The series exhibits a pronounced diurnal cycle, with systematic periods of zero production during night-time and sharp ramp-up and ramp-down phases around sunrise and sunset. Seasonal modulation is also clearly visible, with higher amplitudes during summer months and reduced production during winter. The corresponding time series for the remaining solar plants are provided in Figure A 1 - Figure A 5.

This regular pattern makes solar generation predictable, but uncertainty concentrates around sunrise and sunset transitions. Rapid changes in output during dawn and dusk, combined with weather-driven variability during daylight hours, create regimes in which forecasting errors tend to cluster temporally. These characteristics are representative of utility-scale photovoltaic generation and have been extensively documented in the literature (Pinson, 2013; Soman et al., 2010; Yang et al., 2012).

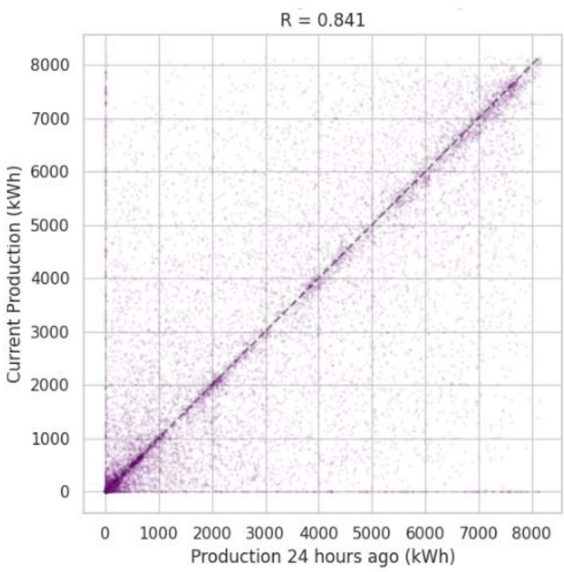

Figure 5 - Lag-24h Autocorrelation of Solar Production (Plant 1 Solar)

The raw time series suggests a high degree of regularity in solar generation and a more erratic behaviour for wind; these differences become clearer when persistence across consecutive days is explicitly examined.

Figure 5 illustrates the lag-24-hour autocorrelation of solar production at Plant 1 Solar. The strong correlation (r = 0.841) indicating the persistence of production patterns across consecutive days, indicating that daily production profiles exhibit strong inter-day dependence (Soman et al., 2010; Yang et al., 2012). Corresponding autocorrelation plots for the remaining solar plants are provided in Figure A 6 - Figure A 10

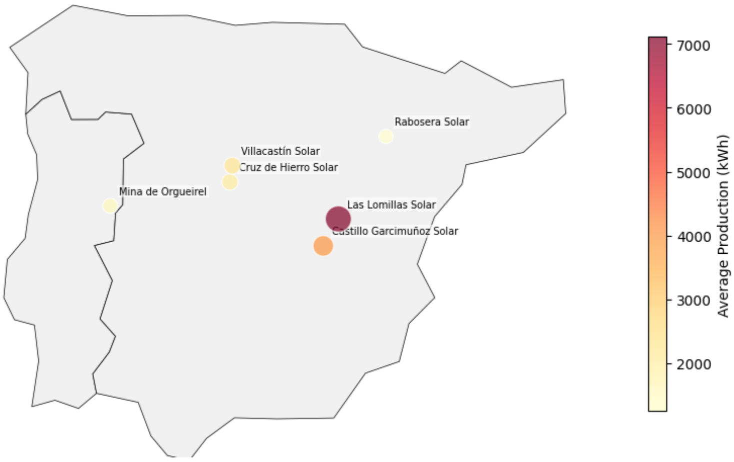

Figure 6 - Map of Average Energy Production for Solar Farms

While the spatial intensity map depicts cross-site variation in average solar production, it does not capture the temporal distribution of output at individual plants. To complement the spatial perspective, the distributional properties of solar generation are examined next.

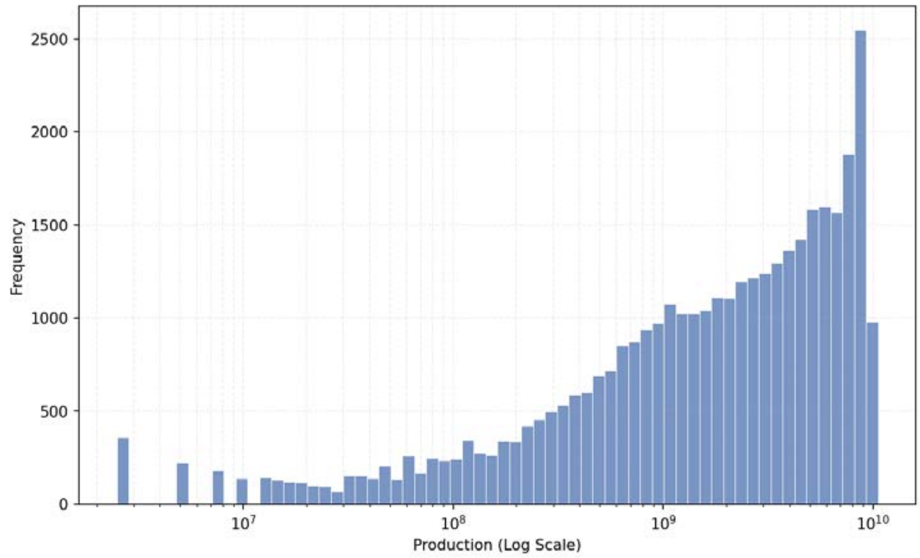

Figure 7 - Hourly distribution of solar generation, logarithmic scale on the x axis, base 10 (Plant 1 Solar)

Figure 7 shows the empirical distribution of hourly solar production at Plant 1 Solar. The histogram reveals a strongly right-skewed distribution, characterised by a pronounced point mass at zero corresponding to night time hours, followed by a long right tail associated with daylight production. This distributional shape reflects the intrinsic intermittency of photovoltaic generation, combining deterministic zero-output periods with a wide range of possible production levels during daytime hours.

This asymmetry has direct implications for forecasting performance and error interpretation. While timing is predictable, uncertainty concentrates in specific regimes, particularly under low-irradiance conditions and during transition periods. This leads to non-Gaussian error structures that are not adequately captured by average metrics alone (Gonçalves et al., 2021; Pinson, 2013). The solar production histogram is dominated by structurally imposed zero-production periods and bounded by a deterministic diurnal cycle. Wind generation, by contrast, lacks an equivalent organizing mechanism, which motivates a direct examination of the temporal evolution of wind output. Equivalent distributional results for other solar plants are provided in Figure A 11 - Figure A 15.

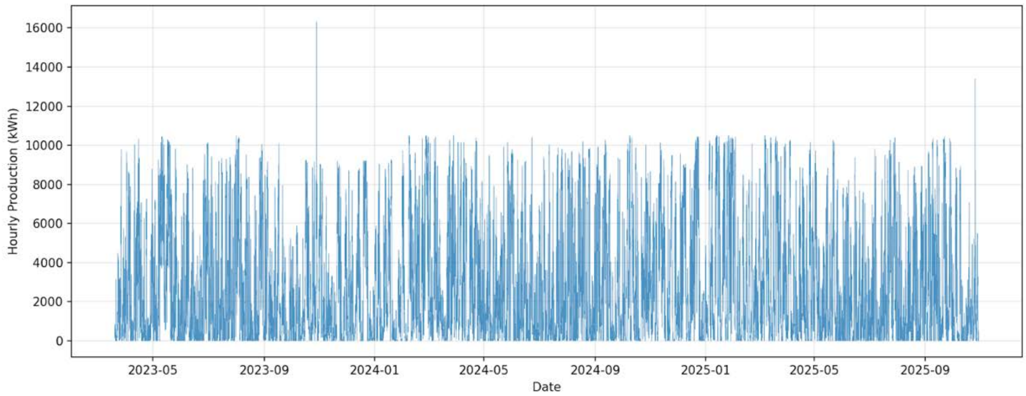

Figure 8 - Hourly Solar Power Production (kWh) for Plant 1 Solar

Figure 8 presents the corresponding hourly wind production time series for the Plant 1 Wind farm within the same hybrid configuration. In contrast to solar generation, wind production displays a markedly less structured temporal behaviour, characterised by fluctuations across multiple time scales, extended periods of low production, and sporadic high-output events. No fixed daily pattern is observed, and variability is primarily driven by evolving meteorological regimes instead to deterministic cycles. Comparable wind production time series for the remaining hybrid asset presented in Figure A 16 Figure A 20.

This signal shows weaker persistence, temporally dispersed uncertainty, and abrupt changes that often lack warning signs in recent data. This complicates short-term predictability and leads to forecasting errors that are less temporally concentrated but often larger in magnitude. These features are consistent with previous studies on wind power variability and forecasting challenges (Yang et al., 2012).

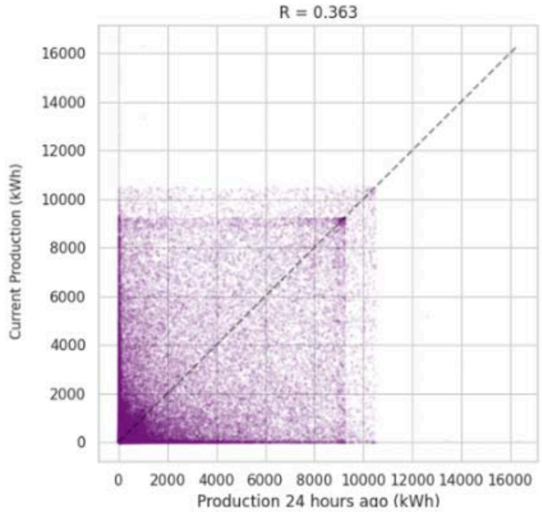

Figure 9 - Lag-24h Autocorrelation of Wind Production (Plant 1 Wind)

Figure 9 presents the lag-24-hour autocorrelation of wind production at the Plant 1 Wind farm. The substantially lower correlation (r = 0.363) confirms the weaker daily persistence of wind generation and the dominant role of evolving atmospheric conditions. Wind output lacks a stable diurnal cycle and is instead shaped by synoptic-scale meteorological regimes. Previous studies document this pattern in wind forecasting literature and associate it with higher variability, regime shifts, and reduced temporal coherence (Pinson, 2013; Soman et al., 2010). The lag-24 autocorrelation analysis was replicated for all wind assets, with detailed results available in Figure A 21 Figure A 25 .

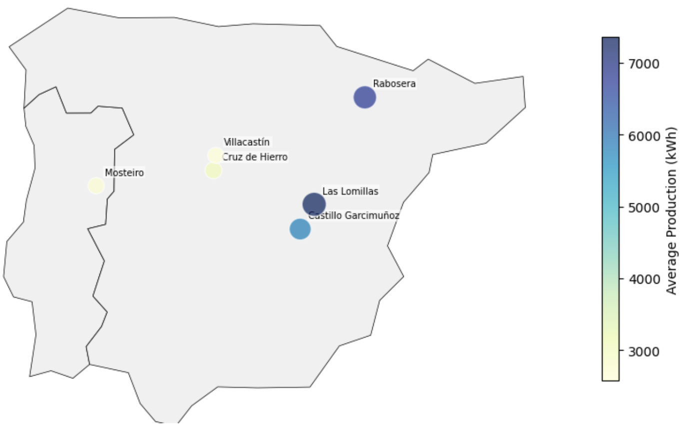

Figure 10 - Map of Average Energy Production for Wind Farms

The spatial intensity map reveals strong location-dependent patterns. However, spatial differences alone do not fully characterise the uncertainty structure of wind production.

To examine the variability, the analysis turns to the distribution of hourly wind generation.

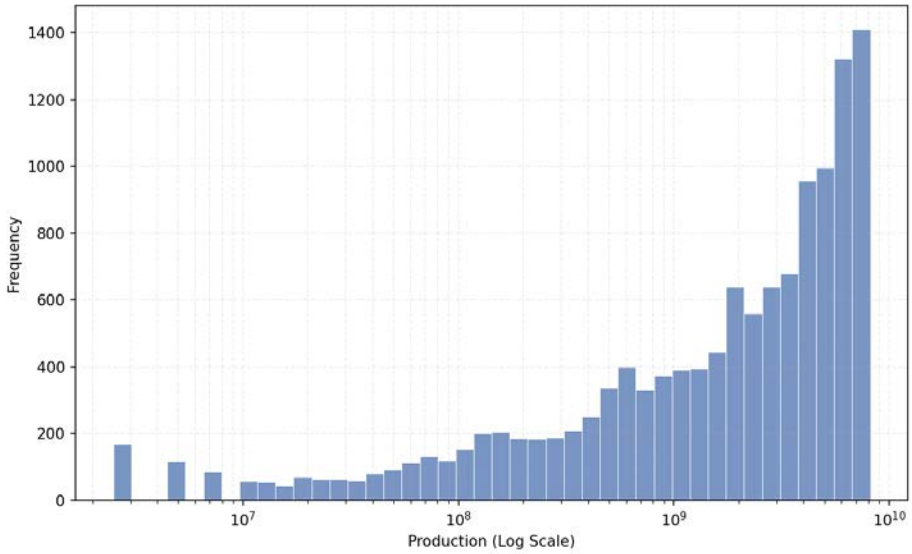

Figure 11 - Hourly Distribution of Wind Generation, logarithmic scale on the x axis, base 10 (Plant 1 Wind)

The empirical distribution of hourly wind production at the Plant 1 Wind farm is shown in Figure 11. The wind distribution displays a markedly broader spread and a heavier tail, reflecting a predominance of low-to-moderate production hours punctuated by sporadic high-output events. Although zero-production periods are not structurally imposed, as in photovoltaic generation, extended low-output regimes remain frequent, resulting in a highly skewed, heavy-tailed distribution. Similar distributional patterns across the other wind plants are documented in Figure A 26 - Figure A 30.

This reveals a fundamentally different form of uncertainty. Instead of concentrating in predictable temporal windows, wind variability is distributed across time and shaped by regime changes that are difficult to anticipate from recent observations, a feature consistently reported in the wind forecasting literature. Combining this uncertainty with the structured solar profile, a clear temporal asymmetry between the two technologies emerges. This asymmetry becomes evident at the intraday scale, where periods of high solar output tend to coincide with reduced wind generation, and vice versa, as illustrated in Figure 12

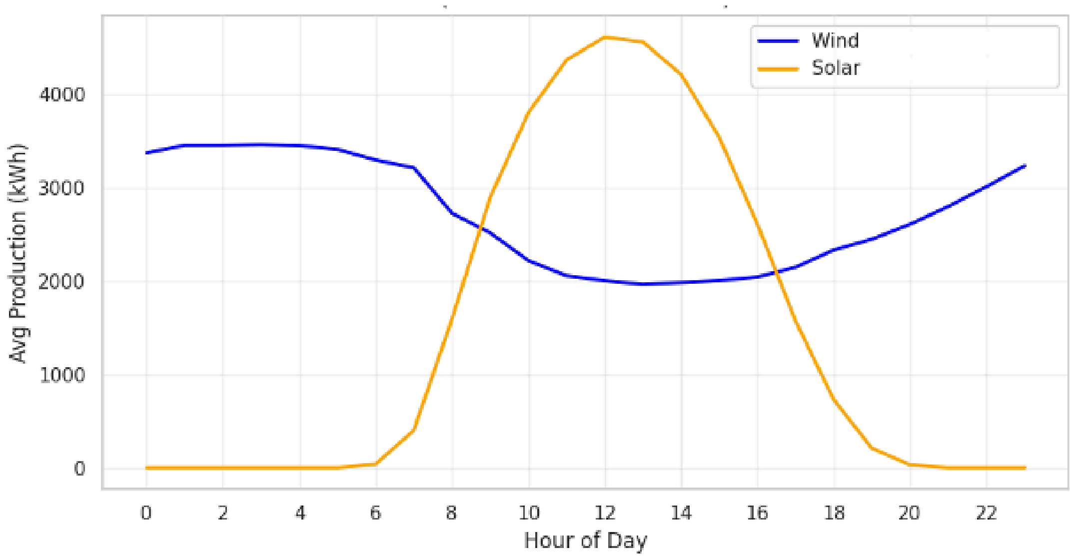

Figure 12 - Average Daily Solar-Wind Complementarity (2023-10-01- 2025-11-01)

Solar production (Plant 1 Solar) is largely confined to daylight hours, while wind generation (Plant 1 Wind) remains active during periods of low or zero irradiance. This asymmetry, illustrated in Figure 12 gives rise to recurrent hourly windows in which declining solar output towards the end of the day is partially offset by sustained or increasing wind generation. Even so, this interaction does not a strong structural anticorrelation, and it does not systematically smooth variability. This reflects a conditional form of complementarity, rather than a structural or systematic anticorrelation between the two technologies. While there are periods during which both technologies exhibit high or low production, such periods are comparatively infrequent. In most hours, solar and wind contribute in opposing temporal regimes, a behaviour consistent with empirical findings in the literature on hybrid systems (Couto & Estanqueiro, 2021).

This distinction matters in systems with high renewable systems, joint low-generation events pose the main risk to balancing and reserve adequacy, not average deviations. The intraday interaction observed here suggests that hybridisation can reduce the frequency of such events without eliminating exposure to them. The compensation effect is uneven across hours and strongly dependent on prevailing wind conditions, implying that complementarity at the hourly scale is probabilistic instead of deterministic. This matters more when moving beyond average daily behaviour to longer temporal horizons, where seasonal structure and interannual variability further condition the effectiveness of hybrid systems.

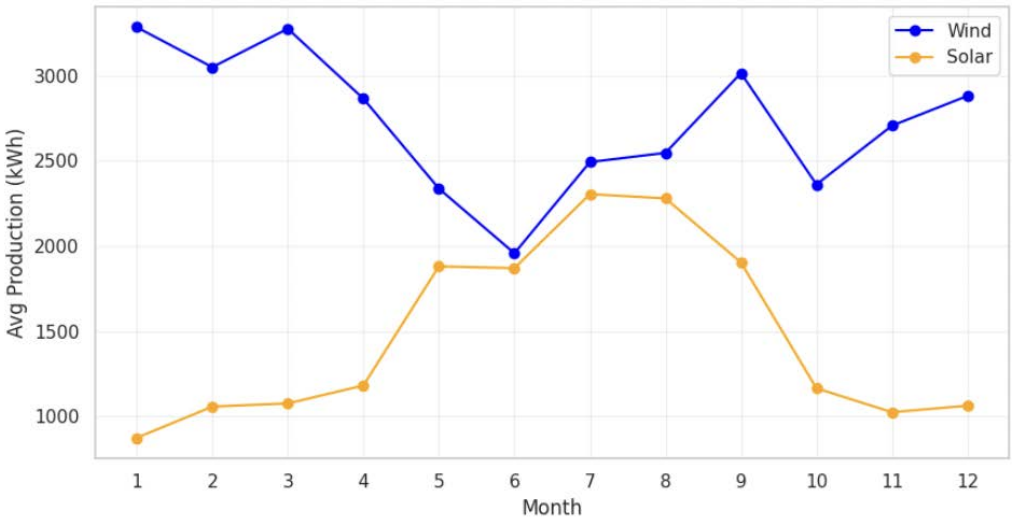

Figure 13 - Monthly Complementarity between Solar and Wind Production

At the seasonal scale, solar and wind generation reveals an apparent offset, with solar peaking in summer and wind producing higher average output during winter and shoulder seasons. This seasonal asymmetry is frequently cited in the literature as a structural advantage of hybrid systems, as it may reduce long-term variability when averaged over extended horizons.

However, seasonal averages alone do not capture the temporal stability of this interaction. Year-by-year analysis reveals important asymmetries emerge between the two technologies.

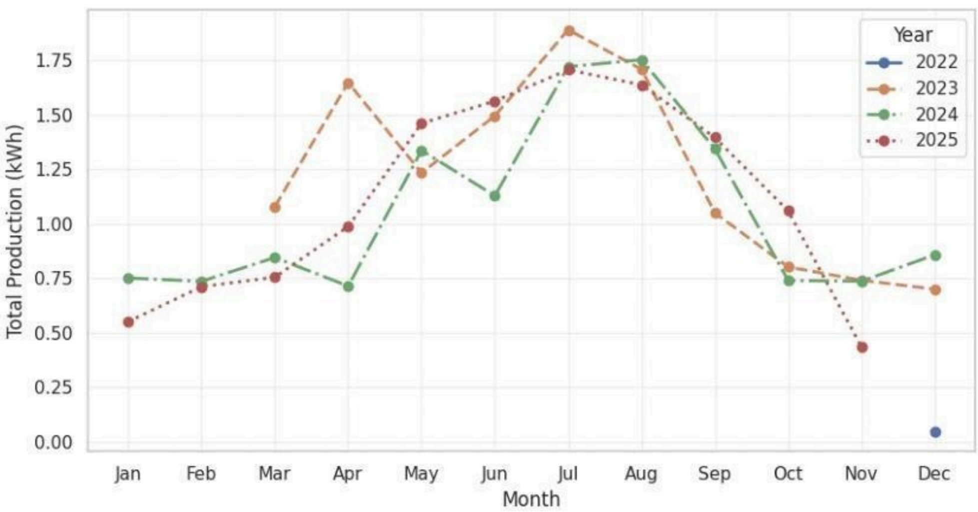

Figure 14 - Year-over-year Solar Production Profiles (Plant 1 Solar)

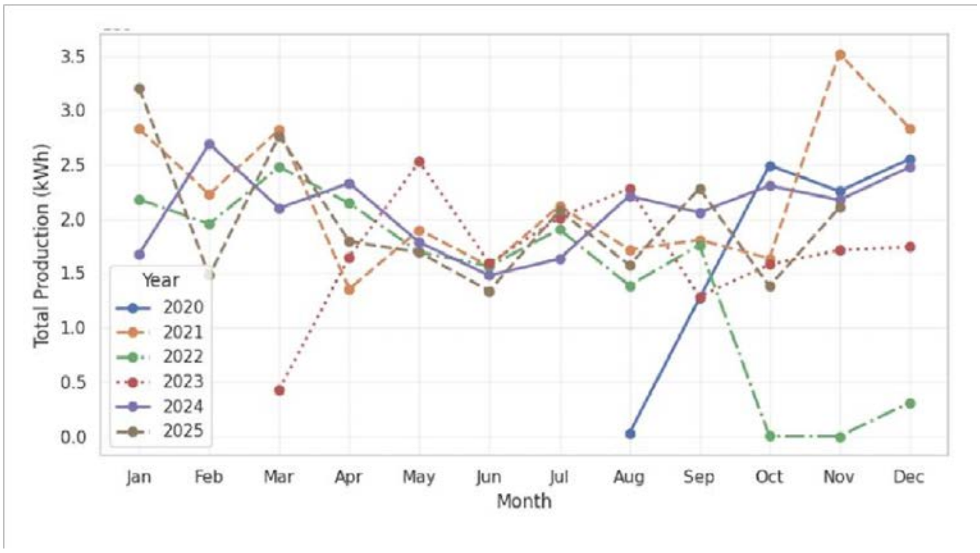

Figure 15 - Year-over-year Wind Production Profiles (Plant 1 Wind)

Solar generation displays a stable and repeatable seasonal profile across years, whereas wind generation represents substantial interannual variability, both in magnitude and in the timing of seasonal peaks and troughs. In some years, wind production reinforces the expected seasonal offset with solar whereas in others it weakens or partially counteracts it. This behaviour reflects the sensitivity of wind resources to large-scale atmospheric regimes and is consistent with findings reported for the Iberian Peninsula (Couto & Estanqueiro, 2021). These patterns indicates that seasonal complementarity is non-stationary and regime dependent. Hybridisation reduces variability on average but does not eliminate exposure to unfavourable years or persistent joint low-production regimes in which both technologies underperform their long-term means. The interannual analysis was replicated for all assets, detailed in Figure A 31 Figure A 35 for Solar profile and Figure A 36 - Figure A 40 for wind.

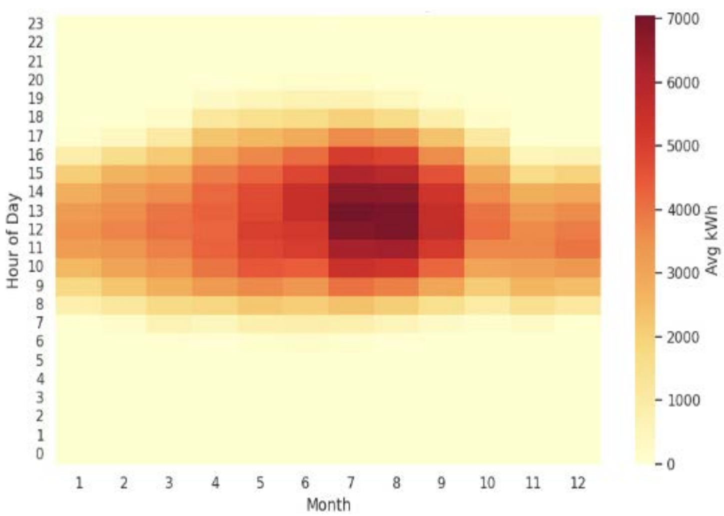

Figure 16 - Hourly-Monthly Production Heatmap for Solar (Plant 1 Solar)

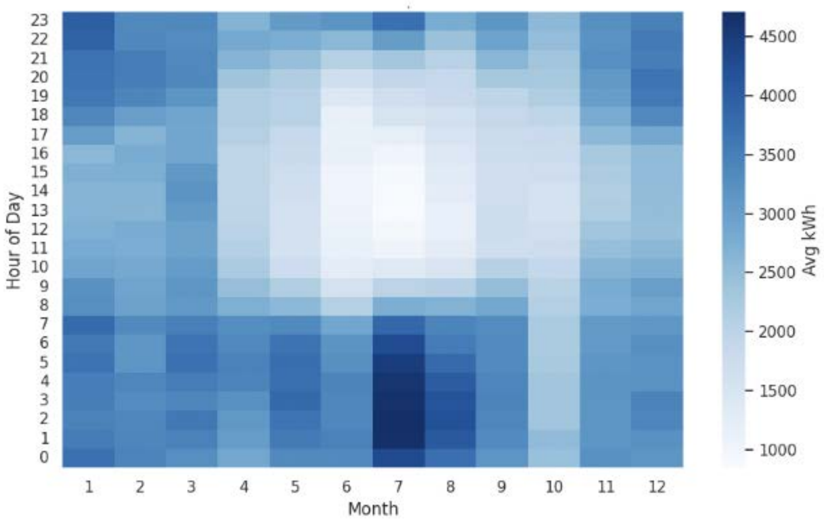

Figure 17 - Hourly-Monthly Production Heatmap for Wind (Plant 1 Wind)

A finer-grained view of temporal structure reveals that complementarity between solar and wind as represented in Figure 16 and Figure 17 is neither uniform nor constant across the year. When production is examined jointly across hours and months, extended periods of low output can be observed for both technologies, alongside episodes of pronounced inter-technology compensation. These patterns indicate that complementarity emerges unevenly in time and is strongly conditioned by both diurnal and seasonal dynamics. A clear pattern still appears across many hour-month combinations, high output from one technology often coincides with low output from the other. In particular, peak solar production around midday and during summer months frequently corresponds to comparatively lower wind output, while wind generation remains active during night-time hours when solar production is structurally zero. This behaviour reflects a form of temporal “fit” between resources, in which compensation operates probabilistically.

This compensation does not prevent joint low-production events, but it does reduce their frequency relative to what would be expected under independent or fully correlated behaviour. The hour-month analysis was systemically extended to all hybrid configurations, with full results available in Figure A 41 Figure A 45 for solar plants and Figure A 46 Figure A 50 for wind plants.

This intra-seasonal and intraday variability highlights a limitation of aggregate metrics. Annual or monthly averages are insufficient to capture the operational value of complementarity, as they obscure the timing and persistence of favourable and unfavourable regimes. This confirms the need for multi-scale analysis when assessing hybrid systems, particularly in forecasting and risk-sensitive applications where the temporal structure of uncertainty is as important as its magnitude (Kong et al., 2019; Soman et al., 2010).

Table 5 - Hybridization matrix (Pearson correlation) - solar and wind production

Wind Farm	Solar Reference	Pearson´s R	P- value
Plant Wind 1	Plant Solar 1	-0.272	< 0.001
Plant 2 Wind	Plant Solar 1	-0.259	< 0.001
Plant 3 Wind	Plant Solar 1	-0.174	< 0.001
Plant 6 Wind	Plant Solar 1	-0.089	< 0.001
Plant 5 Wind	Plant Solar 1	-0.061	< 0.001

All analysed plant pairs, defined as the Plant Solar 1 reference site and each of the five wind farms with available concurrent data, demonstrate negative and statistically significant global Pearson correlation coefficients, indicating a systematic inverse association between solar and wind generation. The moderate magnitude of these correlations confirms that the relationship is moderate in magnitude, implying partial but not deterministic compensation. Periods of simultaneous high or low production do occur, but they are comparatively infrequent, with most hours characterised by opposing temporal regimes. This result is consistent with the temporal patterns identified across intraday and seasonal scales and aligns with previous empirical findings on hybrid renewable systems (Couto & Estanqueiro, 2021; Yang et al., 2012).These structural characteristics define the statistical environment within which forecasting performance and error behaviour must be evaluated, providing a natural transition to the subsequent analysis.

4.2 COMPANY FORECASTING PERFORMANCE AND ERROR CHARACTERISTICS

To contextualize alternative formulations, this section evaluates the company's operational forecasts by analysing the relationship between observed production, forecasted production and the resulting error structure. The objective is not to critique modelling choices, but to characterise the statistical properties of forecast errors under real-world industrial conditions. Observed production refers to measured energy output, the company forecast denotes the values provided by the operator and forecast error is defined as the difference between observed production and the company forecast. The analysis begins with solar generation, as photovoltaic production is characterised by a strong deterministic component associated with the diurnal cycle and seasonal irradiance patterns

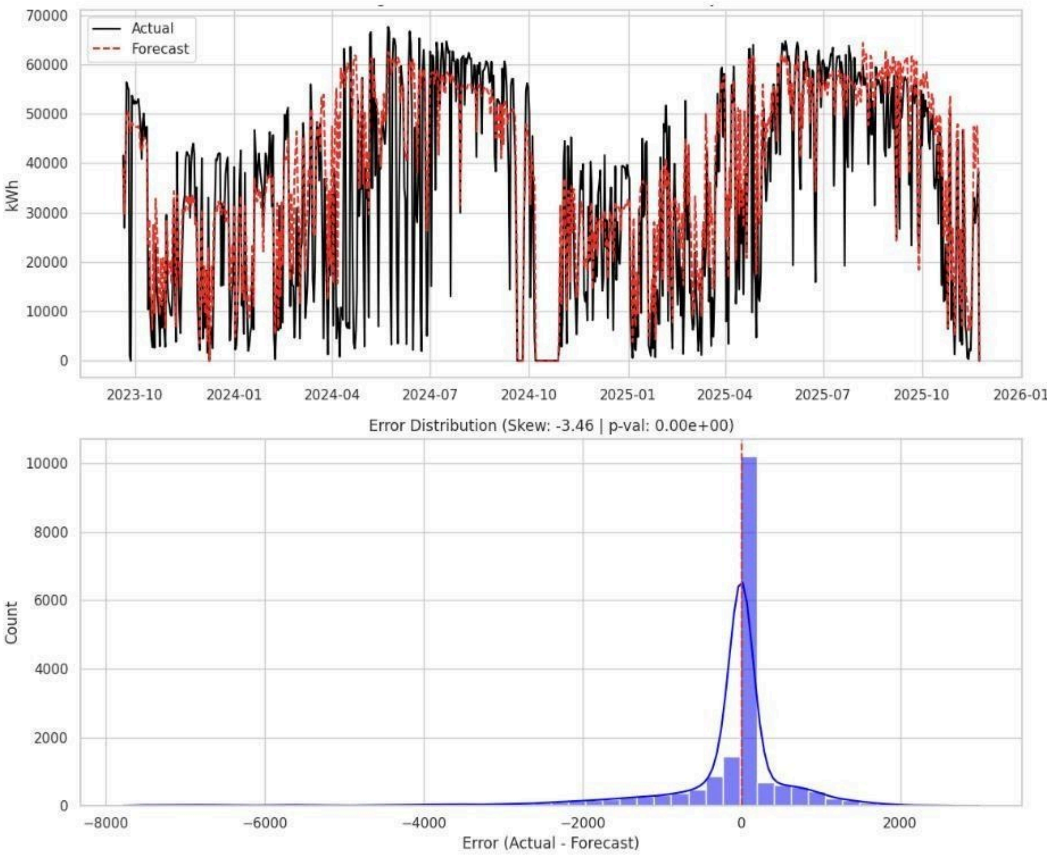

Figure 18 - Plant 1 Solar Production: Observed vs Company Forecast and Errors

Figure 18 compares observed daily solar production at the Plant 1 Solar with the corresponding company forecasts, together with the associated error distribution. The forecasts reproduce the seasonal evolution of solar generation reasonably well, indicating that large-scale irradiance patterns are effectively captured. However, deviations become evident during periods of high production and rapid ramp-ups around peak-generation days.

The error distribution reveals a pronounced asymmetry, with a heavier negative tail indicating systematic underestimation during high-output periods. Such skewness reflects the difficulty to accurately anticipating abrupt irradiance fluctuations and peal generation events. Although average errors remain moderate, extreme deviations contribute disproportionately to total forecasting uncertainty. This behaviour is consistent with established findings in solar forecasting literature, where cloud-driven variability and non-linear irradiance dynamics generate asymmetric and heavy-tailed error structure (Pinson, 2013; Soman et al., 2010).

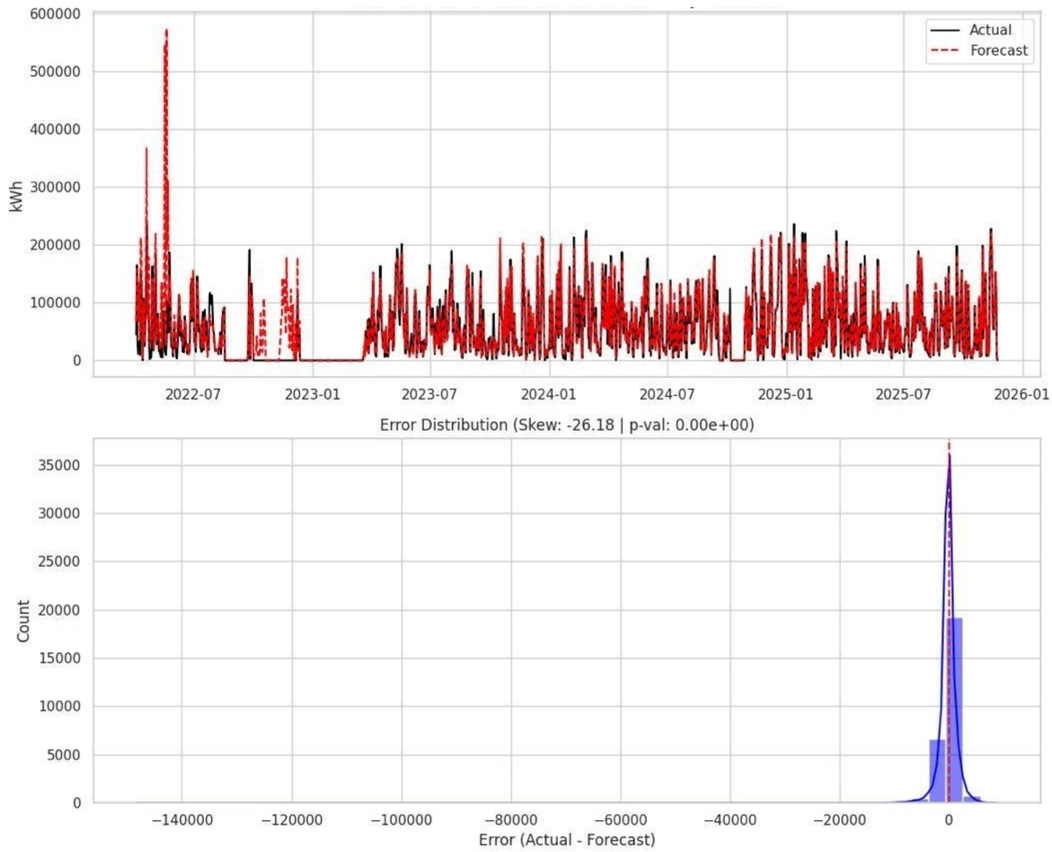

Figure 19 - Plant 1 Wind Production: Observed vs Company Forecast and Errors

Observed and forecasted daily wind production for the Plant 1 Wind farm, as represented by Figure 19, together with the associated error distribution. This site is characterised by frequent low-to-moderate production interspersed with episodic high-output events. While company forecasts capture the general level of activity, large discrepancies are observed during abrupt regime transitions, leading to error distributions with heavy tails and strong skewness. This shows the challenge of anticipating sudden wind ramps and regime shifts, a well-documented challenge in wind forecasting (Gonçalves et al., 2021; Pinson, 2013).

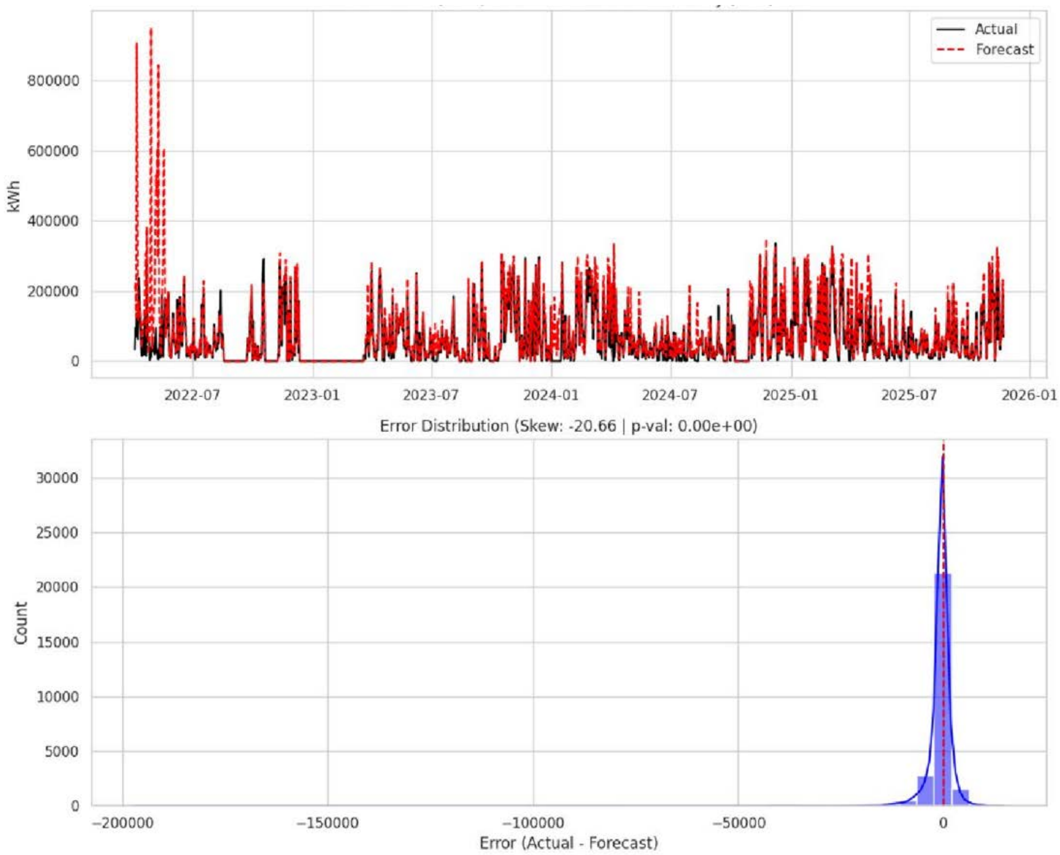

Figure 20 - Plant 3 Wind Production: Observed vs Company forecast and Errors

Figure 20 shows the corresponding results for the Plant 3 Wind farm. In contrast to Plant 1 Wind, this plant exhibits a different production regime, with fewer but more intense high-output events. Forecast errors at Plant 3 Wind display even heavier tails, highlighting how site-specific characteristics, such as exposure, orography, and prevailing wind regimes, directly influence forecast performance. The comparison between. Figure 19 and Figure 20 illustrate that wind forecast errors are not homogeneous across locations, even when produced using the same forecasting framework. This heterogeneity motivated the inclusion of multiple wind plants in the analysis, ensuring that subsequent conclusions regarding hybrid forecasting are not driven by the behaviour of a single wind regime, but instead reflect a broader range of operational conditions.

The analysis of Figure 18 to Figure 20 shows that company forecasts capture average production behaviour for both solar and wind generation, but remain exposed to large and asymmetric errors that concentrate risk in extreme events. For solar generation, uncertainty is primarily associated with high-production periods, whereas for wind generation it is dominated by regime-driven variability and plant specific dynamics. Instead of attempting to further improve the predictability of each technology in isolation, these observations motivate a shift towards a system-level perspective. The observed Vs Forecast analysis was replicated for all remaining wind assets, with detailed results reported in Figure A 51 - Figure A 54

Operational risk in power systems with high renewable penetration is driven primarily by extreme forecast errors not by average deviations. For this reason, the subsequent analysis focuses explicitly on tail behaviour and the structural drivers of extreme imbalance events.

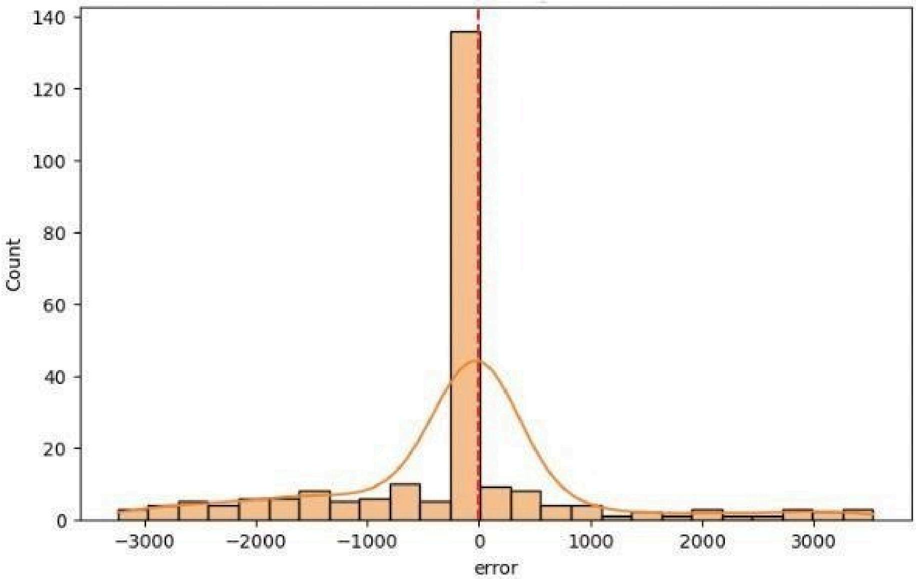

Figure 21 - Solar Error Distributions Based on Company Forecasts (Plant 1 Solar)

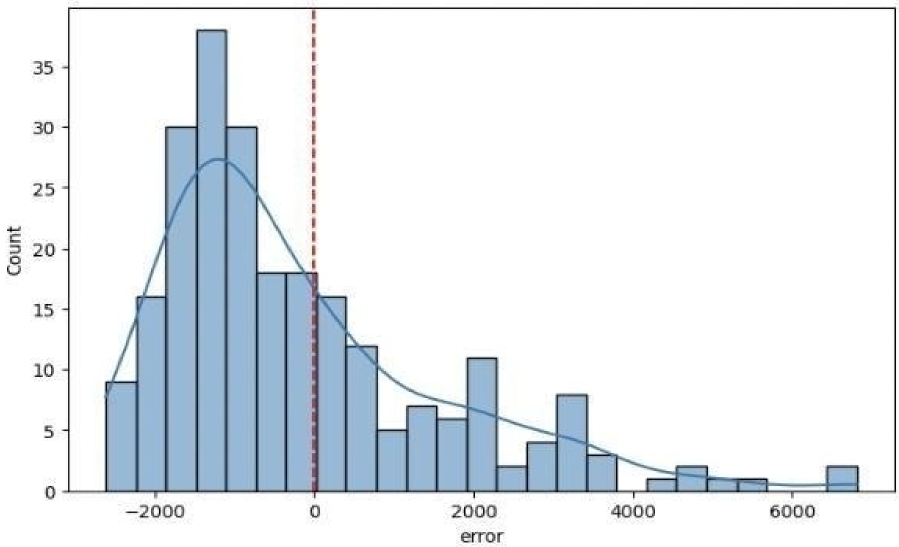

Figure 22 - Wind Error Distributions Based on Company Forecasts (Plant 1 Wind)

The error distributions for solar and wind generation are presented in Figure 21 and Figure 22. Both distribution exhibit clear asymmetry and heavy tails, indicating the forecast errors deviate from Gaussian assumptions. Extreme deviations represent a limited proportion of observations, yet account for a disproportionate share of total error and are operationally relevant in imbalance-driven systems. (Gonçalves et al., 2021; Tina & Gagliano, 2011). The error distribution assessment was replicated across all solar and wind assets. Detailed solar results are documented in Figure A 55 Figure A 59, and wind results in Figure A 60 - Figure A 64.

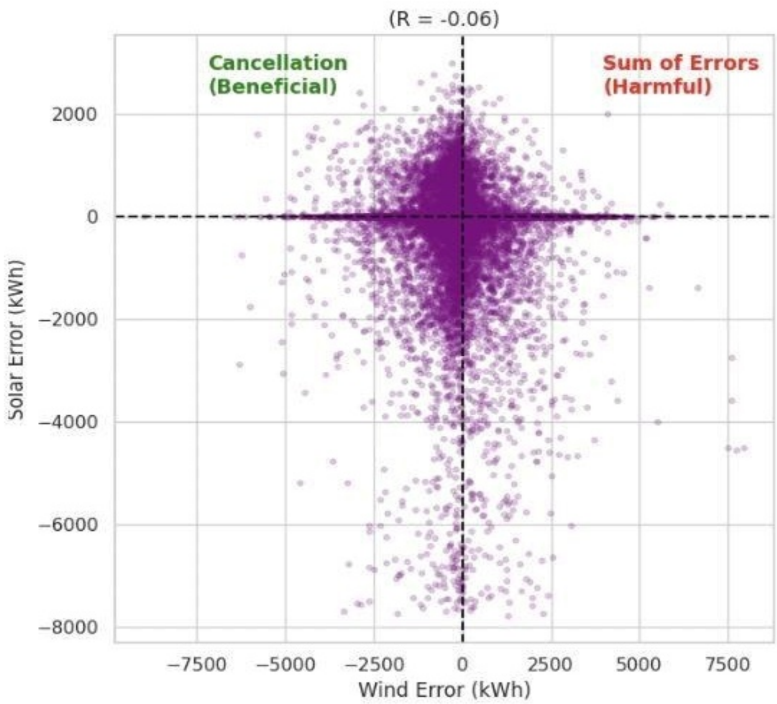

Figure 23 - Complementarity Between Solar and Wind Forecast Errors

Figure 23 presents the joint distribution of solar and wind forecast errors. The near-zero correlation between errors ($r \approx -0.06$) indicates the absence of strong linear dependence across technologies, although interaction effects remain visible in the joint distribution. A substantial number of observations fall in regions where positive and negative errors partially cancel each other, while fewer correspond to simultaneous large errors of the same sign. This behaviour suggests that hybrid configurations may benefit from error interaction effects, even when individual forecasts remain imperfect, an insight aligned with earlier empirical findings on hybrid systems (Yang et al., 2012; Couto & Estanqueiro, 2021). The system-level implications of this interaction are examined by directly comparing isolated and hybrid forecasting formulations

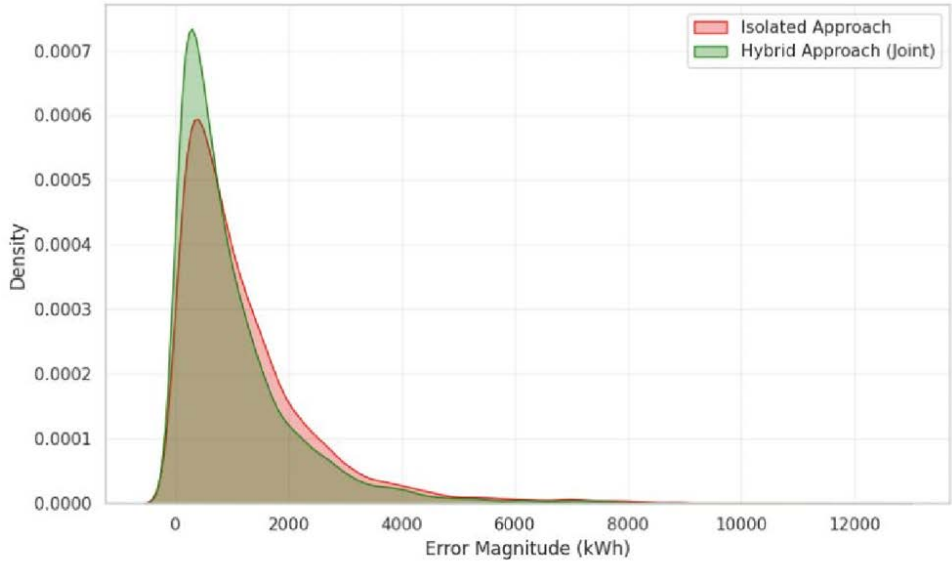

Figure 24 - Error Magnitude Distributions for Isolated and Hybrid Forecasting

Figure 24 contrasts the distribution of error magnitudes obtained when forecasting each technology separately with those obtained when forecasting the aggregated hybrid output. The hybrid formulation exhibits a systematic reduction in the probability of large error magnitudes, in the upper tail of the distribution. This reduction does not stem from uniformly improved predictions, but from a reshaping of uncertainty that reduces exposure to extreme deviations.

The statistical properties of forecast errors at peak events are represented in Table 6 which shows that errors remain asymmetric and non-normal under both formulations. This reinforces the importance of focusing on tail behaviour rather than relying solely on average metrics. Although peak MEA does not uniformly decrease across technologies, the system-level aggregation reduces the joint probability of extreme imbalance events, highlighting that the benefit of hybrid forecasting emerges from uncertainty interaction rather than uniform error minimisation.

Table 6 - Statistical Properties of Forecast Errors at Peak Events

MODEL	TECH	MAE AT PEAKS (kWh)	ASYMMETRY	NORMALITY
Reference	Solar	456	Negative	Rejected
Proposed	Solar	550	Negative	Rejected
Reference	Wind	2400	Negative	Rejected
Proposed	Wind	2253	Positive	Rejected

The overall system-level benefit of hybrid forecasting at the point of interconnection is quantified in Table 7.

Table 7 - Isolated vs Forecasting Perfomance (Hybrid Plant 1)

THESIS RESULTS	TECH
Isolated MAE	2072.11 kWh
Joint MAE	1646.43 kWh
UNCERTAINTY REDUCTION	20.54%

A reduction of 20.54% in MAE is observed when moving from isolated to joint forecasting. This improvement is achieved without materially improving the predictability of solar or wind generation in isolation. The value of the hybrid approach instead emerges from aggregation effects and error interaction. The value of the hybrid formulation does not arise from improved predictability at the component level, but from structural aggregation and error interaction effects at the system boundary.

4.3 Model-based forecasting: performance and uncertainty reshaping

Following the analysis of company forecasts, this section evaluates how forecasting performance can be improved at the modelling level. In particular, this section investigates two complementary questions:

(1) How do different modelling approaches perform when forecasting solar and wind generation individually?
(2) Do improvements at the model level translate into meaningful reductions in uncertainty, especially under operationally relevant conditions?

Before comparing model performance, it is important to clarify how the input data are represented. Renewable generation time series are originally expressed in physical units (kWh) and exhibit strong amplitude variability, zero-production periods and pronounced seasonal modulation. These characteristics, while physically meaningful, pose challenges for learning-based models and can hinder stable optimisation.

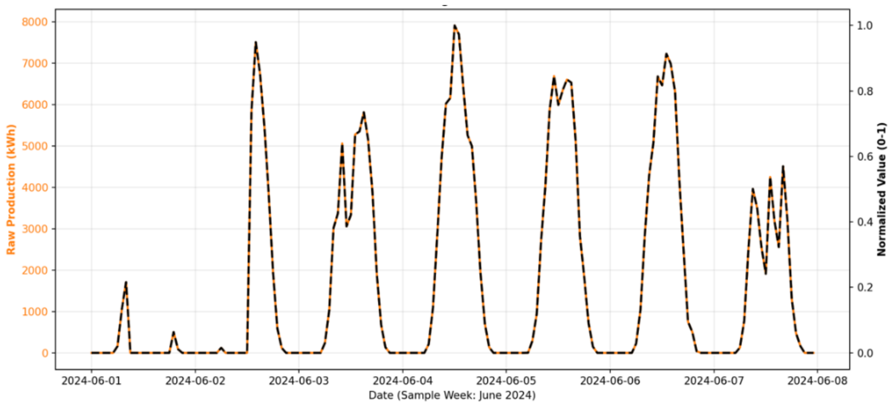

Figure 25 - Raw vs Normalized Solar Production Time Series (Plant 1 Solar)

As illustrated in Figure 25, an excerpt of the original hourly solar production time series alongside its normalised counterpart. The transformation preserves the underlying temporal structure of the signal, including intermittency, ramps and regime changes, while rescaling amplitudes to a numerically stable range. This normalised representation is used consistently across all modelling approaches evaluated in this section, ensuring comparability and robust training behaviour.

The analysis begins with a set of standard baseline models, which serve as a reference against which more advanced approaches can be evaluated. These include classical statistical benchmarks and commonly adopted time-series formulations.

Table 8 - Baseline Forecasting Perfomance (MAE and RMSE)

MODELS	**MAE (kWh)**	**RMSE (kWh)**
Seasonal Naive	54836.2	69044.4
SARIMA (Optimized)	49745.1	61222.0
LSTM (Bidirectional)	46348.2	55879.0

Table 8 reflects the performance of baseline models in terms of MAE and RMSE. As expected, naïve seasonal approaches yield the highest errors, reflecting their inability to capture complex temporal dynamics. Optimised SARIMA models offer moderate improvements, while bidirectional LSTM architectures achieve the lowest errors among the baseline candidates. These results are consistent with findings reported in the forecasting literature, where neural architectures tend to outperform linear models in the presence of non-stationarity and multi-scale temporal patterns. However, as shown later in this section, improved average metrics do not necessarily imply reduced exposure to extreme forecast errors. VAR models were also tested as part of the baseline analysis but are not reported here due to differences in experimental setup, which prevent a direct comparison with the remaining models.

To better understand how errors manifest over time, observed, and forecasted production trajectories were analysed dynamically. A rolling forecasting experiment was conducted to assess how different modelling approaches behave under realistic operational conditions, in which forecasts are repeatedly updated using a fixed historical window. This analysis allows a direct comparison of model responsiveness, stability, and sensitivity to regime changes, beyond what is captured by aggregate error metrics.

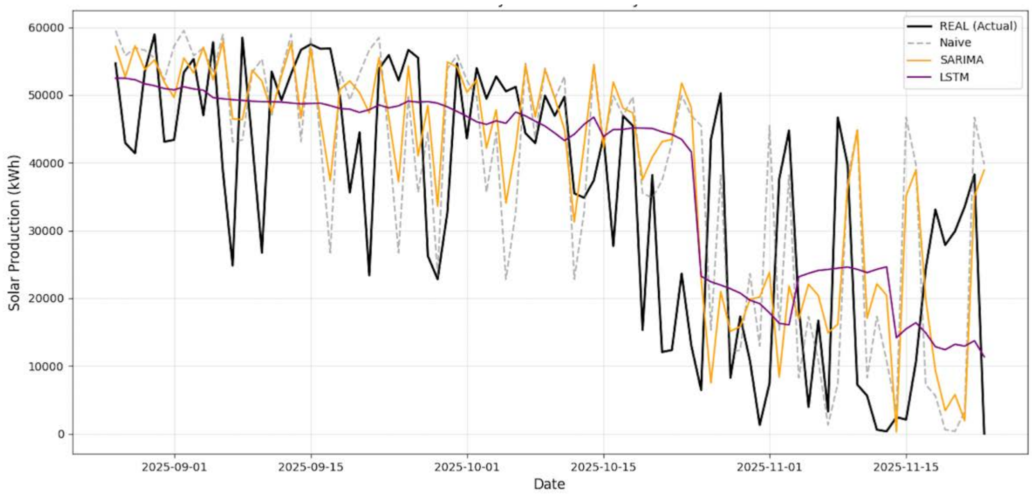

Figure 26 - Rolling Validation: Baseline Models (60d window,10 horizon - Solar)

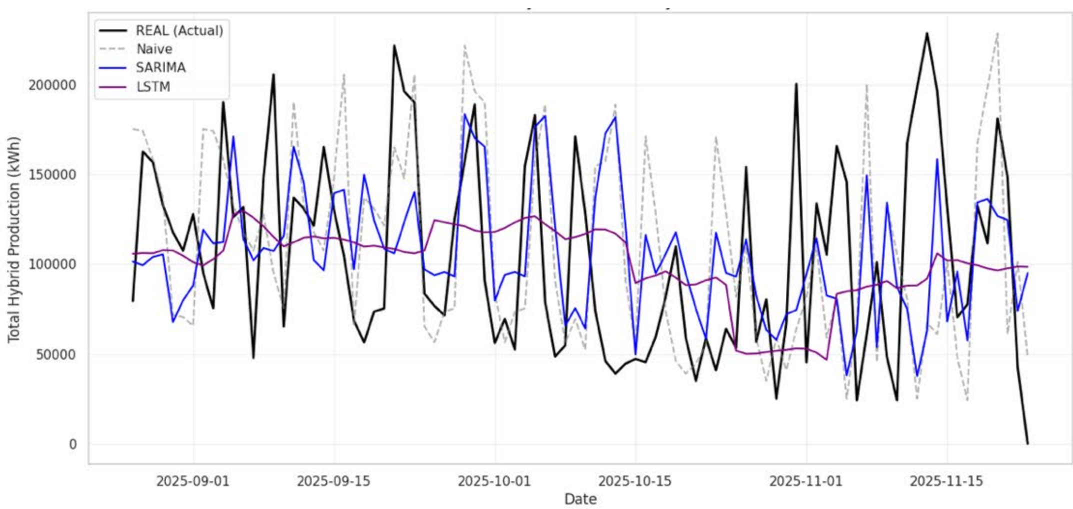

Figure 27 - Rolling Validation: Baseline Models (60d window,10 horizon - Wind)

The behaviour of the baseline forecasting models under a rolling validation framework. While the Seasonal Naive and SARIMA models tend to react sharply to short-term fluctuations, the LSTM produces smoother trajectories that better capture the underlying production dynamics. These differences highlight the trade-off between reactivity and stability across modelling approaches, motivating the selection of neural architectures for subsequent analyses.

When considering the rolling forecast behaviour across different technologies, distinct error dynamics emerge. For solar generation, forecast trajectories exhibit a strong alignment with the underlying diurnal structure, with deviations primarily concentrated around peak production periods. In this case, model differences are mainly reflected in the smoothness and temporal coherence of the forecasts, with neural architectures producing more stable trajectories and fewer abrupt adjustments across successive forecasting windows. In contrast, rolling forecasts for wind and aggregated hybrid production display higher temporal variability and more pronounced regime shifts, leading to larger and more persistent deviations over several days. These differences highlight the technology-dependent nature of forecasting uncertainty and reinforce the need to assess model behaviour dynamically, rather than relying solely on aggregate performance analysis.

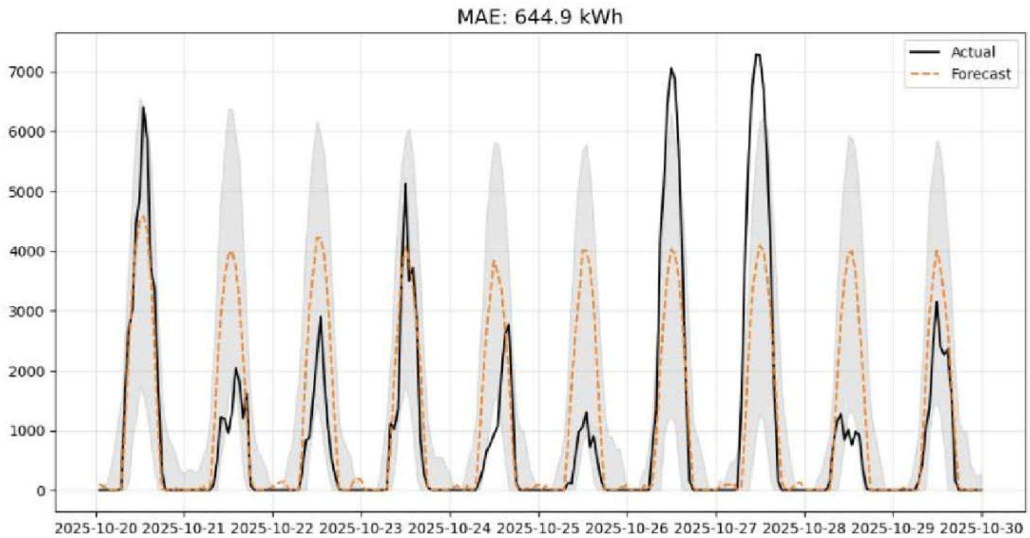

Figure 28 - Observed vs Forecasted Solar Power Production (Plant 1 Solar)

Figure 28 compares observed and forecasted solar production at Plant 1 Solar prior to the application of the proposed modelling framework, including both the ReCycle mechanism and the forecasting model. The grey band represents the predictive uncertainty interval (P5–P95), illustrating that while the central forecast follows the diurnal structure reasonably well, uncertainty remains concentrated around peak production hours. This behaviour mirrors the characteristics identified earlier in company forecasts, confirming that solar uncertainty is structurally linked to high-irradiance regimes rather than to model choice alone. Additional plant-level results are compiled in Figure A 65 - Figure A 69

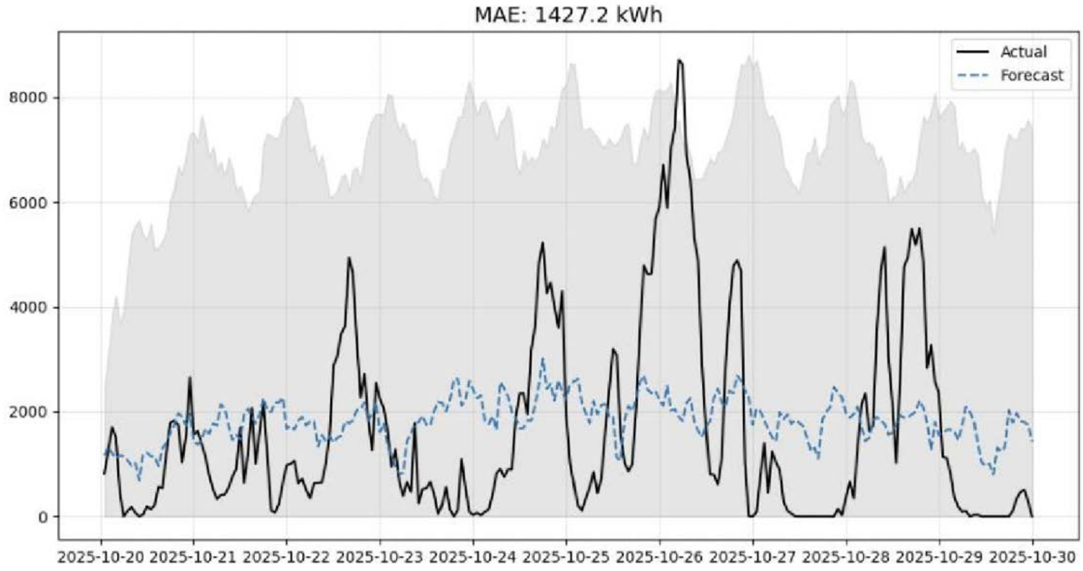

Figure 29 - Observed vs Forecasted Wind Power Production (Plant 1 Wind)

In contrast to solar, Figure 29 demonstrates, forecast errors for wind are less tied to deterministic cycles and instead reflect abrupt regime changes and sustained deviations over several days. Even when average performance is acceptable, large deviations persist, reinforcing the idea that wind forecasting uncertainty is dominated by regime-driven variability. The proposed modelling approach introduces a reformulation of the forecasting task, with the aim of improving robustness rather than purely reducing average error. Results for the remaining plants are provided inFigure A 70 Figure A 74.

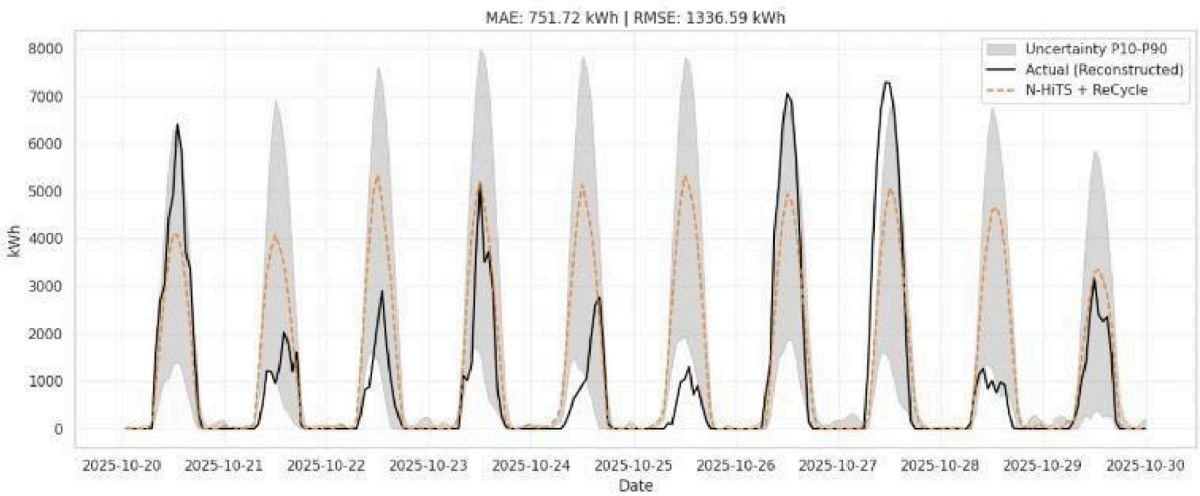

Figure 30 - Observed vs Proposal Model Forecasted Solar (10-day, Plant 1 Solar)

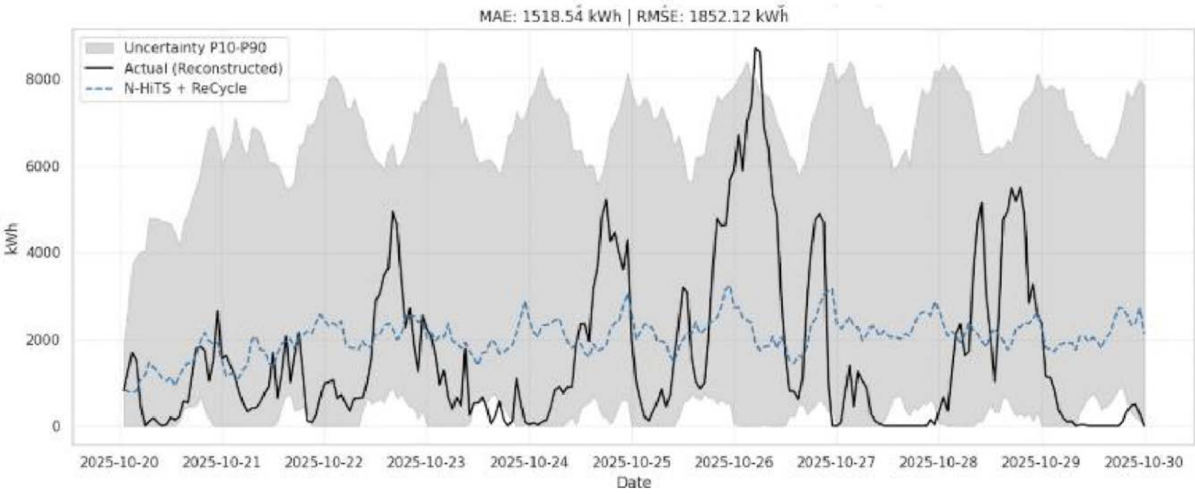

Figure 31 - Observed vs Proposal Model Forecasted Wind (10-day, Plant 1 Wind)

The comparison between observed production and the forecasts generated by the proposal model over a 10-day horizon highlights a qualitative shift in forecasting behaviour under the proposed formulation. Nevertheless, gains in point accuracy remain moderate, the evolution of uncertainty across time is markedly different. Forecast envelopes become more stable and less sensitive to abrupt fluctuations, and extreme deviations appear less persistent during periods associated with higher operational risk. Rather than producing sharper deterministic forecasts, the proposed approach modifies how uncertainty unfolds over time, suggesting that its primary contribution lies in improved robustness rather than marginal gains in average accuracy. This effect of the joint forecasting formulation should not be interpreted as an elimination of large forecast errors, but rather as a redistribution of uncertainty over time.

The impact of this reformulation is quantified using standard performance metrics evaluated at the plant level. Across both technologies, errors remain higher for wind than for solar generation, reflecting the intrinsically higher variability and regime-driven dynamics of wind resources. More specifically, performance remains stable across evaluation periods, indicating that the proposed approach does not introduce additional volatility or degradation under different operating conditions. To complement the qualitative insights from the rolling forecast analysis, quantitative performance metrics are now examined.

Table 9 - Forecasting Accuracy of Proposal Model (10-Day Horizon)

ID	MAE (kWh)	RMSE (kWh)
Plant 1 (Solar)	751.72	1336.59
Plant 2 (Wind)	1518.54	1852.12

Table 10 - Absolute Error of Proposal Model (10-Day Horizon)

ID	MONTH NAME	ABSOLUTE ERROR
Plant 1 (Solar)	October	751.72
Plant 1 (Wind)	October	1518.54

While Table 9 aggregates forecasting accuracy metrics (MAE and RMSE) by plant and technology for the 10-day horizon, providing a quantitative assessment of model performance. Table 10 complements this analysis by presenting absolute forecast errors for the reference hybrid plant over the evaluation month.

Beyond average metrics, the structure of forecast errors provides further insight into model behaviour. Point accuracy improves only moderately. However, changes in the error distribution reveal how uncertainty is reshaped under the proposed formulation. Error behaviour is consistent with earlier distributional findings, confirming that normality assumptions are inappropriate in operational forecasting, a result widely reported in the renewable forecasting literature (Pinson, 2013; Soman et al., 2010).

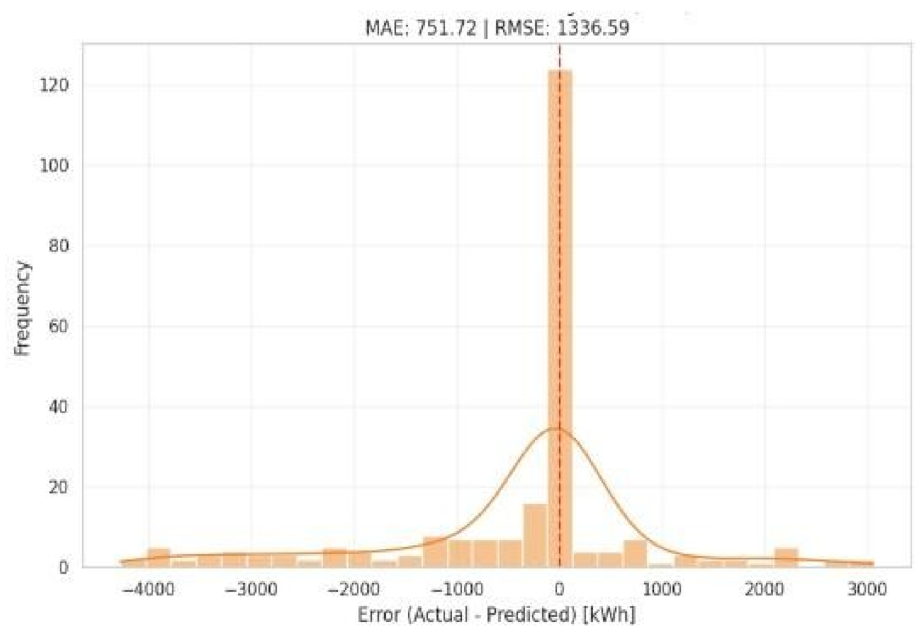

Figure 32 - Error Distribution of Proposal Model (10-day horizon, Plant 1 Solar)

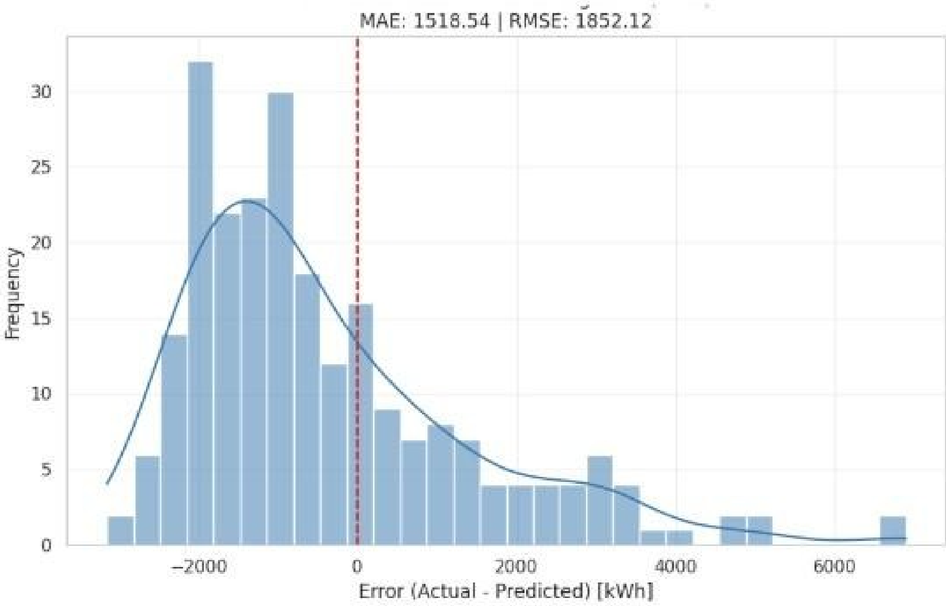

Figure 33 – Error Distribution of Proposal Model (Plant 1 Wind)

Figure 32 and Figure 33 highlight distinct characteristics for solar and wind generation. Solar forecast errors are more tightly concentrated around zero, with extreme deviations occurring less frequently, during high-production periods. This behaviour is consistent with previous studies showing that, while solar forecasting uncertainty increases during peak irradiance, it remains comparatively more bounded than wind-related uncertainty (Kong et al., 2019). Wind forecast errors, in contrast, exhibit heavier tails and stronger skewness, reflecting the persistence of regime-driven variability and site-specific dynamics, a known limitation of wind forecasting even under advanced modelling approaches (Pinson, 2013).

When compared with earlier results, the frequency and persistence of extreme deviations are reduced under the proposed formulation. This reduction does not eliminate large errors. Rather, it redistributes uncertainty over time, reshaping the temporal concentration of risk. In this sense, the benefit of joint formulation emerges not from deterministic accuracy gains, but from altering the statistical structure of forecast uncertainty, leading to fewer sustained extreme events, while extreme deviations may still occur sporadically. Similar effects have been observed in hybrid and aggregated forecasting contexts, where partial error cancellation and temporal smoothing emerge without requiring strong deterministic correlations between technologies (Couto & Estanqueiro, 2021; Yang et al., 2012).

Statistical indicators computed for peak events further confirm these observations. While error distributions remain heavy-tailed across plants and technologies, changes in bias and asymmetry suggest a more balanced error structure for solar generation. This finding reinforces the importance of evaluating forecasting performance beyond mean-based metrics, especially in power systems where operational costs and system risk are dominated by tail events rather than by average forecast accuracy.

Table 11 - Forecast Error Statistics Under Proposed Formulation (10-Day Horizon)

PLANT	TECH	MAE (kWh)	BIAS (kWh)	SKEWNESS	NORMALITY
Plant 1 Wind	Wind	1010	- 269	Negative	Rejected
Plant 2 Wind	Wind	3266	-1545	Strongly Negative	Rejected
Plant 3 Wind	Wind	1707	- 934	Negative	Rejected
Plant 6 Wind	Wind	1382	- 720	Slightly Negative	Rejected
Plant 5 Wind	Wind	3775	-1883	Negative	Rejected
Plant 1 Solar	Solar	456	- 210	Negative	Rejected

The statistical properties in Table 11 show that forecast errors remain non-normal under all configurations, with skewness and bias persisting across technologies. However, the observed reduction in extreme deviations and the stabilisation of error structure indicate that the proposed formulation reshapes uncertainty in a manner that is operationally more favourable. While normality remains rejected across all plants, the magnitude and persistence of extreme deviations are reduced relative to earlier formulations. This result aligns with the broader literature on risk-aware forecasting, which emphasises robustness and tail-risk mitigation as key performance dimensions in high-renewable power systems. While improvements in MAE and RMSE are moderate, the proposed formulation leads to a structural reduction in persistence and temporal clustering of extreme forecast errors (Kong et al., 2019).

Forecast performance is not spatially uniform, even when identical models and methodologies are applied across all sites. Error magnitudes vary significantly between locations, reflecting differences in local resource characteristics, terrain complexity, exposure, and the prevalence of extreme meteorological conditions. This spatial heterogeneity is a well-documented feature of renewable energy forecasting and persists regardless of modelling approach (Pinson, 2013).

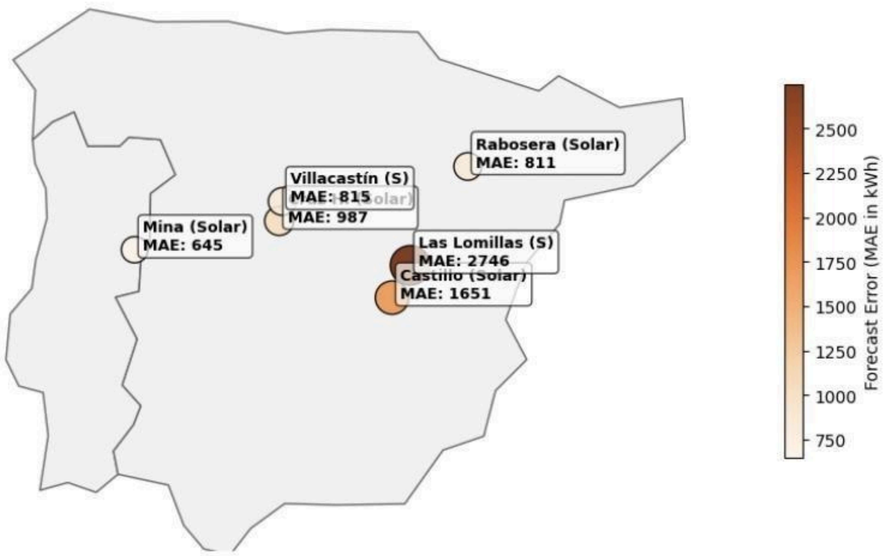

Figure 34 - Forecast Error Across Solar Parks in kWh, Base on the Predicted Data From a Commercial Model

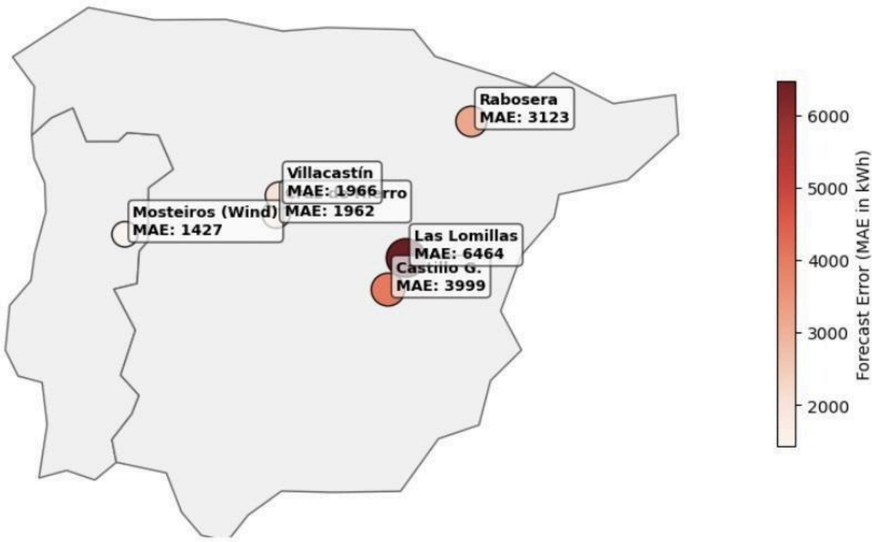

Figure 35 - Forecast Error Across Wind Parks in kWh, Base on the Predicted Data From a Commercial Model

Figure 34 and Figure 35 illustrate the spatial variability of forecasting errors across the analyzed plants. For solar generation (Figure 34) MAE values range from approximately 645 kWh to 2 746 kWh, while for wind generation Figure 35 errors range from about 1 427 kWh to 4 664 kWh.

This variability reflects plant-specific characteristics of renewable generation, such as local irradiance conditions for solar plants and exposure, terrain and prevailing wind regimes for wind farms. Similar spatial heterogeneity in forecast performance has been documented in previous studies and is a known feature of renewable energy forecasting under operational conditions.

Rather than aiming to equalize performance at individual sites, the joint forecasting formulation modifies how location-dependent uncertainties combine at the system level, particularly in contexts where solar–wind complementarity is present (Couto & Estanqueiro, 2021; Yang et al., 2012;).

To ensure that the observed system-level gains are not driven by underperforming component models, the proposed approach is positioned relative to a broader set of forecasting methods reported in the literature.

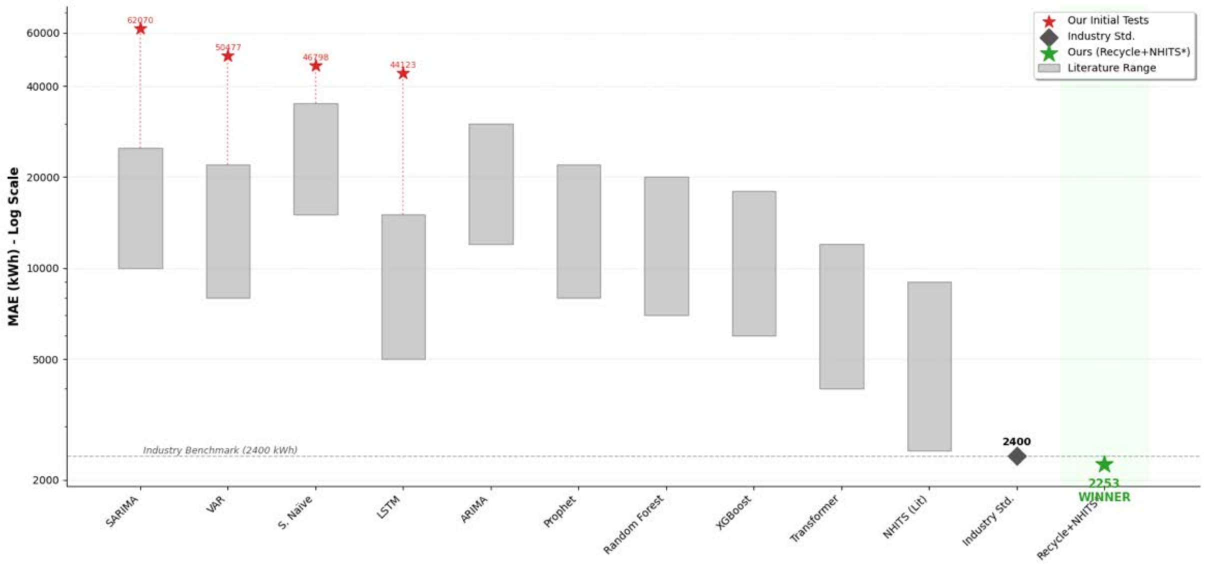

Figure 36 - Comparison of Forecasting Model Performance for Wind Generation

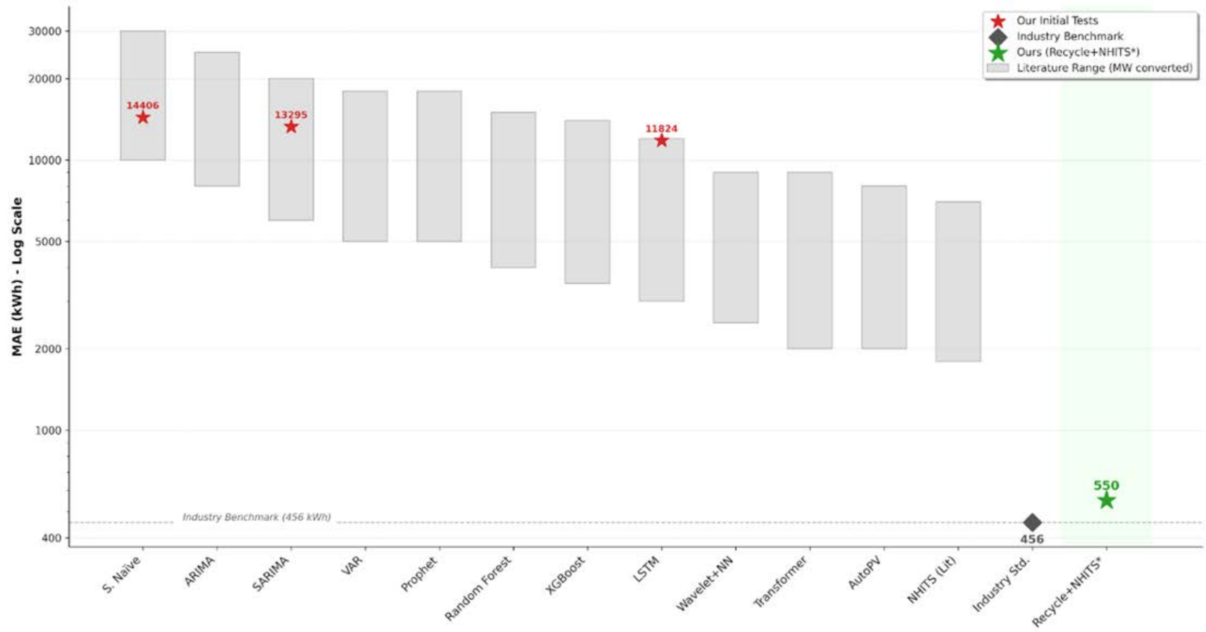

Figure 37 - Comparison of Forecasting Model Performance for Solar Generation

Figure 36 and Figure 37 compare the component-level Mean Absolute Error (MAE) obtained in this study with error ranges reported for comparable forecasting approaches in previous work. The grey shaded bars represent intervals of MAE values compiled from the literature for each model family, including classical statistical models, learning-based approaches, and structured deep learning architectures. For classical time-series formulations, such as persistence-based methods and ARIMA-type models, the reported ranges are derived from comparative studies on renewable energy forecasting (Hong et al., 2020; Pinson, 2013; Soman et al., 2010;). For learning-based models, including recurrent neural networks and LSTM architectures, the reference ranges reflect the variability documented across datasets and forecasting horizons in benchmarking and review studies (Devaraj et al., 2021; Kong et al., 2019; Maragkos & Refanidis, 2025). Structured deep learning models are positioned using results reported in large-scale forecasting competitions and recent evaluations of hierarchical architectures (Benidis et al., 2022; Challu, et al., 2022).

The red stars indicate the component-level forecasting performance obtained in this study across five independent training runs, conducted to account for the stochastic nature of learning-based models. The green markers denote the Industry Standard, corresponding to the operational forecasts provided by an external EDP forecasting provider to the utility for market participation and system operation. The comparison shows that the proposed models achieve MAE values that fall within the ranges reported in the literature and are, in several cases, comparable to operational forecasting performance. This positioning confirms that subsequent reductions in system-level uncertainty do not arise from weak component forecasts, but from the formulation of the forecasting problem itself.

When considered jointly, the results presented in this chapter indicate that improvements in forecasting performance are not driven solely by increasing model complexity, but by how well model structure aligns with the temporal and stochastic characteristics of renewable generation. While advanced architectures are necessary to capture non-stationary and multi-scale dynamics, their main contribution lies in how uncertainty and interaction effects are represented across technologies, rather than in altering the intrinsic predictability of individual resources. Within this framework, NHITS is particularly well suited to the application considered. Its hierarchical, multi-resolution architecture enables the coherent representation of short-term variability, diurnal cycles, and longer-term temporal patterns, while maintaining competitive component-level performance under operational conditions. When embedded in the proposed forecasting formulation, NHITS contributes to reshaping the distribution and persistence of forecast errors at the point of interconnection, providing a clear conceptual bridge to the hybrid forecasting results analyzed in the following section, where uncertainty reduction and operational implications are examined explicitly.

5. CONCLUSIONS

This dissertation began with a fundamental question: To what extent does joint forecasting of solar and wind generation in hybrid power plants at the point of interconnection reduce predictive uncertainty compared to isolated forecasting in the Iberian Peninsula?

The results indicate that the formulation of the problem at the point of interconnection plays a decisive role in shaping predictive performance. When solar and wind generation are forecast jointly rather than independently, aggregate uncertainty is systematically reduced (Cerqueira, et al., 2020; Hyndman & Athanasopoulos, 2021). For the hybrid configuration analyzed, joint forecasting achieved a reduction of 20.54% in Mean Absolute Error (MAE) relative to isolated technology forecasts (Hong et al., 2020; Hyndman, & Koehler, 2006). This improvement does not result from artificially smoothing variability. Instead, it emerges from how the stochastic behavior of solar and wind generation interacts when evaluated at the system boundary. The magnitude and consistency of observed uncertainty reduction exceeded initial expectations, particularly given the limited feature space and absence of detailed operational inputs.

Two complementary mechanisms explain this outcome. The first is production complementarity, where solar irradiance-driven cycles and wind atmospheric regimes partially offset one another at the system boundary (Yang et al., 2012). The second outcome is error complementarity, where forecast deviations in one technology compensate deviations in the other, reducing aggregate uncertainty during extreme imbalance events (OMIE, 2024; Pinson, 2013).

These findings demonstrate that the forecasting gains in hybrid systems cannot be attributed exclusively to increasing model complexity. While advanced architectures are required to represent non-stationary and multi-scale temporal dynamics, their effectiveness is contingent upon alignment between model structure, signal preparation, and the stochastic coupling across technologies. Complexity alone does not generate value, structural coherence does.

Among the evaluated models, NHITS achieved consistent results across technologies, forecasting horizons, and locations. Its hierarchical and multi-resolution design reflects the multi-scale temporal characteristics of renewable generation time series, including short-term variability, diurnal cycles, and seasonal dynamics (Challu, et al., 2022; Benidis et al., 2022). Nevertheless, the relative performance gains associated with joint forecasting remained robust across hyperparameter configurations, reinforcing that the principal contribution of this work lies in system-level problem reformulation, not in extreme hyperparameter optimisation (Benidis et al., 2022). The robustness of the

observed improvements across forecasting horizons and validation splits further indicates that gains are structural and not incidental.
Pre-modelling strategies also contributed to stability and interpretability. Residual decomposition through the ReCycle approach reduced structural noise and isolated dominant cyclic components prior to model training (Weyrauch et al., 2024). This confirms that forecasting performance depends not only on model architecture, but also on how the signal is structured and prepared before training.

Spatial heterogeneity provides additional context. Forecast accuracy differs across locations due to resource regimes, terrain exposure, and meteorological conditions (Hong et al., 2020; Pinson, 2013). Joint forecasting does not eliminate this heterogeneity instead, it modifies how local uncertainties interact at the system boundary. Sites with stronger temporal complementarity between solar and wind generation exhibit larger improvements, while sites with weaker coupling still benefit from consistent, though more moderate, reductions in aggregate error.

The primary contribution of this dissertation is therefore both empirical and conceptual, demonstrating that variability in hybrid systems is structurally emergent and not solely driven by model specification. The results show that hybrid forecasting must be framed as a system-level problem, where predictive behavior arises from technological interaction at the point of interconnection instead of isolated component modelling. In hybrid renewable systems, forecasting performance depends less on algorithmic sophistication alone and more on how variability is defined, structure and aggregated at system boundary.

5.1 LIMITATIONS

The empirical analysis is geographically concentrated in the Iberian Peninsula and primarily centered on a hybrid plant with complete and consistent operational data. Renewable generation patterns, complementarity effects, and forecast error structures are inherently location-dependent (Couto & Estanqueiro, 2021; Hong et al., 2020; Pinson, 2013). Differences in resource variability and market rules limit the direct transferability of the quantified uncertainty reduction to other regions. The empirical evidence should therefore be interpreted as region-specific rather than universally transferable. The analysis is therefore geographically biased toward Iberian climatic and market conditions.

Benchmarking relied on literature-reported state-of-the-art ranges and forecasts from an independent industrial provider. This provides a robust positioning of the obtained results within known performance intervals (Benidis et al., 2022; Devaraj et al., 2021; Maragkos & Refanidis, 2025). No external architectures were reimplemented under identical data conditions. As model comparisons are sensitive to dataset characteristics and experimental design (Bergmeir, & Benítez, 2012; Cerqueira, et al., 2020), results should be interpreted as contextual rather than strictly comparative.

The evaluation framework is deterministic and restricted to point-forecast metrics (MAE and RMSE) (Hyndman, & Koehler, 2006). These metrics do not characterize predictive distributions or tail risk, which are operationally relevant in high-renewable systems (Hong et al., 2020; Pinson, 2013). Probabilistic evaluation using scoring rules such as CRPS (Gneiting, & Raftery, 2007) was outside the scope of this study.

Data constraints were significant. Models relied exclusively on operational production and forecast time series. Installed capacity and dynamic available capacity at each plant were not fully observable, preventing the separation of meteorological forecast uncertainty from deviations driven by curtailment or operational constraints.

Detailed plant-level availability data were not accessible throughout the observed period, Therefore, the analysis assumed that installed capacity was continuously available 100% over the course of the dissertation. In practice, generation capacity may not have always operated at its maximum level due to scheduled maintenance, forced outages or other operational constraints. This simplification may have introduced some degree of bias into the evaluation, as it was not possible to adjust observed production relative to the effectively available capacity at each point in time. Part of the forecast error may therefore reflect variations in operational availability rather than exclusively meteorological uncertainty, implicitly incorporating real-world industrial operating conditions into the assessment framework.

While the limited availability of exogenous features constrained modelling richness, it also strengthens the robustness of the findings. The observed system-level gains emerge under data-constrained industrial conditions, suggesting that hybrid reformulation retains value even in feature-scarce environment.

Hyperparameter optimization was performed manually. Deep learning architectures, including NHITS, do not yet have universally transferable hyperparameter standards across domains (Benidis et al., 2022). Although multiple training runs were performed and consistent procedures were adopted, the optimization process remains partly empirical and sensitive to design choices. Additionally, forecasting results depend on the adopted experimental design, including window selection, horizon definition, and validation strategy (Bergmeir, & Benítez, 2012; Cerqueira, et al., 2020). What is identified as “best performing” is inherently relative to the adopted training and evaluation framework.

These limitations define the scope of the study. The results demonstrate system-level uncertainty reduction can be achieved under geographically constrained, deterministic, and data-limited industrial conditions, but broader generalization requires caution. For confidentiality reasons, plant identities and certain details were anonymized. This does not affect the statistical results, it limits direct plant level comparability and full external replication.

5.2 FUTURE WORK

Building upon the system-level perspective developed in this dissertation, several research directions emerge. First, data standardisation in hybrid solar-wind forecasting should be prioritised. The absence of harmonised benchmark datasets with consistent variable definitions, temporal alignment procedures and evaluation protocols continues to limit reproducibility and cross-study comparability. Establishing structured and openly accessible repositories for hybrid plants would significantly strengthen methodological transparency and accelerate cumulative progress in the field.

Further performance gains may arise from systematic refinement of the forecasting pipeline. Deep learning architectures remain sensitive to hyperparameter configuration, and energy forecasting lacks universally transferable optimization standards. Automated search strategies could improve robustness and reduce empirical tuning. Integrating richer inputs including high-resolution meteorological variables, curtailment signals, plant-level availability indicators, effective capacity constraints and technical operating limits, would allow models to better represent operational nonlinearities and distinguish meteorological variability from capacity-driven effects.

Advancing beyond deterministic point forecasts is essential for operational relevance. Probabilistic frameworks, quantile regression, distributional neural networks and adversarial learning approaches can explicitly model predictive uncertainty and tail risk. This is particularly critical in high-renewable systems, where extreme deviations drive imbalance exposure and market risk (Pinson, 2013; OMIE, 2024).

Future development should also deepen the integration between data-driven learning and physical structure. Expanding historical production depth, incorporating exogenous atmospheric and system-level predictors, and embedding simplified representations of irradiance physics, turbine behavior and saturation effects would move forecasting toward hybrid symbolic-statistical paradigms. In such frameworks, domain knowledge and machine learning jointly constrain the learning process rather than acting as separate modelling philosophies.

The role of forecasting horizon also deserves further investigation. While this dissertation focuses on short-term operational horizons, the degree of solar–wind complementarity may evolve across longer lead times. Understanding how interaction effects and uncertainty propagation change across temporal scales would refine the system-level perspective introduced in this work.

From an operational standpoint, integrating joint solar–wind forecasting with battery energy storage systems (BESS) represents a natural extension of the proposed framework. Combining forecasting, dispatch optimization and storage control within a unified architecture would enable temporal complementarity to be actively managed at the point of interconnection, transforming it from a passive structural property into an operationally exploitable mechanism.

At present, large-scale BESS deployment remains economically constrained in many hybrid plants, where curtailment through temporary turbine shutdown or inverter limitation is often preferred due to capital costs. Despite current economic constraints, storage integration may unlock additional system-level value by reducing renewable curtailment and enabling active management of temporal complementarity.

Economic valuation within the MIBEL market remains an open avenue. Assessing how the observed 20.54% reduction in uncertainty translates into imbalance cost mitigation would provide a direct economic interpretation of the results. This requires modelling imbalance pricing structures, bidding strategies and contractual exposure. Given the asymmetric nature of imbalance costs in sequential electricity markets, such valuation could reveal the economic relevance of extreme-error mitigation beyond average performance improvements.

Finally, extending the framework to spatial aggregation and portfolio-level forecasting would clarify how uncertainty propagates across geographically distributed hybrid assets and how hybridization interacts with geographic diversification. Collectively, these directions extend the system-level paradigm introduced in this dissertation beyond forecasting accuracy toward integrated uncertainty management in hybrid renewable systems, reinforcing the evolving role of hybrid power plants in supporting grid stability and operational resilience, particularly in weak-grid contexts.

BIBLIOGRAPHICAL REFERENCES

Almeida, L. M. B. (2026). *Forecasting hybrid power plant generation* [Computer software]. GitHub. https://github.com/leonormabalmeida-ai/Forecasting_Hybrid_Power_Plant_Generation

Bandara, K., Bergmeir, C., & Smyl, S. (2020). *Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach*. *Expert Systems with Applications, 140*, 112896. https://doi.org/10.1016/j.eswa.2019.112896

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., Gasthaus, J., & Januschowski, T. (2022). Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys, 55*(6), 1–36. https://doi.org/10.1145/3533382

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192-213. https://doi.org/10.1016/j.ins.2011.12.028

Bett, P. E., & Thornton, H. E. (2016). The climatological relationships between wind and solar energy supply in Britain. *Renewable Energy, 87*, 96–110. https://doi.org/10.1016/j.renene.2015.10.006

Caruana, R. (1997). Multitask learning. *Machine Learning, 28*(1), 41–75. https://doi.org/10.1023/A:1007379606734

Cerqueira, V., Torgo, L., & Soares, C. (2020). *Evaluating time series forecasting models: An empirical study on performance estimation methods*. Machine Learning, *109*, 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Challu, C., Olivares, K., Lobo, J. L., & Garza, J. (2022).N-HiTS: Neural hierarchical interpolation for time series forecasting. arXiv. https://arxiv.org/abs/2201.12886

Chinaris, P. P., Psarros, G. N., & Papathanassiou, S. A. (2024). *Hybridization of wind farms with co-located PV and storage installations*. *Renewable Energy, 240*, 122057. https://doi.org/10.1016/j.renene.2024.122057

Couto, A., & Estanqueiro, A. (2021). *Assessment of wind and solar PV local complementarity for the hybridization of the wind power plants installed in Portugal*. *Journal of Cleaner Production, 319*, 128728. https://doi.org/10.1016/j.jclepro.2021.128728

Das, K., Sharma, H., Mukherjee, S., & Glavaski, S. (2025). Research challenges and opportunities of utility-scale hybrid power plants. *WIREs Energy and Environment*, e70001. https://doi.org/10.1002/wene.70001

Deshmukh, M. K., & Deshmukh, S. S. (2008). *Modeling of hybrid renewable energy systems*. *Renewable and Sustainable Energy Reviews, 12*(1), 235–249. https://doi.org/10.1016/j.rser.2006.07.011

Devaraj, J., Elavarasan, R. M., Shafiullah, G., Jamal, T., & Khan, I. (2021). A holistic review on energy forecasting using big data and deep learning models. *International Journal of Energy Research, 45*(9), 13489–13530. https://doi.org/10.1002/er.6679

Dufo-López, R., Bernal-Agustín, J. L., & Mendoza, F. (2009). *Design and economical analysis of hybrid PV-wind systems connected to the grid for the intermittent production of hydrogen*. *Energy Policy, 37*, 3082–3095. https://doi.org/10.1016/j.enpol.2009.03.059

Frew, B., Sergi, B., Denholm, P., Cole, W., Gates, N., Levie, D., & Margolis, R. (2021). *The curtailment paradox in the transition to high solar power systems*. Joule, 5(5), 1143–1167. https://doi.org/10.1016/j.joule.2021.03.021

Gneiting, T., & Raftery, A. E. (2007). *Strictly proper scoring rules, prediction, and estimation*. *Journal of the American Statistical Association*, 102(477), 359–378. https://doi.org/10.1198/016214506000001437

Gonçalves, C., Cavalcante, L., Brito, M., & Pinson, P. (2021). *Forecasting conditional extreme quantiles for wind energy*. *Electric Power Systems Research, 190*, 106636. https://doi.org/10.1016/j.epsr.2020.106636

Hassan, Q., Algburi, S., Sameen, A. Z., Salman, H. M., & Jaszczur, M. (2023). *A review of hybrid renewable energy systems: Solar and wind powered solution - Challenges, opportunities, and policy implications.* Results in Engineering, 20, 101621. https://doi.org/10.1016/j.rineng.2023.101621

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy, 7*, 376–388. https://doi.org/10.1109/OAJPE.2020.3029979

Huang, Q., Shi, Y., Wang, Y., Lu, L., & Cui, Y. (2015). *Multi-turbine wind-solar hybrid system*. *Renewable Energy, 76*, 401–407. https://doi.org/10.1016/j.renene.2014.11.060

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jornal de Negócios. (2023, January 12). *EDP liga à rede primeira central híbrida com energia solar e eólica na Península Ibérica*. https://www.jornaldenegocios.pt/empresas/energia/detalhe/edp-liga-a-rede-primeira-central-hibrida-com-energia-solar-e-eolica-na-peninsula-iberica

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid, 10*(1), 841–851. https://doi.org/10.1109/TSG.2017.2753802

Lund, H., Østergaard, P. A., Connolly, D., & Mathiesen, B. V. (2017). *Smart energy and smart energy systems*. *Energy, 137*, 556–565. https://doi.org/10.1016/j.energy.2017.05.123

Maragkos, N., & Refanidis, I. (2025). A comparative evaluation of time-series forecasting models for energy datasets. Computers, 14(7), 246. https://doi.org/10.3390/computers14070246

Marugán, A. P., Márquez, F. P. G., Pérez, J. M. P., & Ruiz-Hernández, D. (2018). A survey of artificial neural networks in wind energy systems. *Applied Energy, 228*, 1822-1836. https://www.sciencedirect.com/science/article/abs/pii/S0306261918311048

Nixtla. (2015). NeuralForecast [Computer Software]. GitHub. https://github.com/Nixtla/neuralforecast

OMIE. (2024). *Integrated report 2023*. https://www.omie.es/sites/default/files/2024-05/OMI_reporting_integrado_2023_EN.pdf

Pinson, P. (2013). *Wind energy: Forecasting challenges for its operational management*. arXiv. https://arxiv.org/abs/1312.6471

Pinson, P. (2006). *Estimation of the uncertainty in wind power forecasting* (Doctoral dissertation). École des Mines de Paris. https://pastel.hal.science/pastel-00002187v1/document

Polo, J., Bernardos, A., Fabregat-Santiago, C., et al. (2025). Hybrid wind–solar generation and analysis for the Iberian Peninsula: A case study. *Energies, 18*(15), 3966. https://doi.org/10.3390/en18153966

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv*. https://arxiv.org/abs/1706.05098

Shahnazian, F., Das, K., Yan, R., & Sørensen, P. (2024). *Aspects of relevance of hybrid power plants in control and stability of weak grids*. *Energies, 17*, 6353. https://doi.org/10.3390/en17246353

Shih, S.-Y., Sun, F.-K., & Lee, H.-Y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning, 108*(8–9), 1421–1441. https://doi.org/10.1007/s10994-019-05815-0

Silva-Rodriguez, J., Raffoul, E., & Li, X. (2024). LSTM-based net load forecasting for wind and solar power-equipped microgrids. *arXiv*. https://arxiv.org/abs/2408.00136

Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *Proceedings of the North American Power Symposium (NAPS 2010)*. https://doi.org/10.1109/NAPS.2010.5619586

Stefenon, S. F., Seman, L. O., Yamaguchi, C. K., Coelho, L. dos S., Mariani, V. C., Matos-Carvalho, J. P., et al. (2025). Neural hierarchical interpolation time series (NHITS) for reservoir level multi-horizon forecasting in hydroelectric power plants. *Energies, 18*(1), 93
https://ieeexplore.ieee.org/document/10938105

Taieb, S. B., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2011). *A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition*. arXiv. https://arxiv.org/abs/1108.3259

Taieb, S. B., & Hyndman, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting, 30*(2), 382–394. https://doi.org/10.1016/j.ijforecast.2013.07.005

Tina, G. M., & Gagliano, S. (2011). *Probabilistic modelling of hybrid solar/wind power system with solar tracking system*. *Renewable Energy, 36*, 1719–1727. https://doi.org/10.1016/j.renene.2010.12.001

Wang, J., Liu, F., Song, Y., & Li, J. (2021). A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy, 304*, 117766. https://doi.org/10.1016/j.apenergy.2021.117766

Weyrauch, A., Steens, T., Taubert, O., Hanke, B., Eqbal, A., Götz, E., Streit, A., Götz, M., & Debus, C. (2024). ReCycle: Fast and efficient long time series forecasting with residual cyclic transformers. *arXiv*. https://arxiv.org/abs/2405.03429

Xie, Y., Zhang, Z., Wang, S., Liu, J., & Guo, H. (2023). An overview of deterministic and probabilistic forecasting of wind energy. *iScience, 26*, 106060. https://www.sciencedirect.com/science/article/pii/S2589004222020776

Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T. C., & Coimbra, C. F. M. (2012). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy, 86*(2), 622–633. https://www.sciencedirect.com/science/article/abs/pii/S0038092X17310022

Yang, H., Lu, L., & Zhou, W. (2007). *A novel optimization sizing model for hybrid solar-wind power generation system*. *Solar Energy, 81*, 76–84. https://doi.org/10.1016/j.solener.2006.06.010

Zhang, Y., Wang, J., & Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews, 32*, 255–270. https://doi.org/10.1016/j.rser.2014.01.033

APPENDIX A

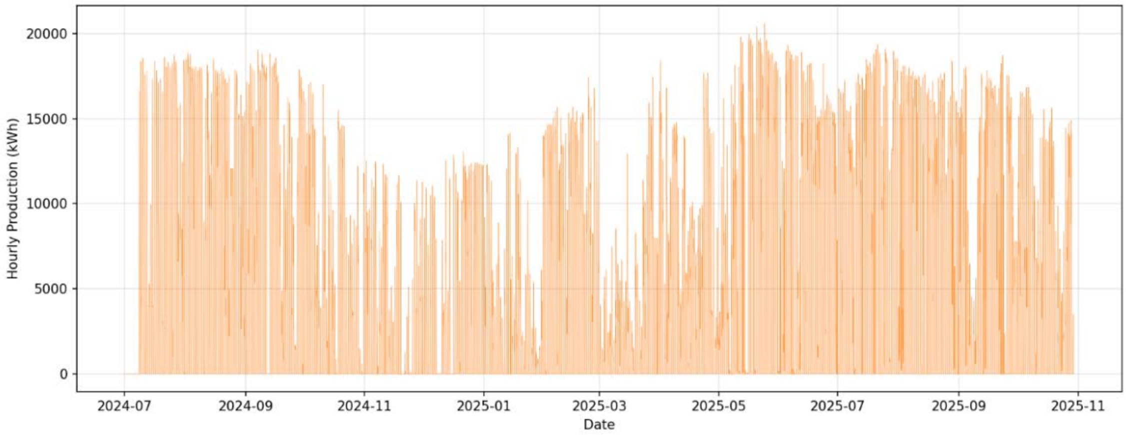

Figure A 1 - Hourly Solar Power Production (kWh) for Plant 2 Solar

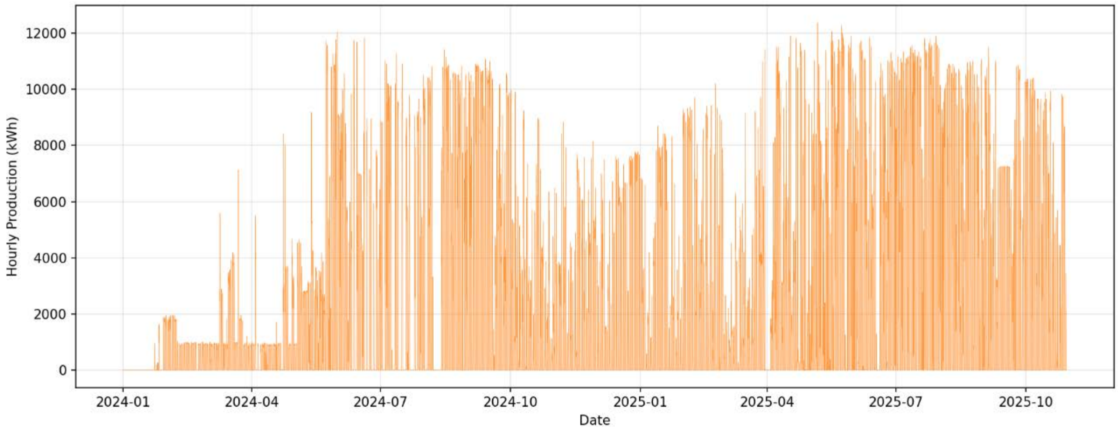

Figure A 2 - Hourly Solar Power Production (kWh) for Plant 3 Solar

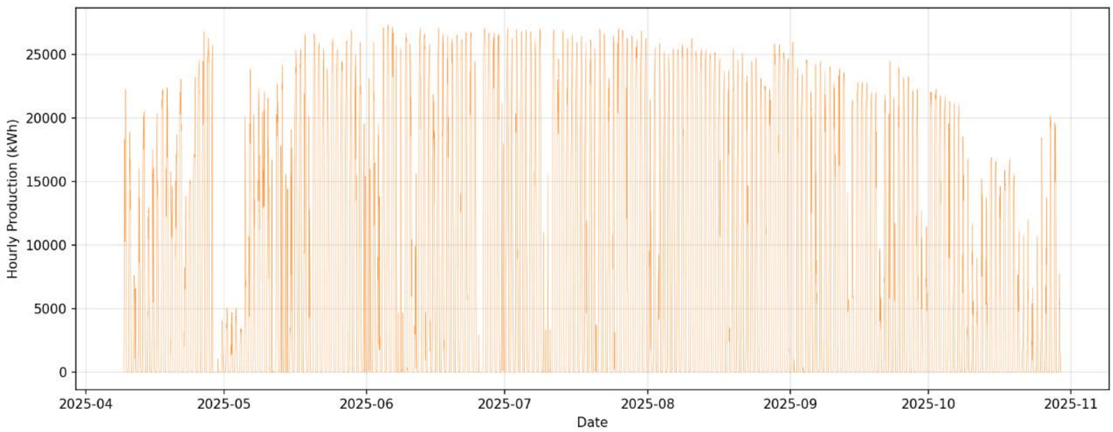

Figure A 3 - Hourly Solar Power Production (kWh) for Plant 4 Solar

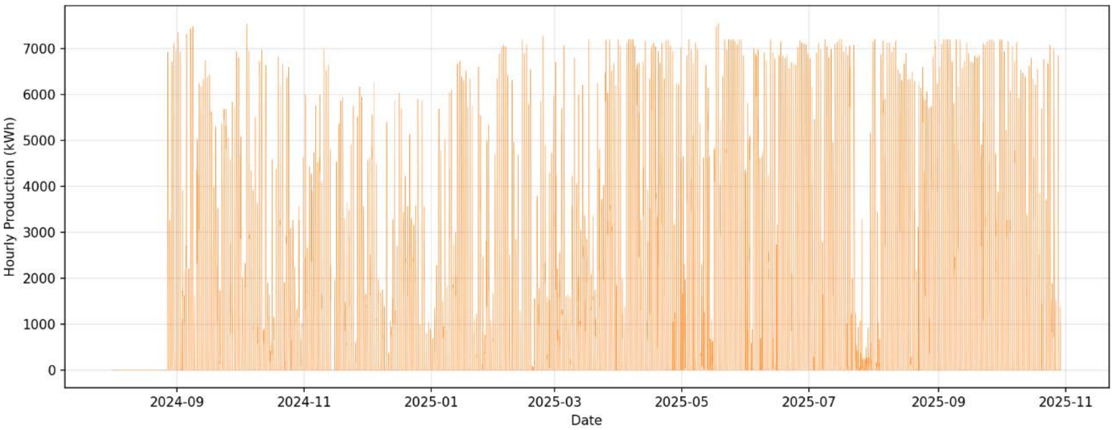

Figure A 4 - Hourly Solar Power Production (kWh) for Plant 5 Solar

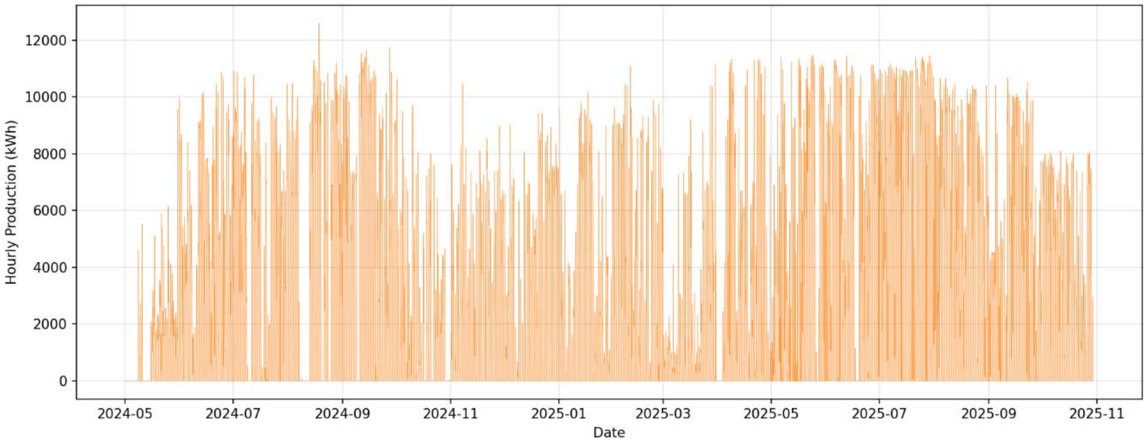

Figure A 5 - Hourly Solar Power Production (kWh) for Plant 6 Solar

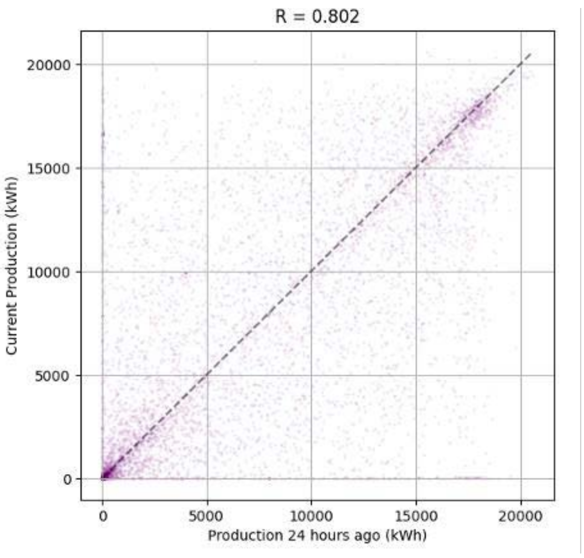

Figure A 6 - Lag-24h Autocorrelation of Solar Production (Plant 2 Solar)

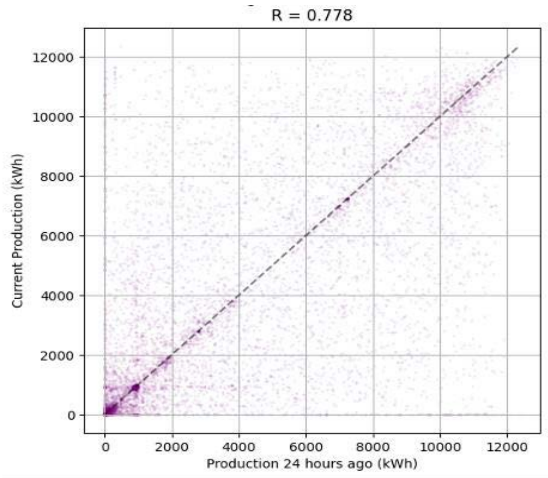

Figure A 7 - Autocorrelation (Lag 24h) for Solar Plant (Plant 3 Solar)

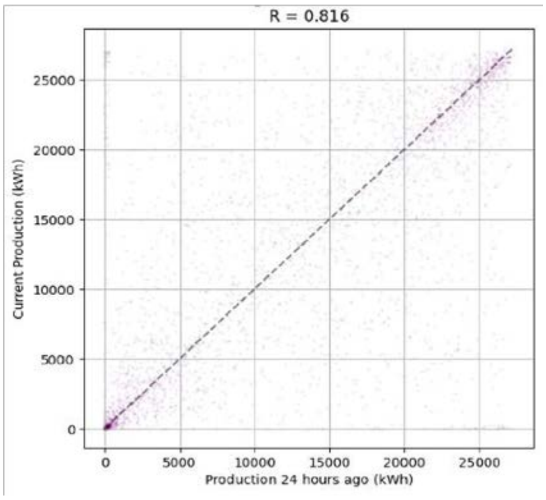

Figure A 8 - Lag-24h Autocorrelation of Solar Production (Plant 4 Solar)

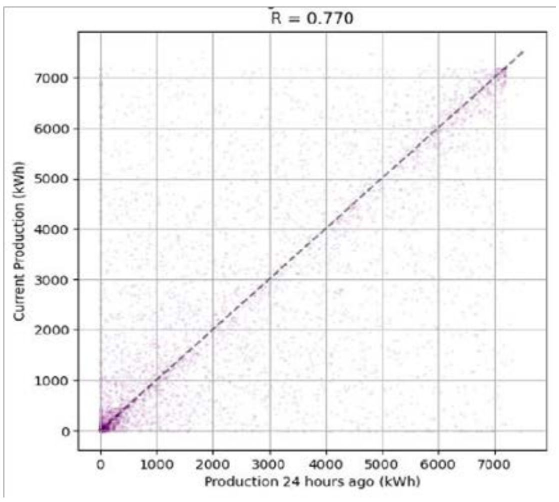

Figure A 9 - Lag-24h Autocorrelation of Solar Production (Plant 5 Solar)

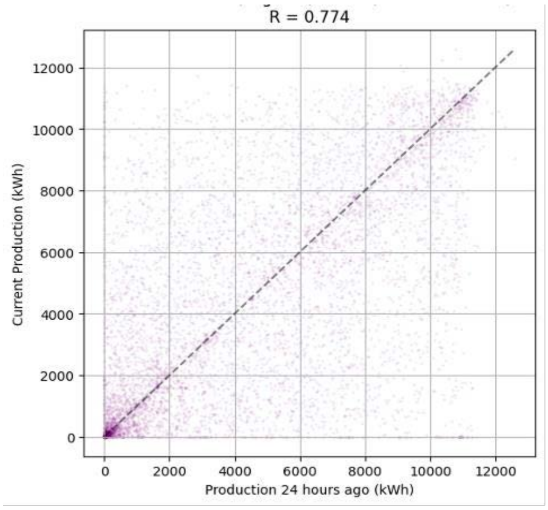

Figure A 10 - Lag-24h Autocorrelation of Solar Production (Plant 6 Solar)

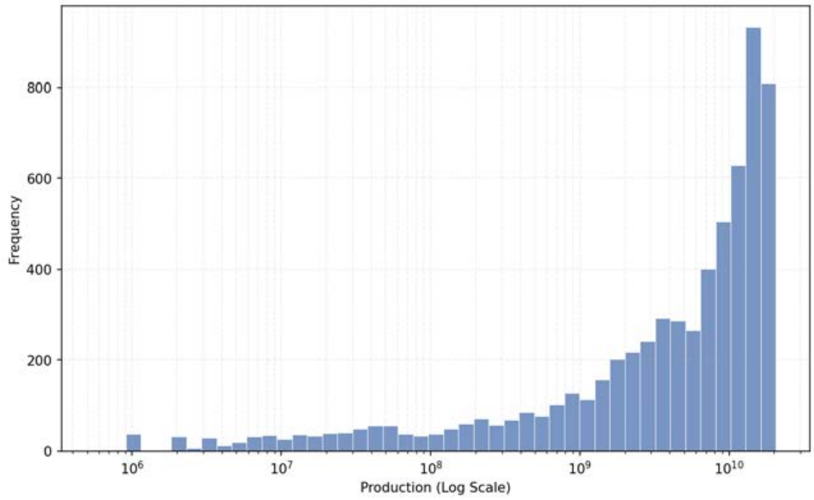

Figure A 11 - Hourly Distribution of Solar Generation, logarithmic scale on the x axis, base 10 (Plant 2 Solar)

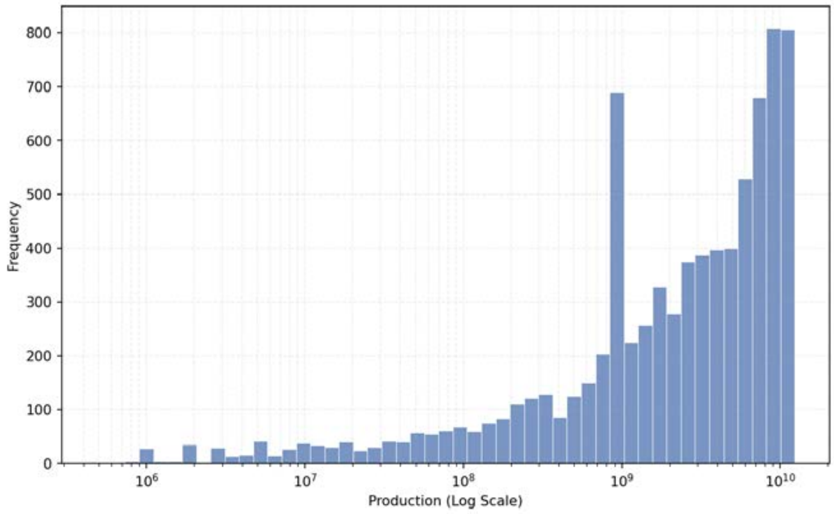

Figure A 12 - Hourly Distribution of Solar Generation, logarithmic scale on the x axis, base 10 (Plant 3 Solar)

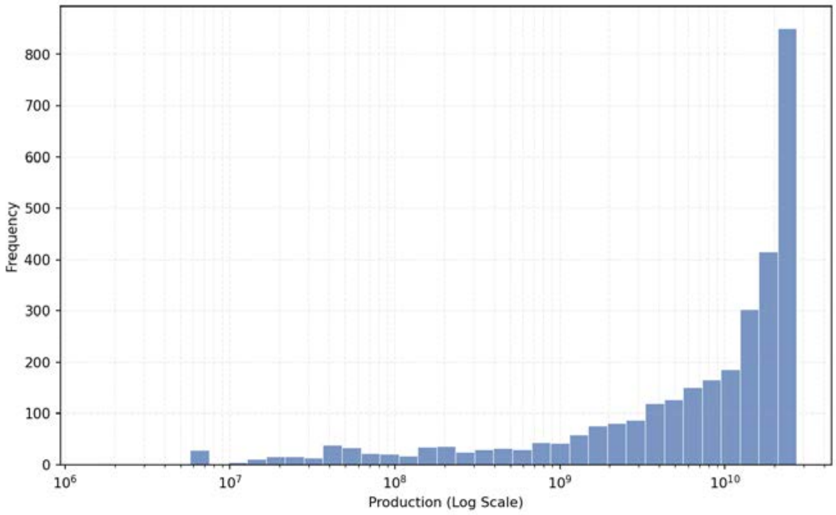

Figure A 13 - Hourly Distribution of Solar Generation, logarithmic scale on the x axis, base 10 (Plant 4 Solar)

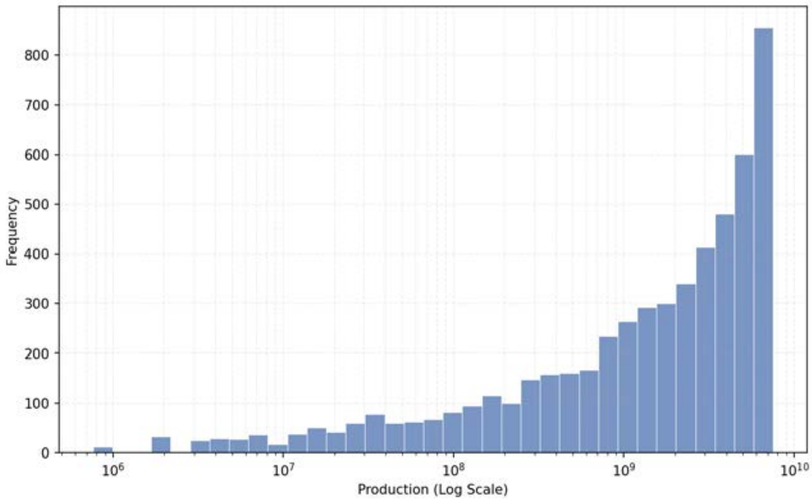

Figure A 14 - Hourly Distribution of Solar Generation, logarithmic scale on the x axis, base 10 (Plant 5 Solar)

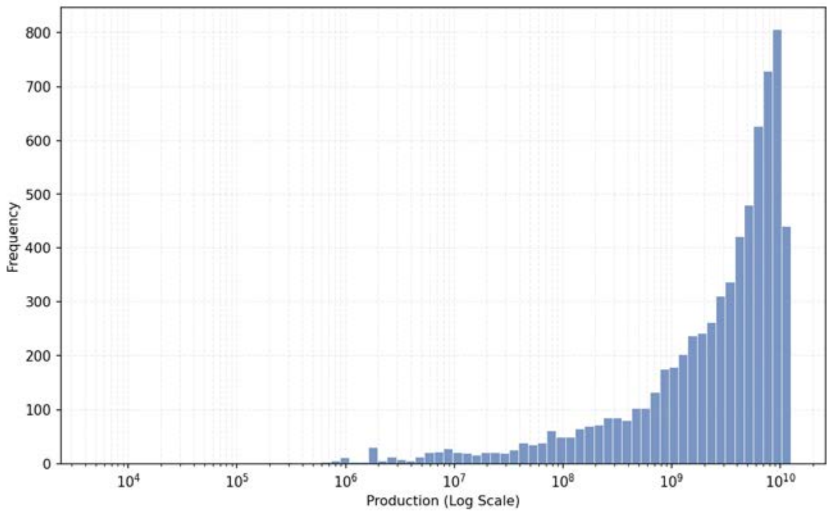

Figure A 15 - Hourly Distribution of Solar Generation, logarithmic scale on the x axis, base 10 (Plant 6 Solar)

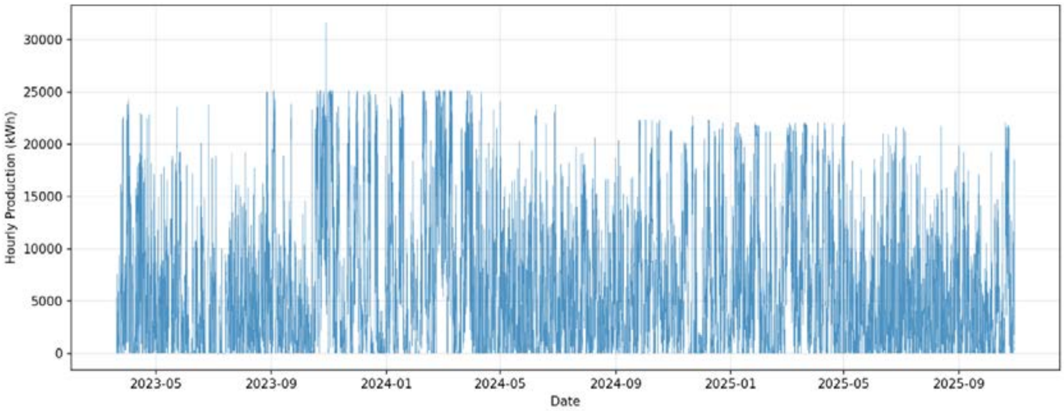

Figure A 16 - Hourly Solar Power Production (kWh) for Plant 2 Solar

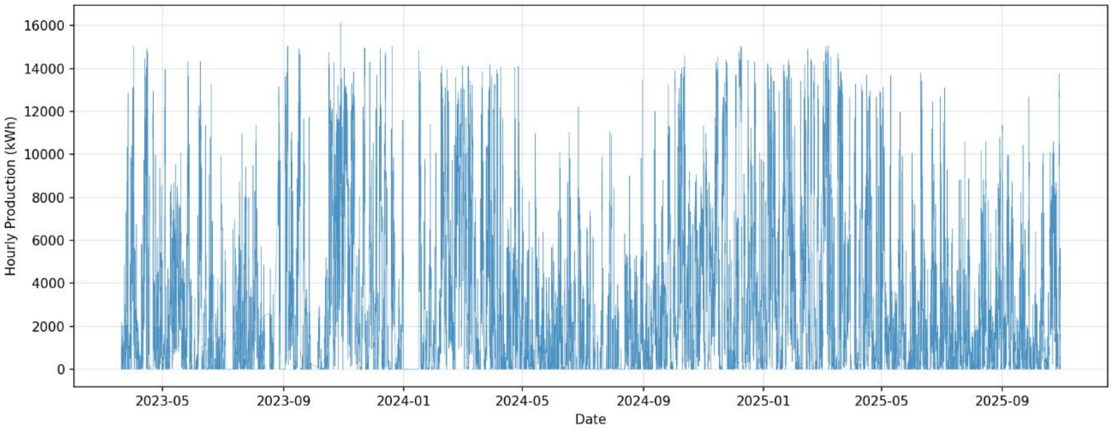

Figure A 17 - Hourly Solar Power Production (kWh) for Plant 3 Solar

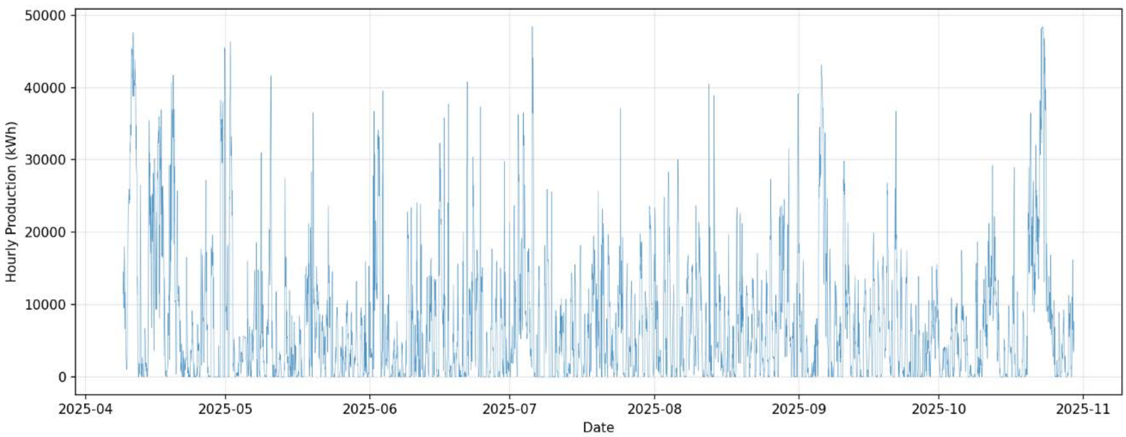

Figure A 18 - Hourly Solar Power Production (kWh) for Plant 4 Solar

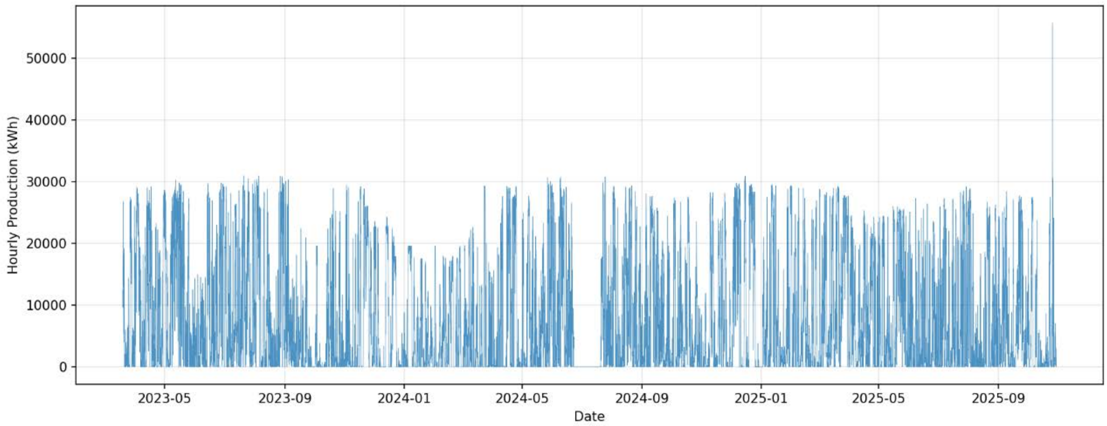

Figure A 19 - Hourly Solar Power Production (kWh) for Plant 5 Solar

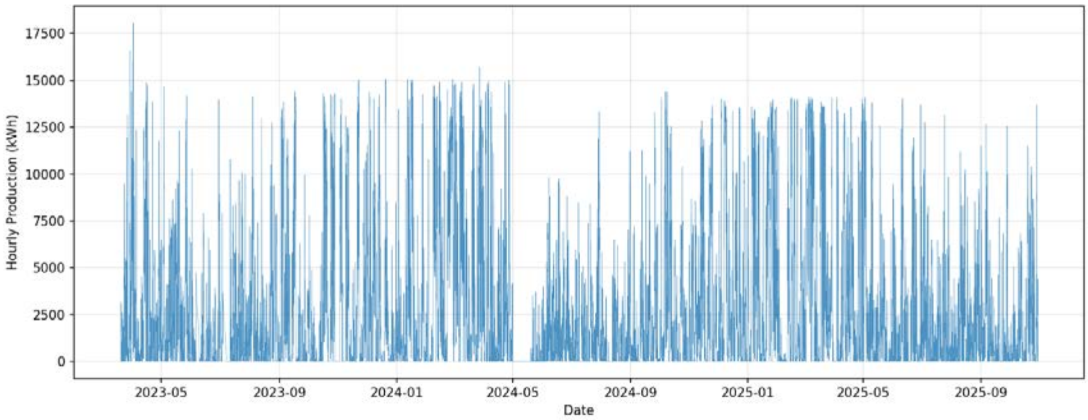

Figure A 20 - Hourly Solar Power Production (kWh) for Plant 6 Solar

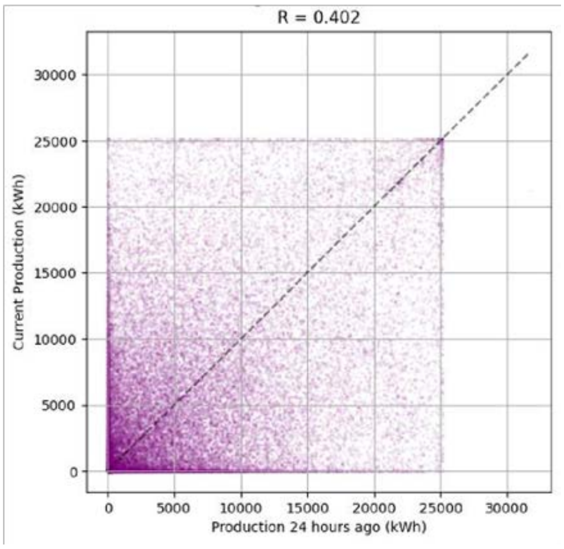

Figure A 21 - Lag-24h Autocorrelation of Wind Production (Plant 2 Wind)

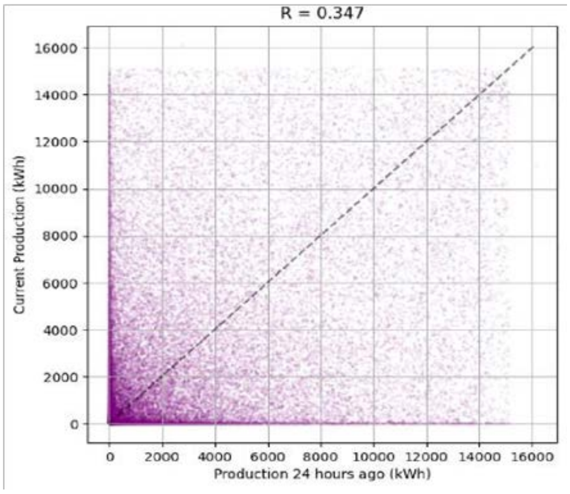

Figure A 22 - Lag-24h Autocorrelation of Wind Production (Plant 3 Wind)

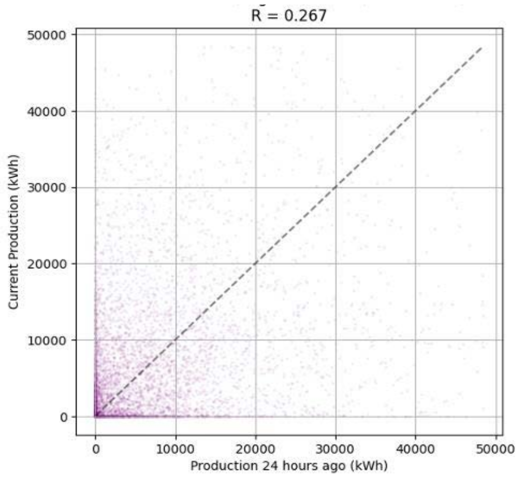

Figure A 23 - Lag-24h Autocorrelation of Wind Production (Plant 4 Wind)

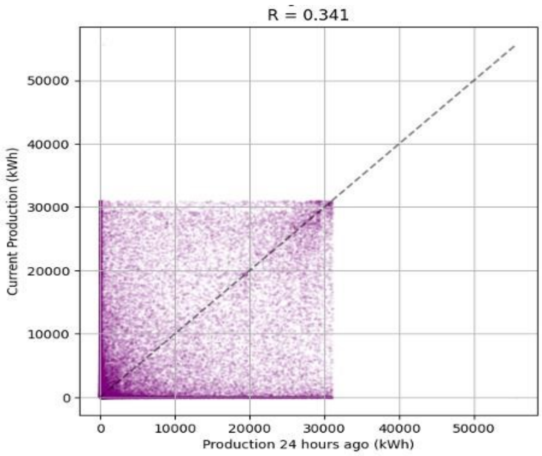

Figure A 24 - Lag-24h Autocorrelation of Wind Production (Plant 5 Wind)

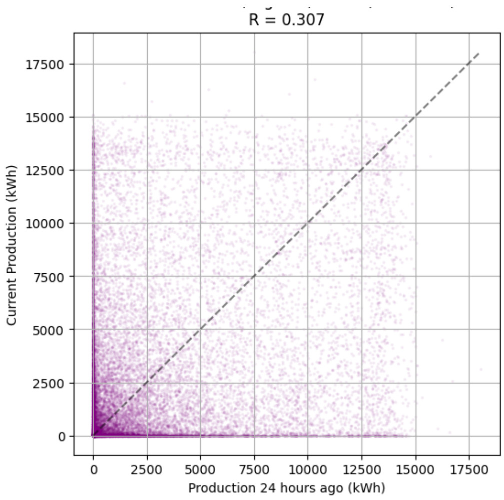

Figure A 25 - Lag-24h Autocorrelation of Wind Production (Plant 6 Wind)

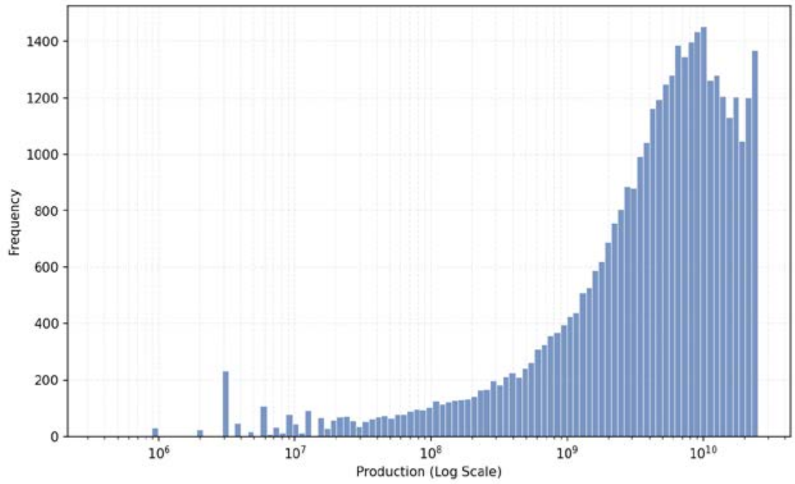

Figure A 26 - Hourly Distribution of Wind Generation, logarithmic scale on the x axis, base 10 (Plant 2 Wind)

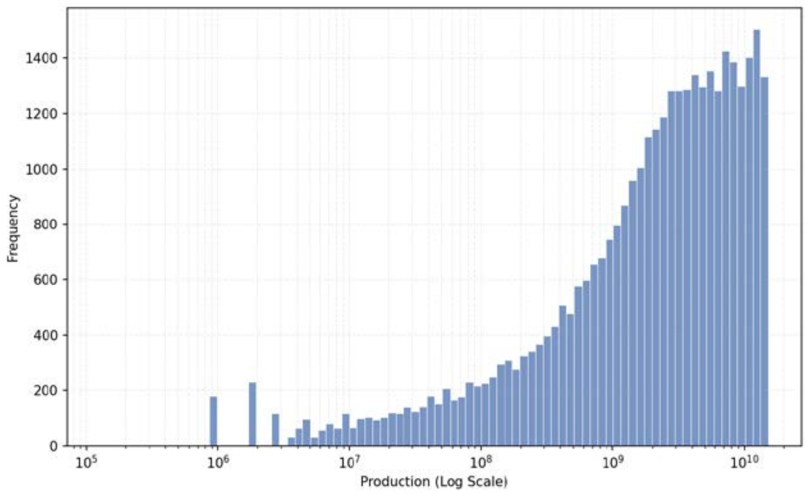

Figure A 27 - Hourly Distribution of Wind Generation, logarithmic scale on the x axis, base 10 (Plant 3 Wind)

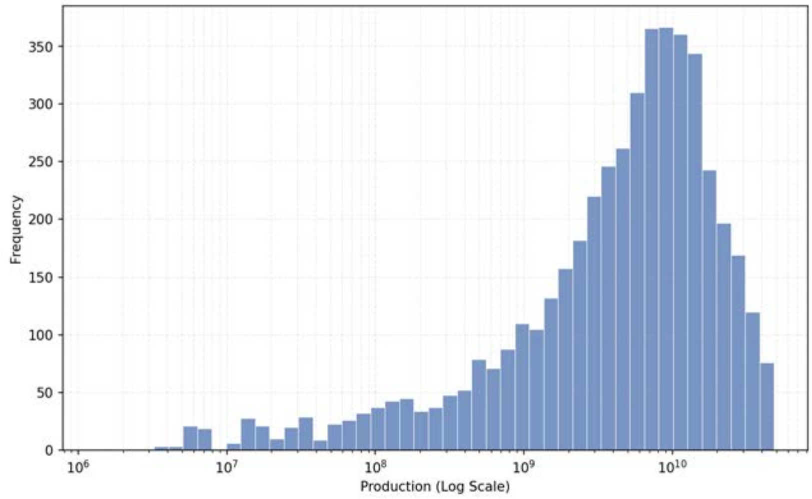

Figure A 28 - Hourly Distribution of Wind Generation, logarithmic scale on the x axis, base 10 (Plant 4 Wind)

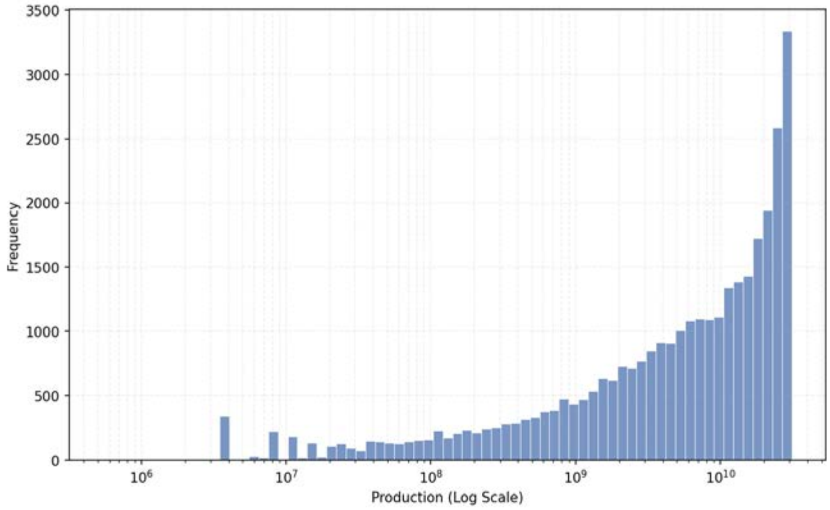

Figure A 29 - Hourly Distribution of Wind Generation, logarithmic scale on the x axis, base 10 (Plant 5 Wind)

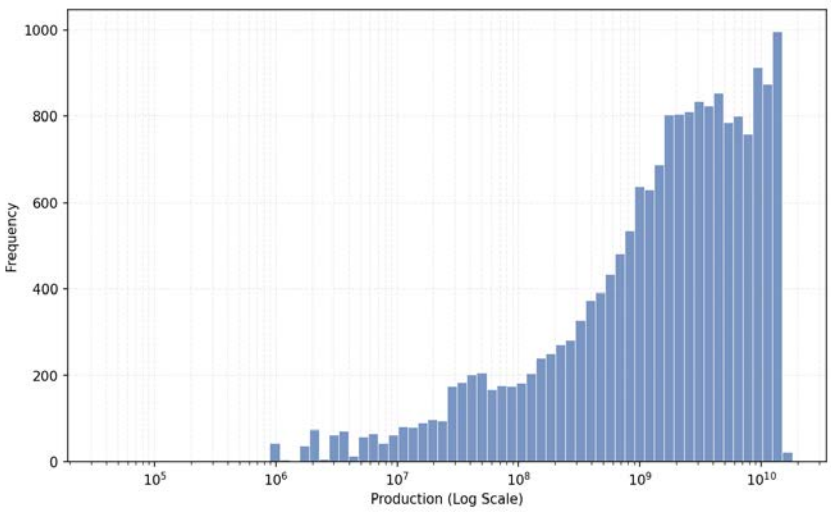

Figure A 30 - Hourly distribution of Wind Generation (Plant 6 Wind)

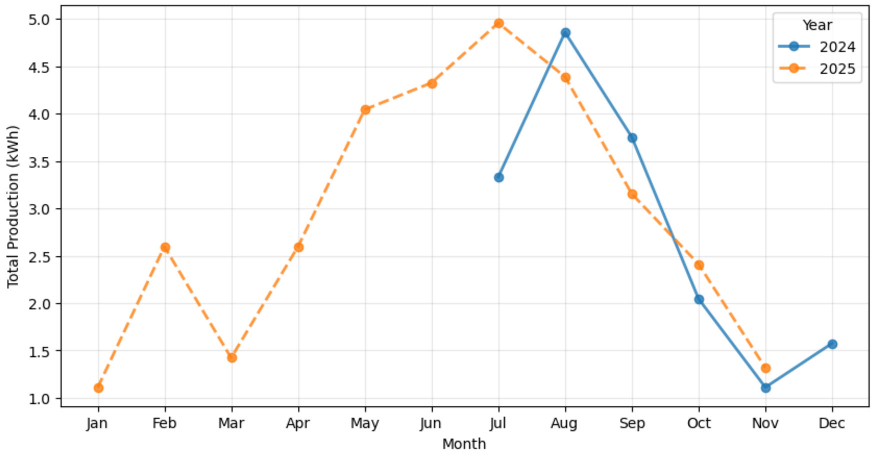

Figure A 31 - Year-over-year Solar Production Profiles (Plant 2 Solar)

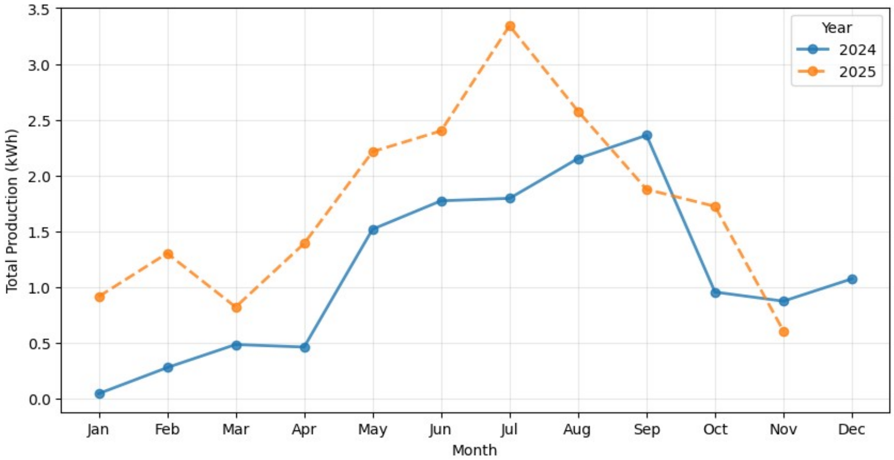

Figure A 32 - Year-over-year Solar Production Profiles (Plant 3 Solar)

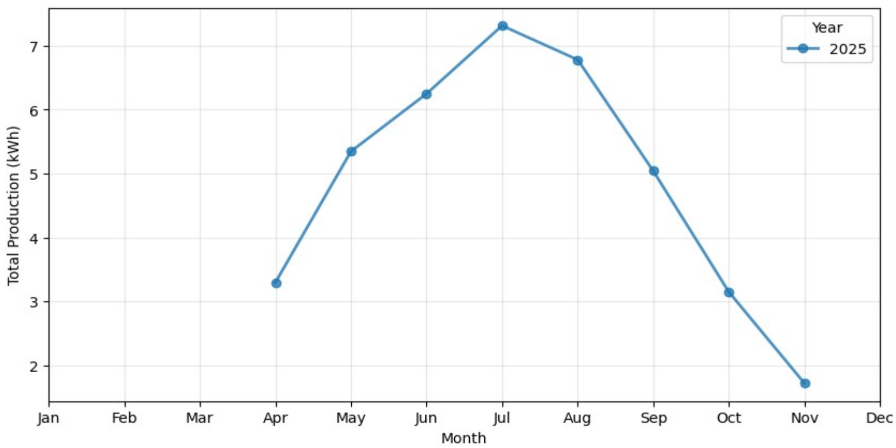

Figure A 33 - Year-over-year Solar Production Profiles (Plant 4 Solar)

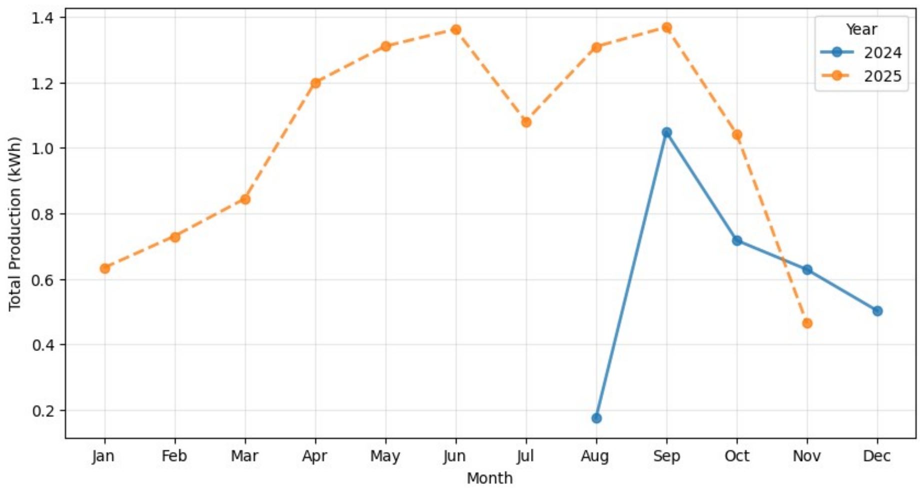

Figure A 34 - Year-over-year Solar Production Profiles (Plant 5 Solar)

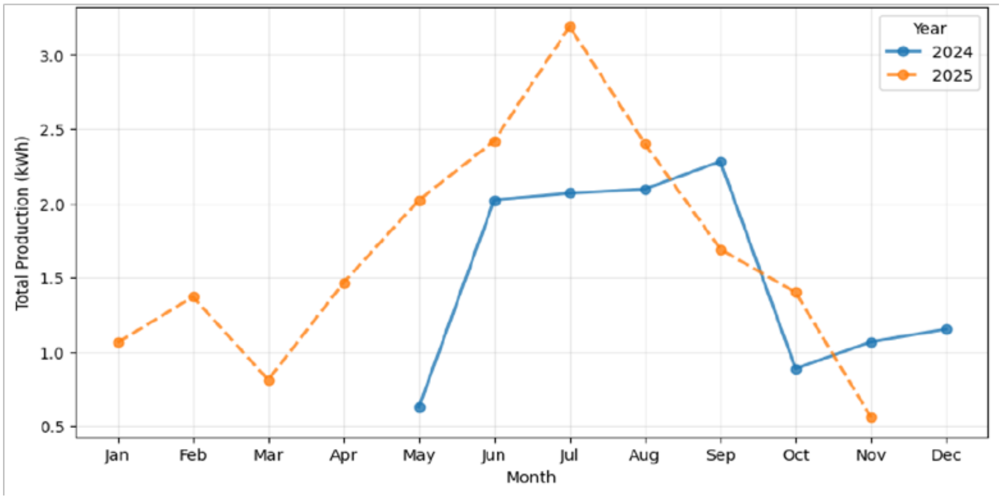

Figure A 35 - Year-over-year Solar Production Profiles (Plant 6 Solar)

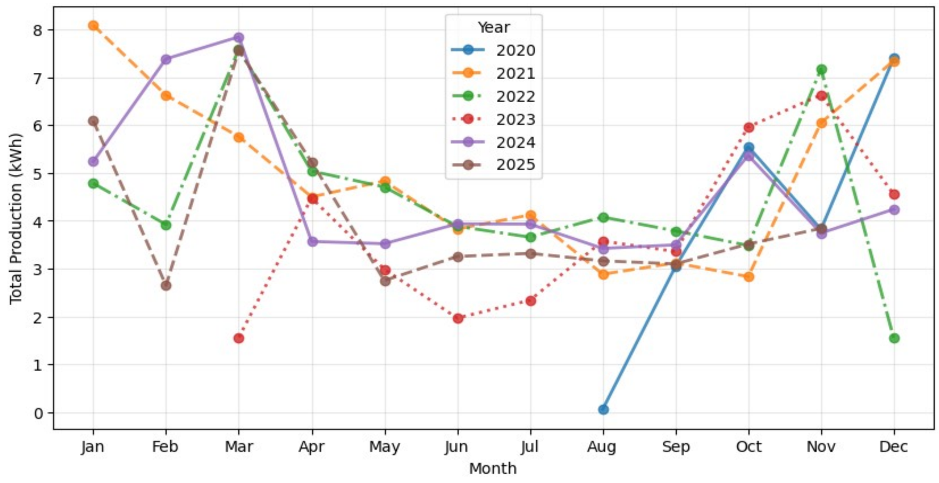

Figure A 36 - Year-over-year Wind Production Profiles (Plant 2 Wind)

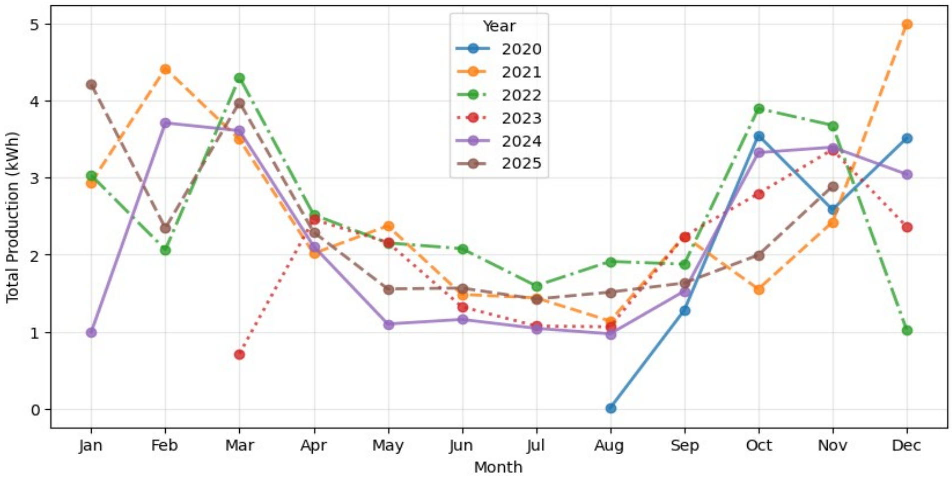

Figure A 37 - Year-over-year Wind Production Profiles (Plant 3 Wind)

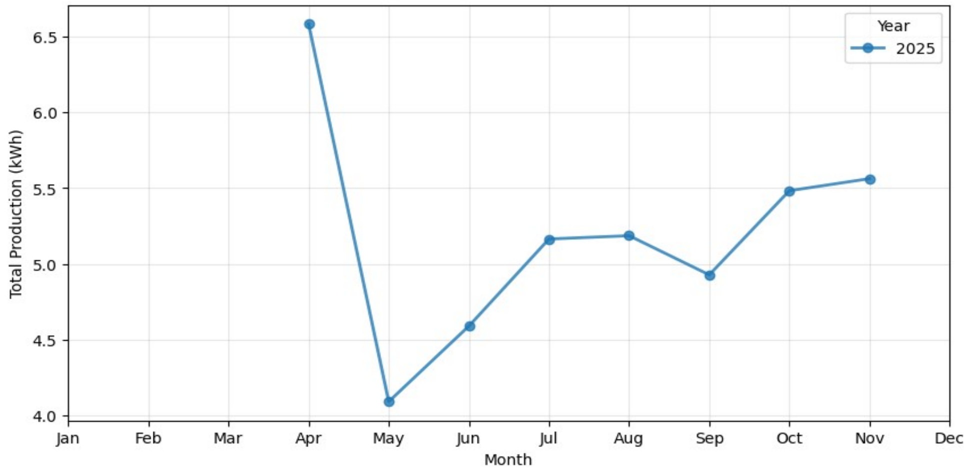

Figure A 38 - Year-over-year Wind Production Profiles (Plant 4 Wind)

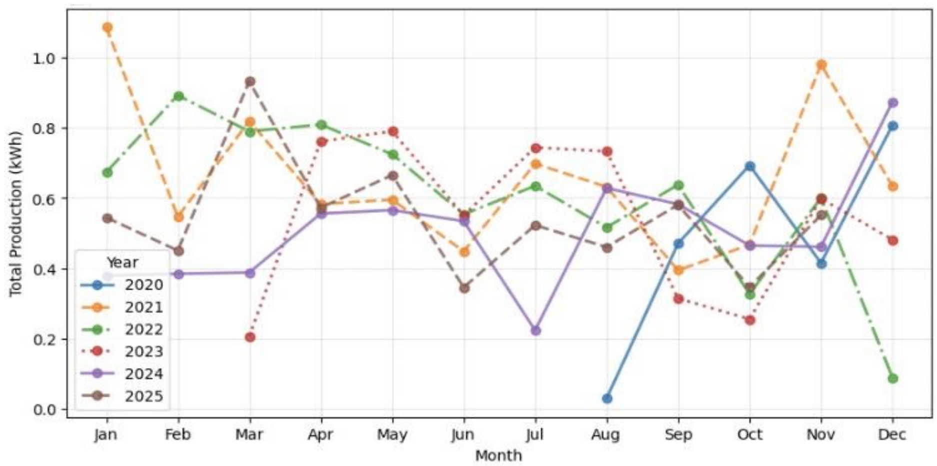

Figure A 39 - Year-over-year Wind Production Profiles (Plant 5 Wind)

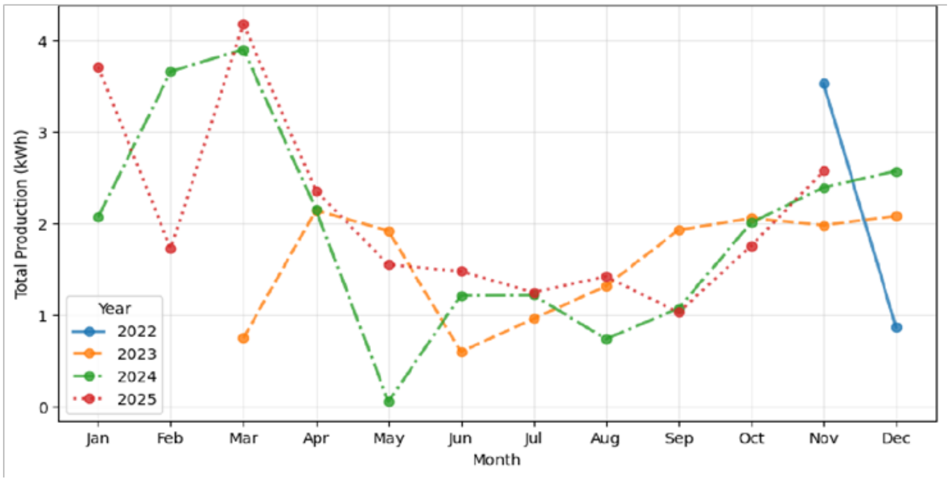

Figure A 40 - Year-over-year Wind Production Profiles (Plant 6 Wind)

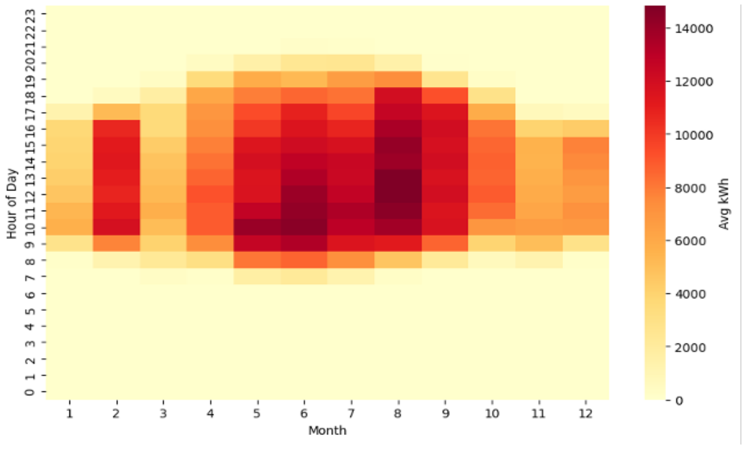

Figure A 41 - Hourly–Monthly Production Heatmap for Solar (Plant 2 Solar)

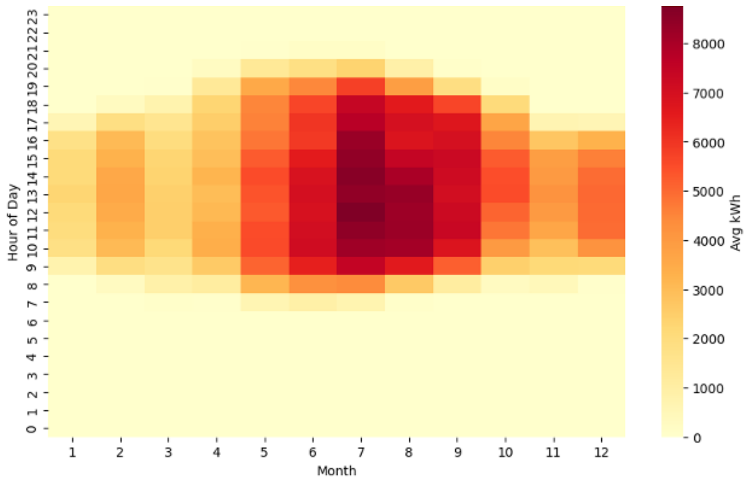

Figure A 42 - Hourly–Monthly Production Heatmap for Solar (Plant 3 Solar)

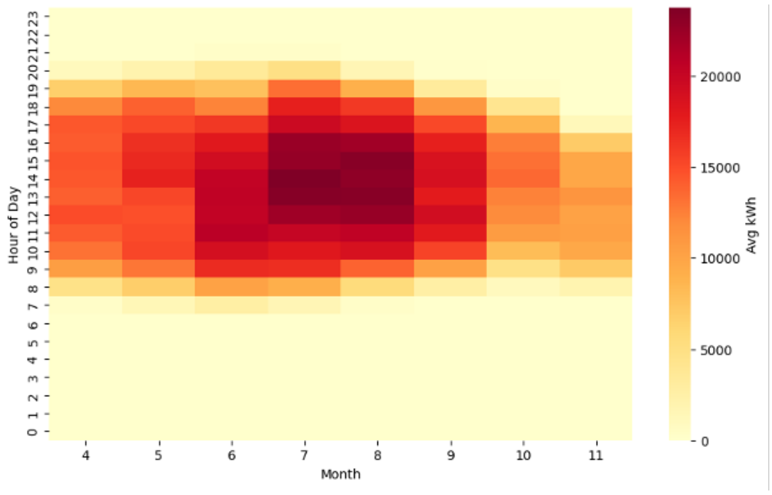

Figure A 43 - Hourly–Monthly Production Heatmap for Solar (Plant 4 Solar)

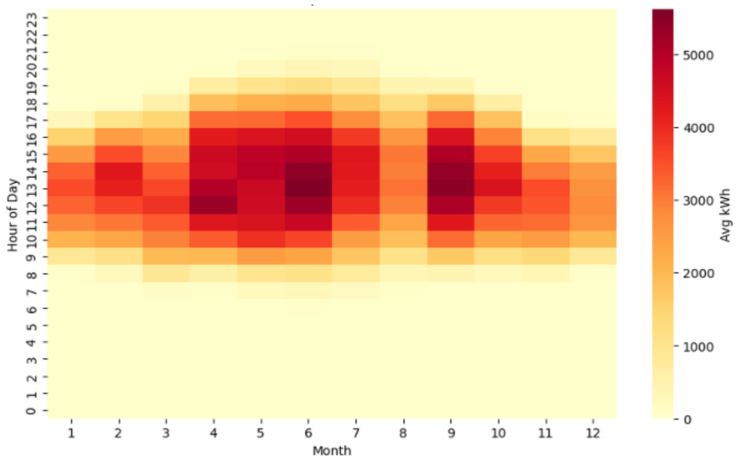

Figure A 44 - Hourly–Monthly Production Heatmap for Solar (Plant 5 Solar)

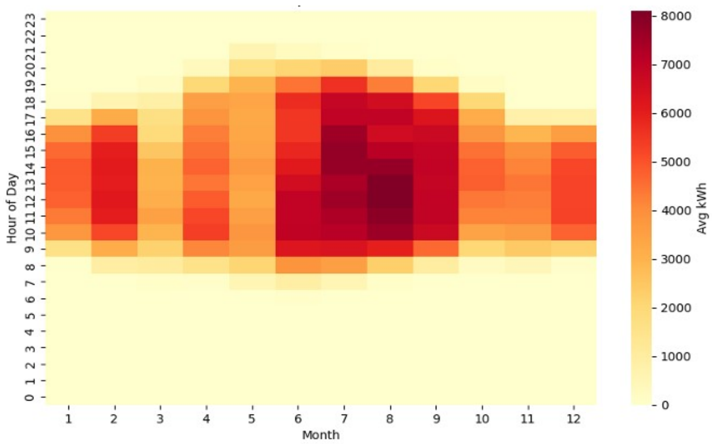

Figure A 45 - Hourly–Monthly Production Heatmap for Solar (Plant 6 Solar)

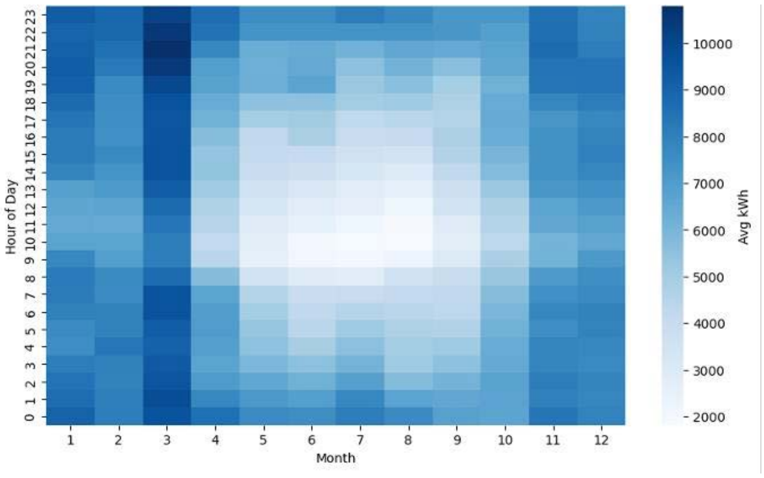

Figure A 46 - Hourly–Monthly Production Heatmap for Wind (Plant 2 Wind)\

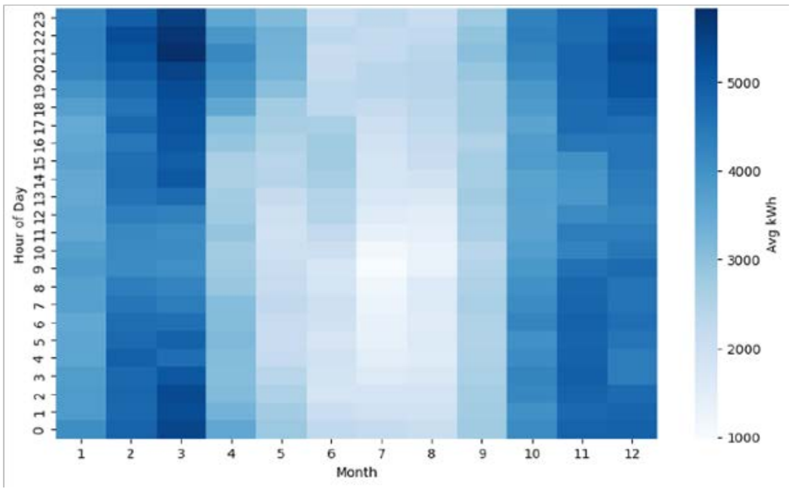

Figure A 47 - Hourly–Monthly Production Heatmap for Wind (Plant 3 Wind)

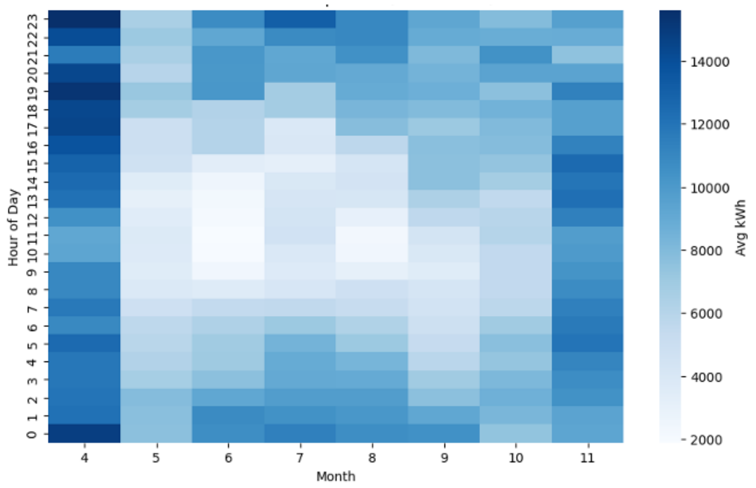

Figure A 48 - Hourly–Monthly Production Heatmap for Wind (Plant 4 Wind)

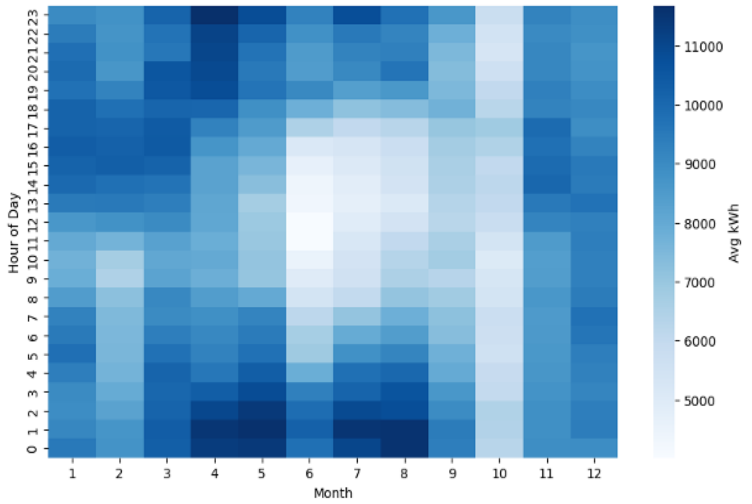

Figure A 49 - Hourly–Monthly Production Heatmap for Wind (Plant 5 Wind)

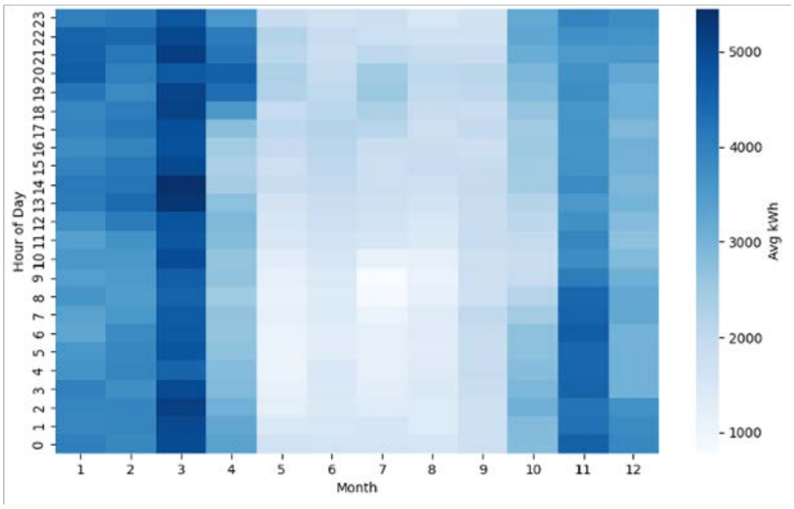

Figure A 50 - Hourly–Monthly Production Heatmap for Wind (Plant 6 Wind)

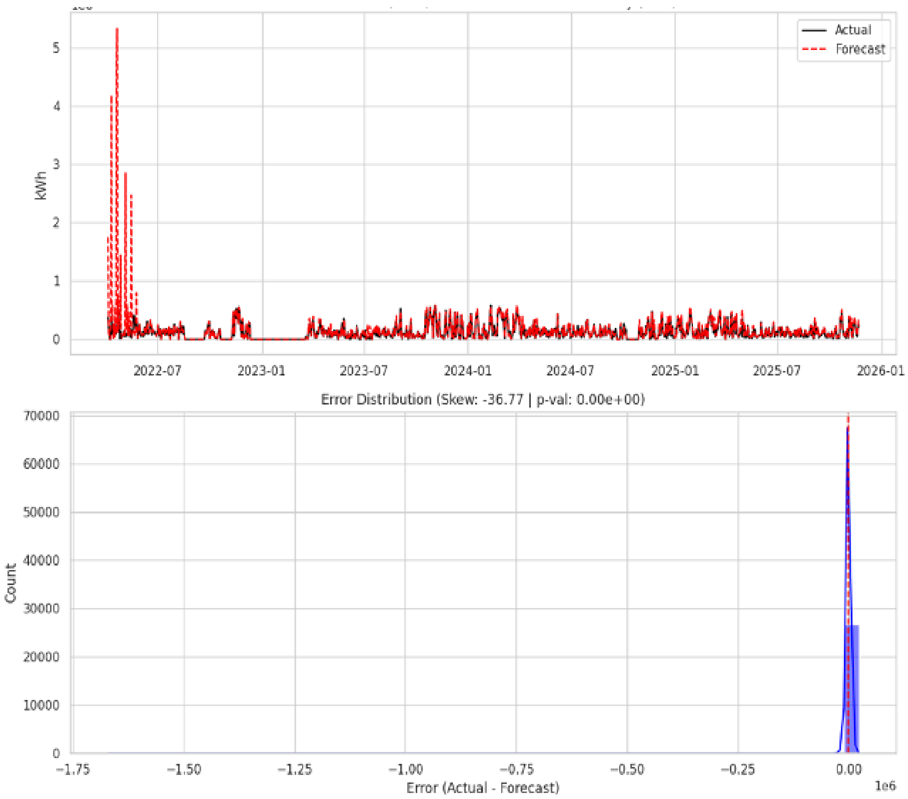

Figure A 51 - Plant 2 Wind Production: Observed vs Company Forecast and Errors

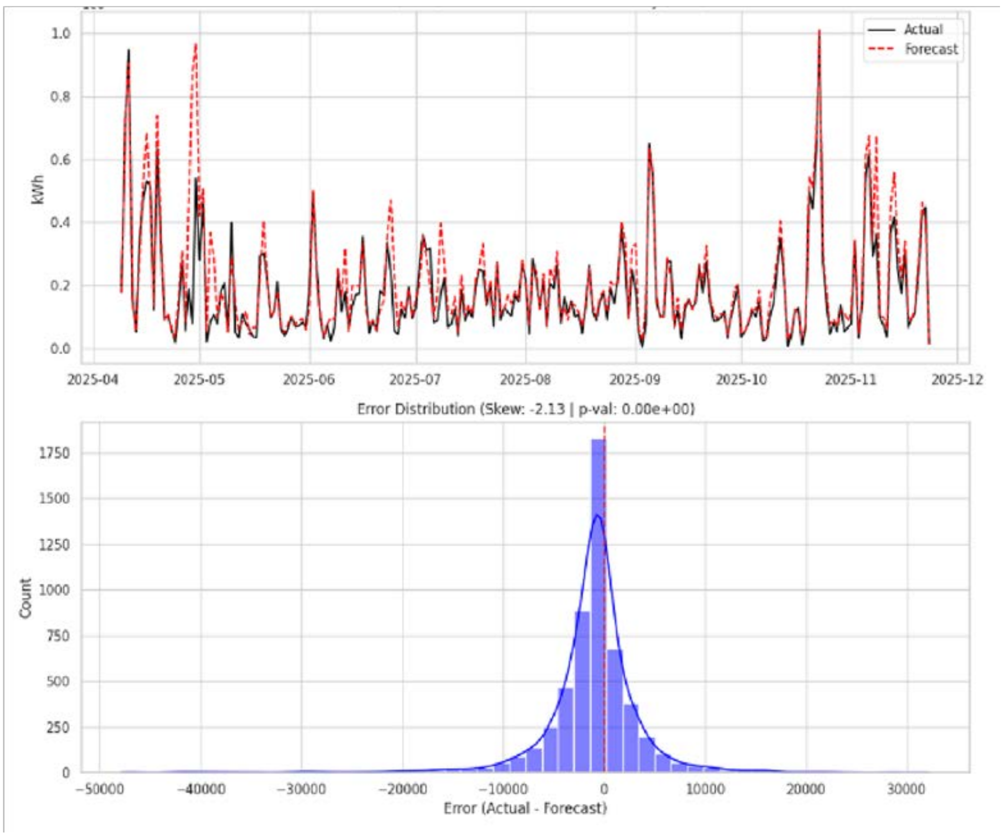

Figure A 52 - Plant 4 Wind Production: Observed vs Company Forecast and Errors

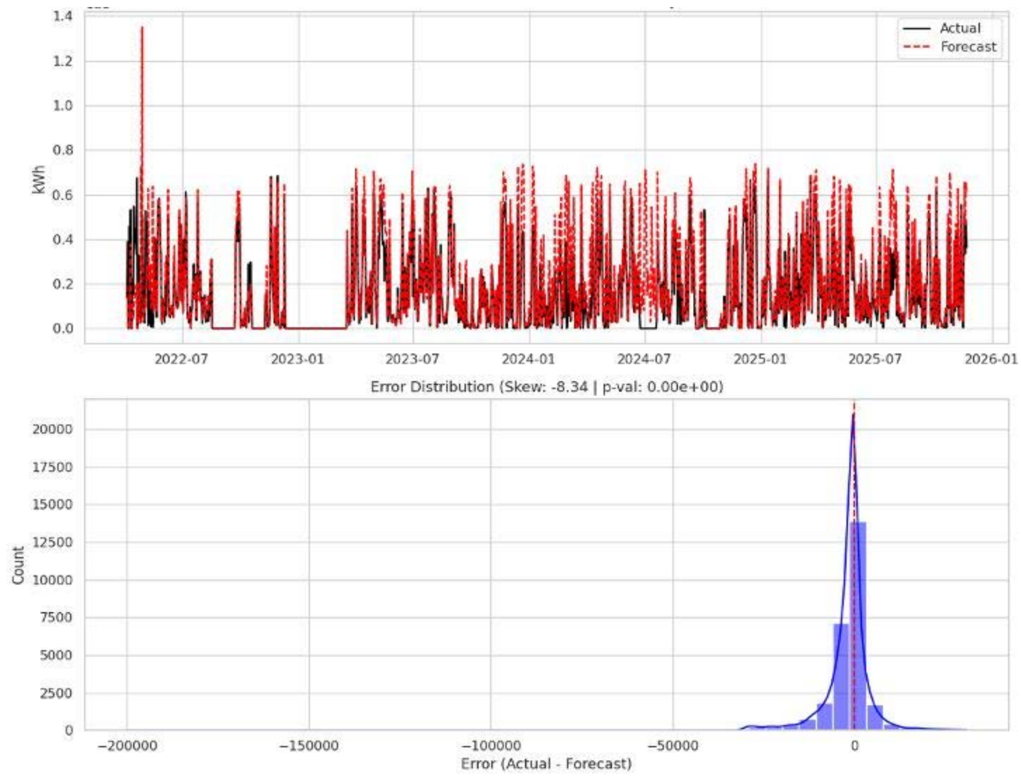

Figure A 53 - Plant 5 Wind Production: Observed vs Company Forecast and Errors

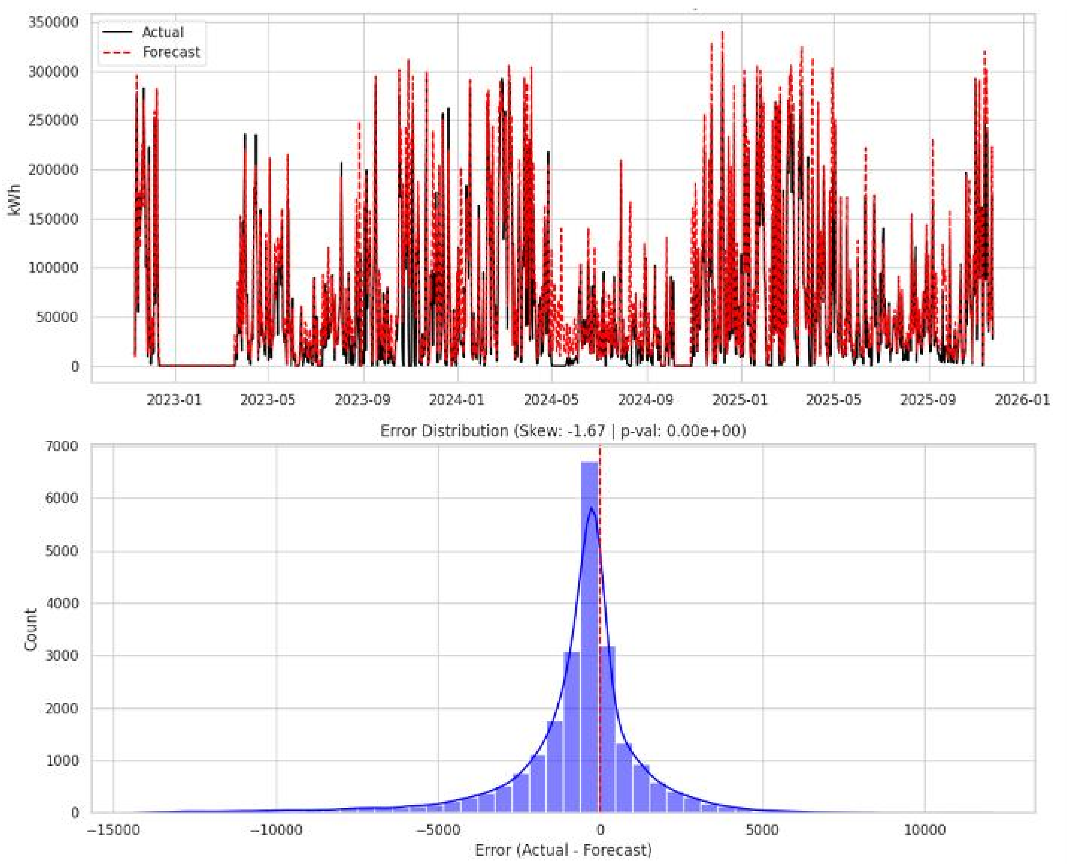

Figure A 54 - Plant 6 Wind Production: Observed vs Company Forecast and Errors

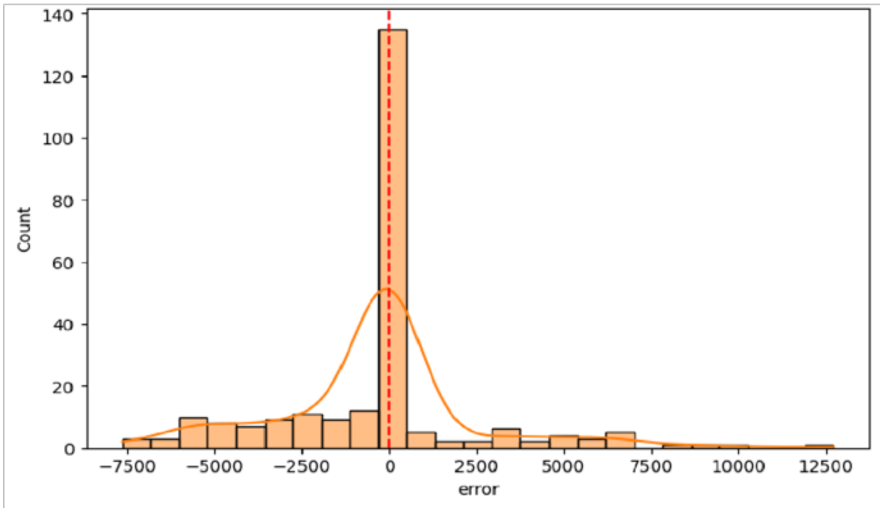

Figure A 55 - Solar Error Distributions Based on Company Forecasts (Plant 2 Solar)

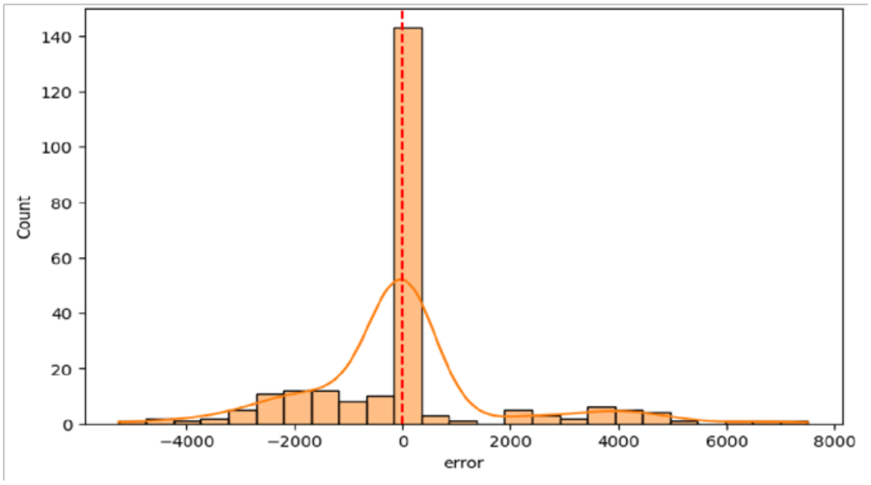

Figure A 56 - Solar Error Distributions Based on Company Forecasts (Plant 3 Solar)

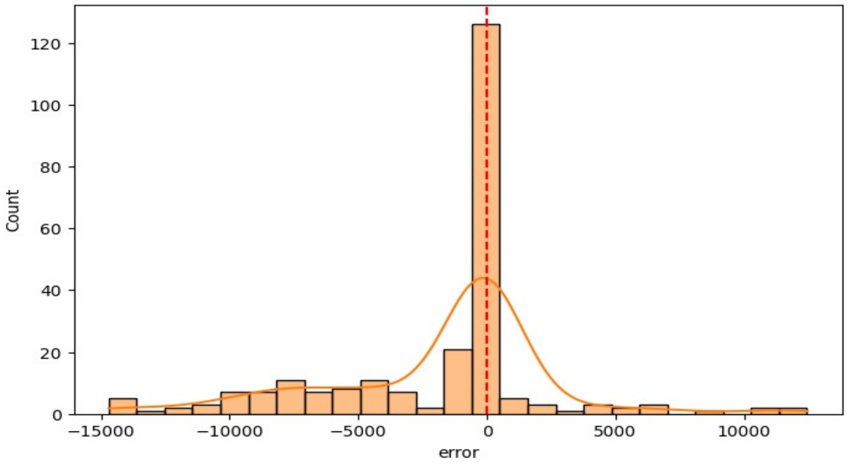

Figure A 57 - Solar Error Distributions Based on Company Forecasts (Plant 4 Solar)

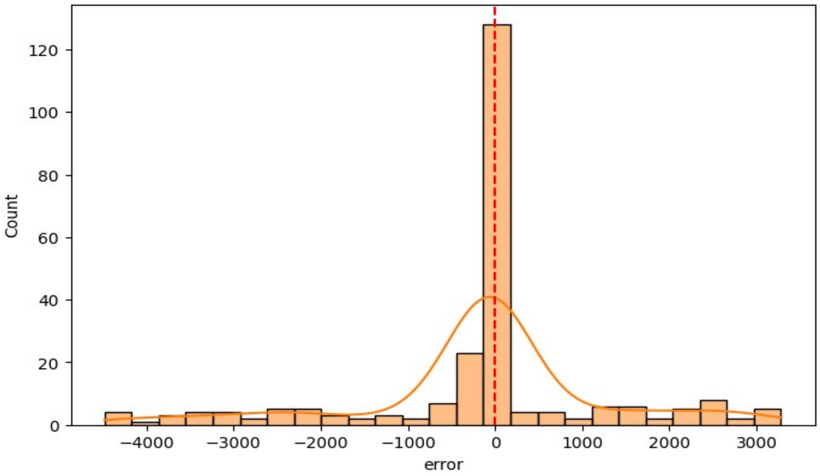

Figure A 58 - Solar Error Distributions Based on Company Forecasts (Plant 5 Solar)

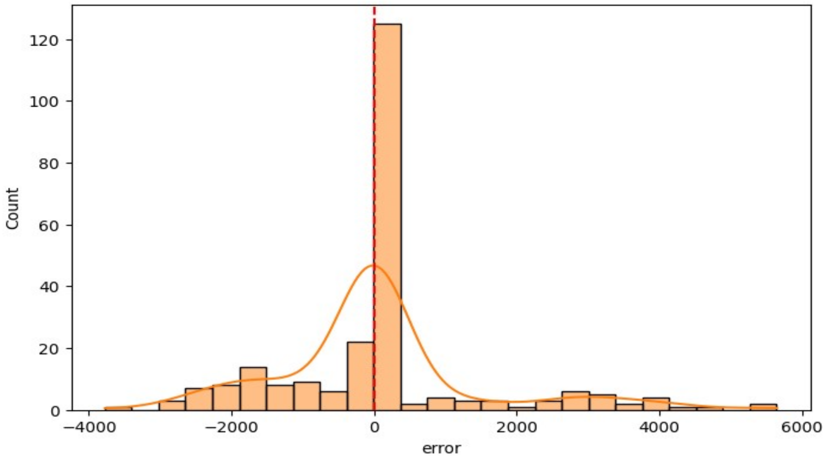

Figure A 59 - Solar Error Distributions Based on Company Forecasts (Plant 6 Solar)

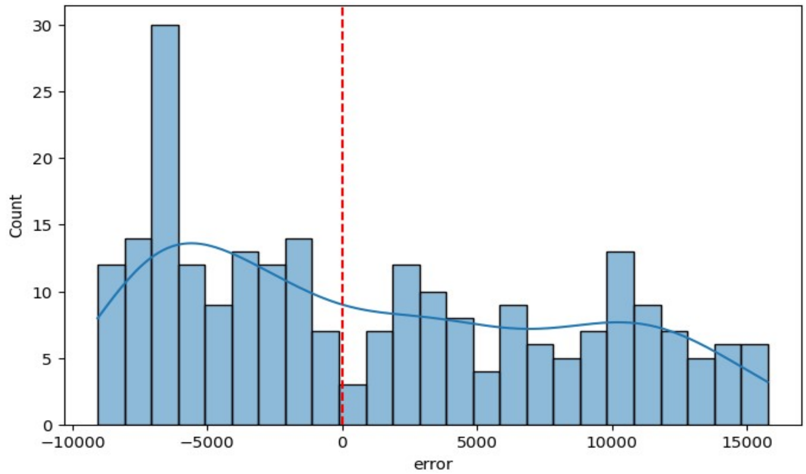

Figure A 60 - Wind Error Distributions Based on Company Forecasts (Plant 2 Wind)

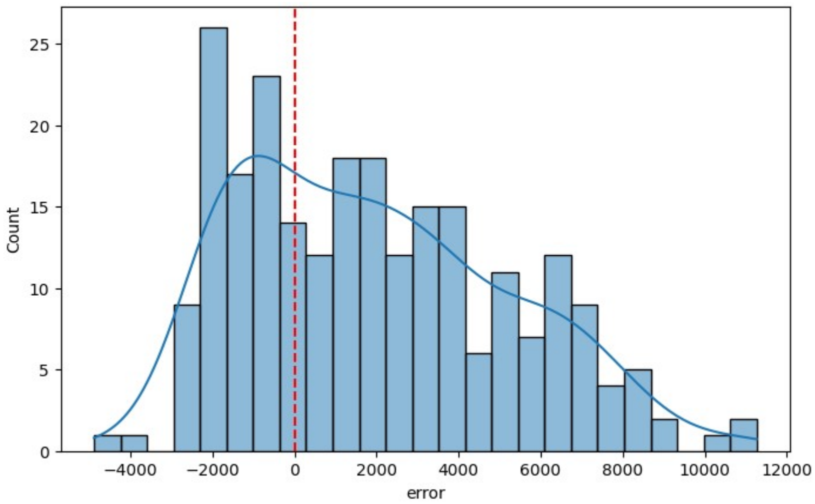

Figure A 61 - Wind Error Distributions Based on Company Forecasts (Plant 3 Wind)

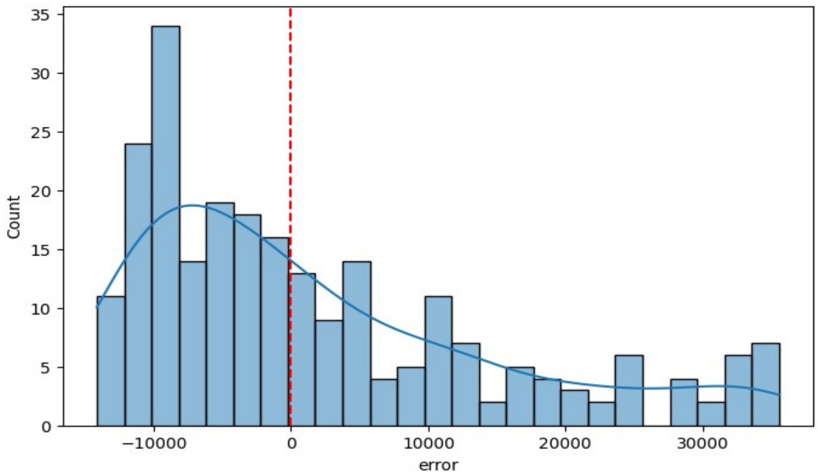

Figure A 62 - Wind Error Distributions Based on Company Forecasts (Plant 4 Wind)

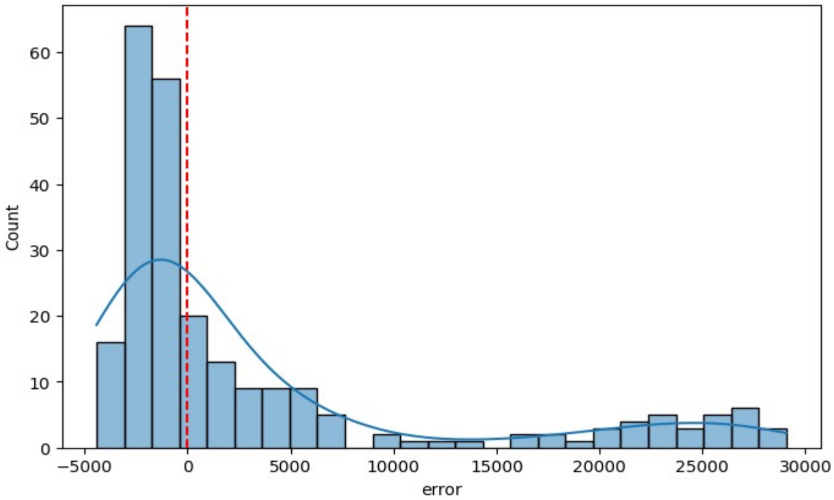

Figure A 63 - Wind Error Distributions Based on Company Forecasts (Plant 5 Wind)

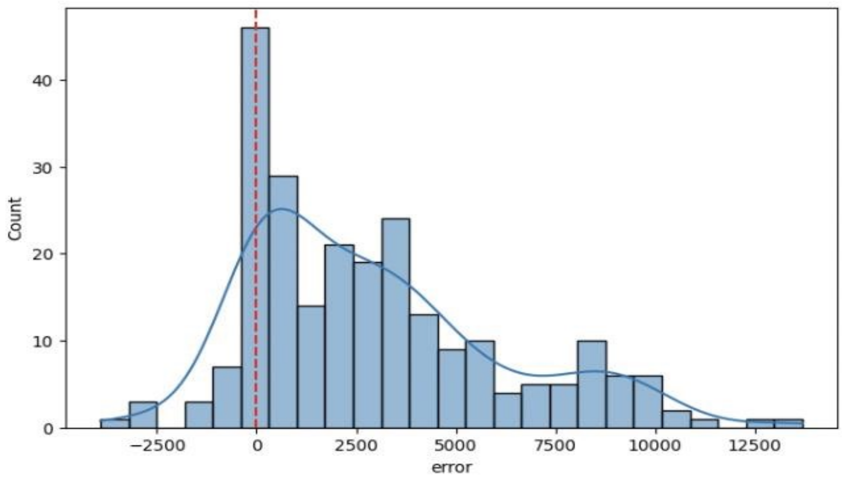

Figure A 64 - Wind Error Distributions Based on Company Forecasts (Plant 6 Wind)

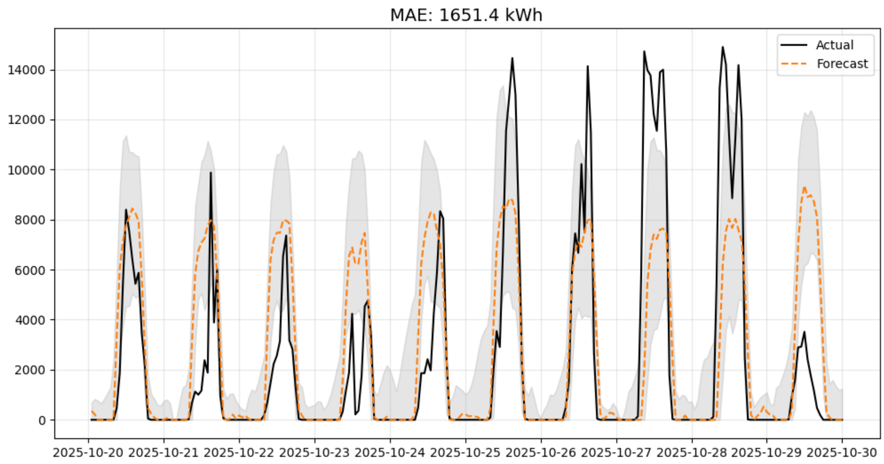

Figure A 65 – Observed vs Forecasted Solar Power Production (Plant 2 Solar)

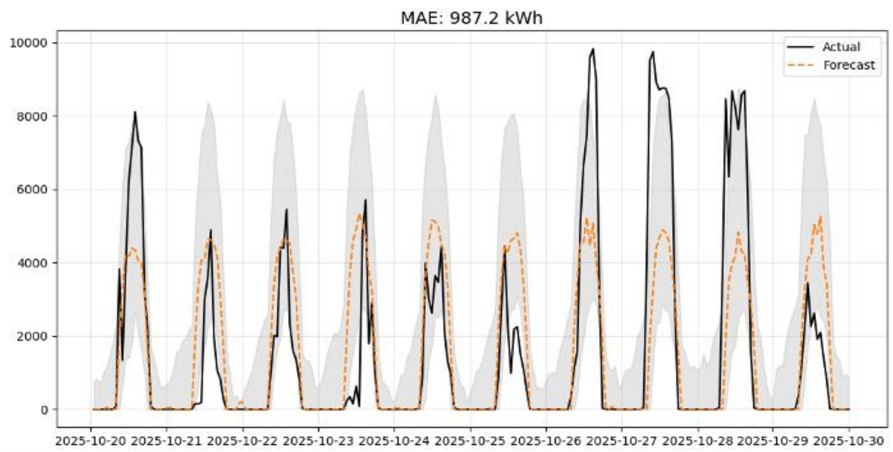

Figure A 66 - Observed vs Forecasted Solar Power Production (Plant 3 Solar)

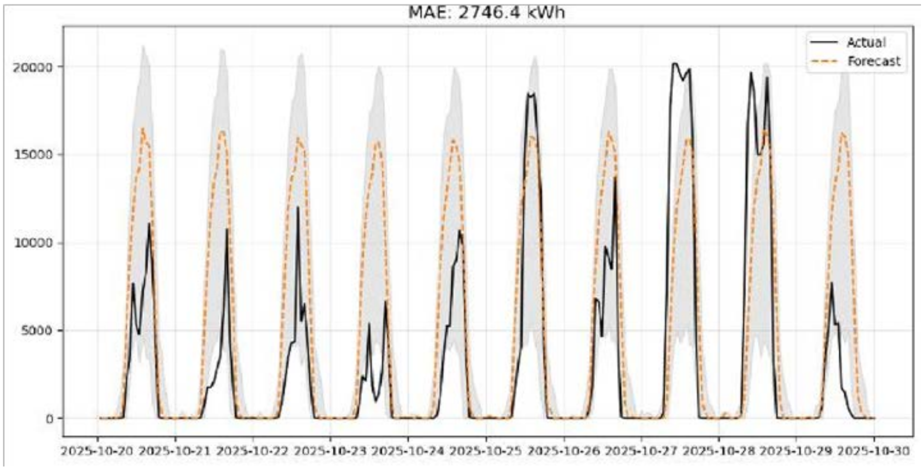

Figure A 67 - Observed vs Forecasted Solar Power Production (Plant 4 Solar)

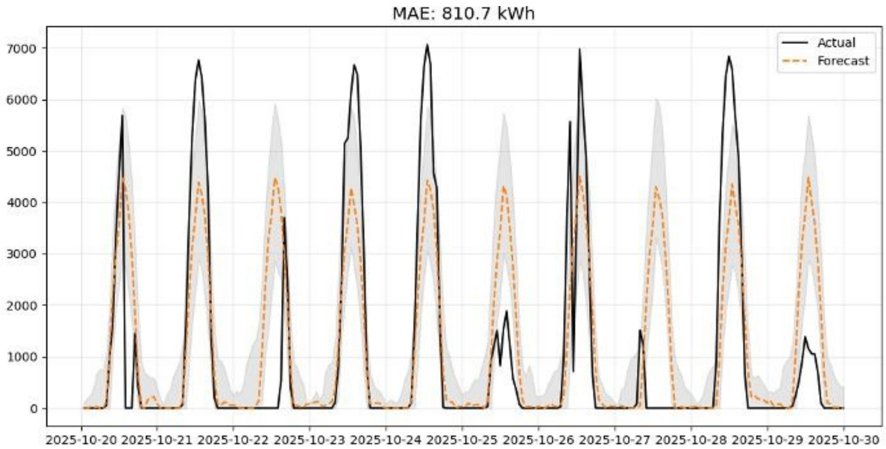

Figure A 68 - Observed vs Forecasted Solar Power Production (Plant 5 Solar)

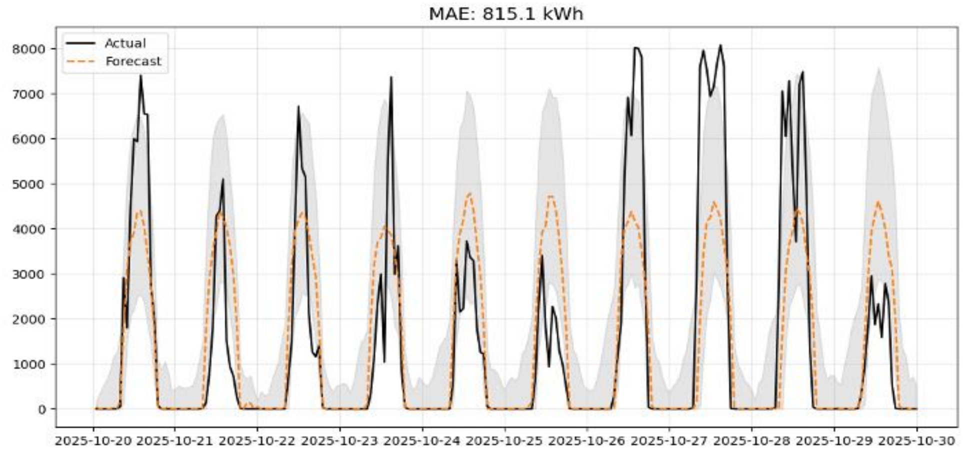

Figure A 69 - Observed vs Forecasted Solar Power Production (Plant 6 Solar)

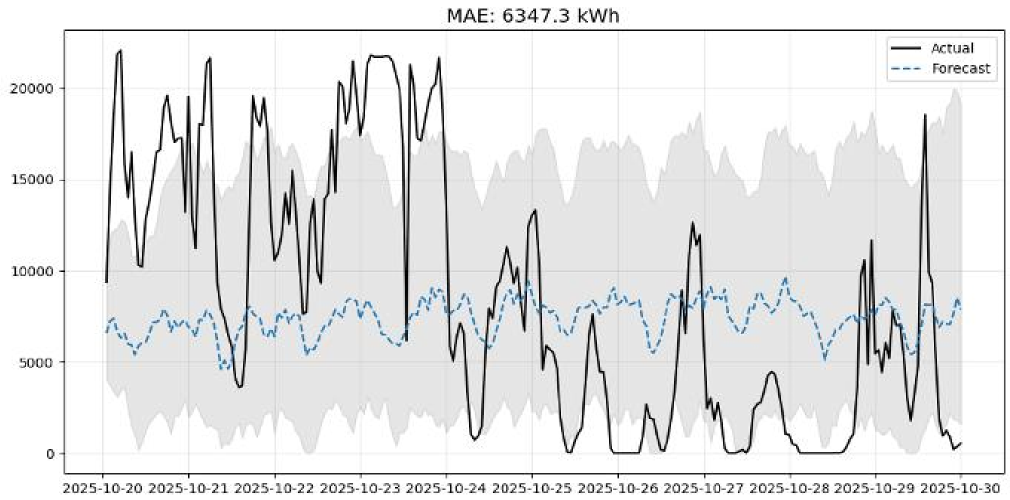

Figure A 70 - Observed vs Forecasted Wind Power Production (Plant 2 Wind)

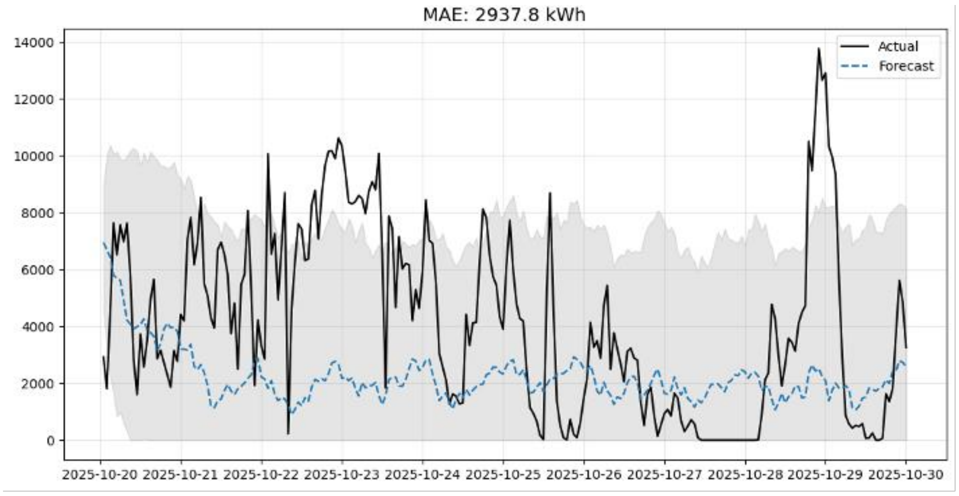

Figure A 71 - Observed vs Forecasted Wind Power Production (Plant 3 Wind)

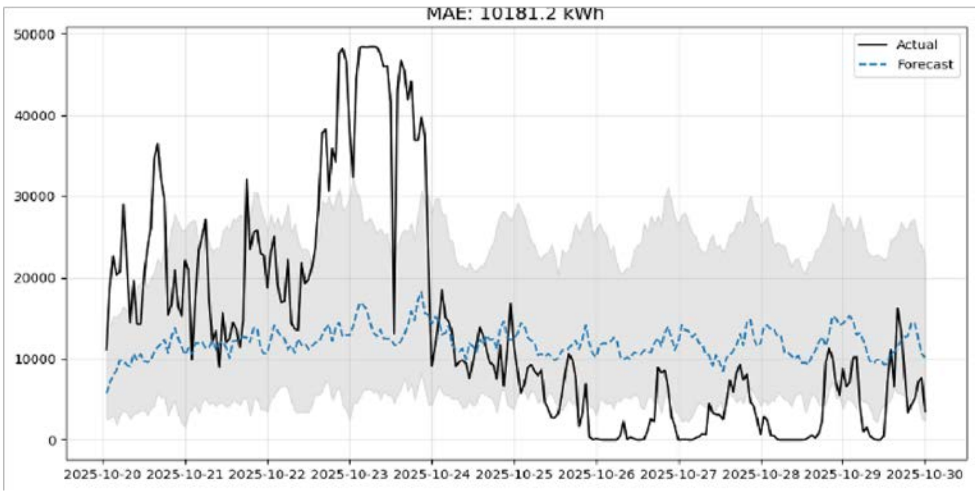

Figure A 72 - Observed vs Forecasted Wind Power Production (Plant 4 Wind)

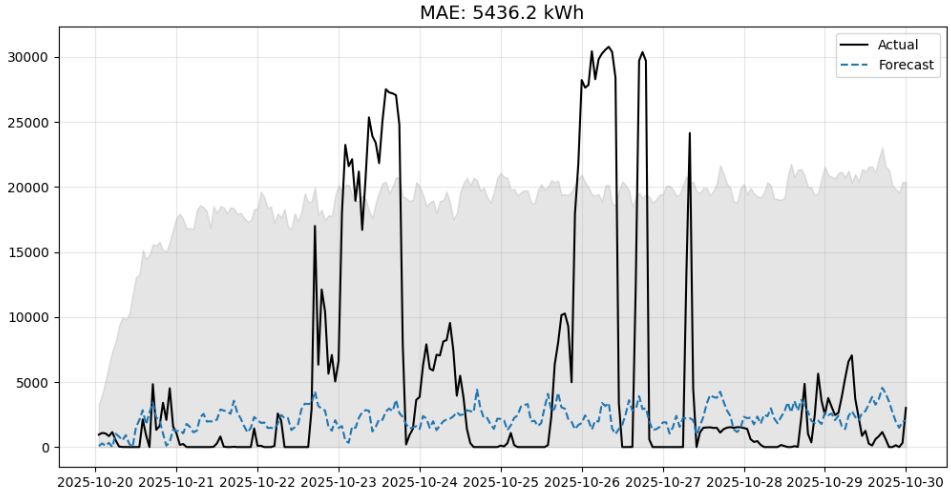

Figure A 73 - Observed vs Forecasted Wind Power Production (Plant 5 Wind)

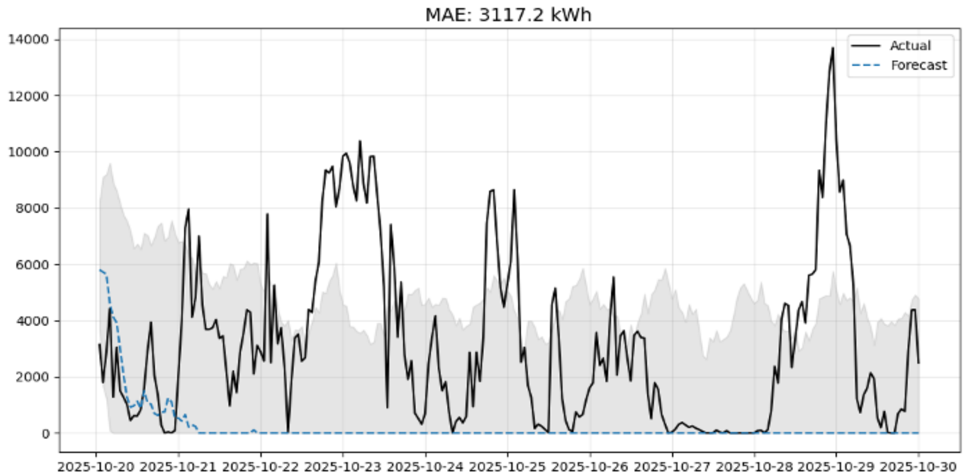

Figure A 74 - Observed vs Forecasted Wind Power Production (Plant 6 Wind)

ANNEX A

Outlook

RE: NOVA IMS | Ethics Committee - NEED REVIEW

From Ethics Committee <ethicscommittee@novaims.unl.pt>
Date Fri 27/02/2026 12:48
To leonorbaptistadealmeida@gmail.com <leonorbaptistadealmeida@gmail.com>; Leonor Mergulhao Alves Baptista de Almeida <20240172@novaims.unl.pt>; Flavio Luis Portas Pinheiro <fpinheiro@novaims.unl.pt>
Cc Ethics Committee <ethicscommittee@novaims.unl.pt>

Dear Leonor Almeida,
Dear professor Flávio Pinheiro,

Thank you for filling out the Research Ethics Checklist. After reviewing your request, you can proceed with the study, as we do not foresee any major ethical concerns with the project.

Project No.: **DSCI2026-2-228321**

Project Title: **A comparative analysis of solar and wind generation in hybrid power plants in the Iberian Peninsula**

Principal Researcher: **Leonor Mergulhão Alves Baptista de Almeida**

according to the regulations of the Ethics Committee of NOVA IMS and MagIC Research Center this project was considered to meet the requirements of the NOVA IMS Internal Review Board, being considered **APPROVED** on 27/02/2026.

It is the Principal Researcher's responsibility to ensure that all researchers and stakeholders associated with this project are aware of the conditions of approval and which documents have been approved.

The Principal Researcher is required to notify the Ethics Committee, via amendment or progress report, of

- Any significant change to the project and the reason for that change;
- Any unforeseen events or unexpected developments that merit notification;
- The inability of the Principal Researcher to continue in that role or any other change in research personnel involved in the project.

Lisbon, 27/02/2026
NOVA IMS Ethics Committee
ethicscommittee@novaims.unl.pt

This email serves as formal proof of ethical approval. If required for inclusion in a thesis, dissertation, or any other academic documentation, a PDF version of this message may be created and attached accordingly.

Cristina Oliveira
Gestora executiva do centro de investigação MagIC / Executive Research Manager of the Information Management Research Center (MagIC)
Tel. +351 213 828 610 (ext.: 2180)

Data with Purpose.

Campus de Campolide, 1070-312 Lisboa, Portugal
novaims.unl.pt