INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Photovoltaic Power Forecasting at national level using Ensemble Models

Inês Sofia Calmeirão Sabarigo

Master in Data Science

Supervisors:
Doctor Diana Elisabeta Aldea Mendes, Associate Professor,
ISCTE-Instituto Universitário de Lisboa

Doctor Nuno Miguel Pinho da Silva,
R&D Nester

October, 2025

BUSINESS SCHOOL

TECNOLOGIAS E ARQUITETURA

Department of Quantitative Methods for Management and Economics

Department of Information Science and Technology

Photovoltaic Power Forecasting at national level using Ensemble Models

Inês Sofia Calmeirão Sabarigo

Master in Data Science

Supervisor:
Doctor Diana Elisabeta Aldea Mendes, Associate Professor,
ISCTE-Instituto Universitário de Lisboa

Doctor Nuno Miguel Pinho da Silva,
R&D Nester

October, 2025

Acknowledgements

The completion of this dissertation could not have been accomplished without the contribution of a group of people, to whom I would like to take the opportunity to express my sincere gratitude.

First, I would like to thank my supervisors. My sincere thanks to Professor Diana Aldea Mendes for her guidance, support, and constant availability. Her invaluable insights and teachings have contributed significantly to this work and my academic and personal development. With the same appreciation, I thank Dr. Nuno Pinho da Silva for his thoughtful inputs and the guidance he kindly provided.

I would also like to express my sincere gratitude to my parents and brother. Their unconditional support led me forward and helped me overcome the challenges that would come along the way. They were always my safe haven.

I am also grateful to my Évora friends, Inês Neves, Maria Vornicescu, Marta Ribeiro, and Marta Sousa, for all the moral support and simply for being there. Their words of encouragement kept me motivated throughout this journey.

Resumo

A inclusão de sistemas fotovoltaicos na rede elétrica portuguesa tem vindo a aumentar significativamente nos últimos anos, o que, devido à sua natureza volátil, aumenta a necessidade de uma previsão precisa para manter o bom funcionamento e gestão da rede elétrica. Apesar da extensa investigação que evidencia as potencialidades de modelos *ensemble* para a previsão de energia ao nível da estação fotovoltaica, existe uma falta generalizada de investigação ao nível nacional ou multirregional. Para endereçar esta lacuna, a presente dissertação pretende explorar as capacidades dos modelos *ensemble* para realizar previsões do dia seguinte da potência fotovoltaica gerada à escala nacional portuguesa. Para concretizar este objetivo, os modelos *ensemble* XGBoost, *Random Forest*, LigthGBM, e um modelo *Stacking* foram testados. A metodologia definida focou-se na avaliação destes modelos sob diferentes configurações, como o número de *lags* usados como input nos modelos, o tamanho do conjunto de treino e diferentes números de passos a prever à frente. Foram testadas uma abordagem univariada, incluindo apenas as observações históricas da potência fotovoltaica, e uma abordagem multivariada, incluindo variáveis meteorológicas. Os resultados foram comparados com modelos *benchmark* e revelaram que os modelos *ensemble* foram eficazes na previsão de potência fotovoltaica à escala nacional. Também foi demonstrado que a exclusão de variáveis com pouca importância, baseada na análise SHAP, resultou numa melhor performance da previsão em diversas configurações. O XGBoost foi o mais eficaz, alcançando na sua configuração com melhor performance: um RMSE de 82.13, um MAE de 49.84, e um MAPE de 12.10%.

Palavras-chave: Previsão de Energia Fotovoltaica; Séries Temporais; Modelos *Ensemble*; Previsão do dia seguinte

Abstract

The inclusion of photovoltaic (PV) systems in Portugal’s power grid has been increasing significantly in recent years, which, due to its volatile nature, increases the need for an accurate forecast to maintain the power grid's good functioning and effective management. Although there is extensive research that shows the potentialities of ensemble models for PV power prediction at the plant level, there is an overall lack of research at the national or multi-regional level. To address this gap, this dissertation aims to explore the ensemble models' capabilities in predicting one-day-ahead PV power generated at Portugal’s national level. To do so, ensemble models such as XGBoost, Random Forest, LightGBM, and a Stacking approach were tested. The defined methodology focused on evaluating the different ensemble models under varying configurations, including the number of lags used as model input, the training dataset size, and different steps-ahead setups. Moreover, both an univariate approach, which includes only solar power or lag-related features, and a multivariate approach, where weather variables are also included, were tested. The results were compared with benchmark models and revealed that the ensemble models proved effective in PV power forecasting at a national scale. It was also demonstrated that excluding low-importance features, guided by SHAP analysis, led to improved forecasting performance in several configurations. Out of the tested ensemble models, the most accurate was XGBoost, achieving in its best-performing configuration an RMSE of 82.13, an MAE of 49.84, and a MAPE of 12.10%.

Keywords: Photovoltaic Power Forecasting; Time series; Ensemble Models; One-day-ahead

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACF: Autocorrelation Function

AI: Artificial Intelligence

AIC: Akaike Information Criterion

ADF: Augmented-Dickey-Fuller

ARIMA: Auto-Regressive Integrated Moving Average

DGEG: Direção Geral de Energia e Geologia

EFB: Exclusive Feature Bundling

GB: Gradient Boosting

GOSS: Gradient-based One-Side Sampling

KPSS: Kwiatkowski–Phillips–Schmidt–Shin

LightGBM: Light Gradient-Boosting Machine

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

ML: Machine Learning

PACF: Partial Autocorrelation Function

PP: Phillips-Perron

PV: Photovoltaic

RF: Random Forest

RMSE: Root Mean Square Error

SARIMA: Seasonal Auto-Regressive Integrated Moving Average

SHAP: SHapley Additive exPlanations

SLR: Systematic Literature Review

UCM: Unobserved Components Model

XGBoost: eXtreme Gradient Boosting

CHAPTER 1

Introduction

1.1. Background

Renewable energy expansion has become a common target for all nations as a way of mitigating global climate change and maintaining energy supply security. With the advancement of Photovoltaic (PV) technology and its declining costs over time, solar power generation has been experiencing significant growth worldwide (Liu et al., 2023a). As briefly explained by Ahmed et al. (2020), photovoltaic technology generates electricity through the charge build-up caused by the excitation of free electrons present in the semiconductor material, which is triggered by photons from sunlight. PV panels can be installed in different types of configurations, such as rooftops and open spaces, and at various scales. However, the power generated by these systems is challenging to predict due to its dependency on environmental factors (Rai et al., 2022); it has fluctuations caused by different effects, such as cloud coverage and variations in solar irradiance. A reliable PV output forecast will significantly mitigate the negative impacts of its uncertainty, enhance stability, and improve economic viability (Ahmed et al., 2020). Consequently, research on solar power forecasting has been done by exploring various forecasting models, such as physical, statistical, and machine learning models.

Physical models, which are based on mathematical equations, are considered complex, since they need detailed geographic and meteorological data to predict PV power (Liu et al., 2023b). Abdelmoula et al. (2022) implemented a physical model as a reference. Ultimately, that model demonstrated the worst performance among all the applied methods.

Traditional time series models, such as Auto-Regressive Integrated Moving Average (ARIMA) and the seasonal version of the ARIMA model (SARIMA), have also been applied in this context. Singh and Pozo (2019), as cited by Wang et al. (2022), used an Auto-Regressive Moving Average (ARMA) model for hourly PV power prediction. However, these traditional methods are not suitable for nonlinear relationships, which is the case with the relationship between PV power and the factors that affect it (Zhang & Zhu, 2022).

Alternatively, Machine Learning (ML) models can better detect more complex and nonlinear patterns between variables (Gaboitaolelwe et al., 2023). Başaran et al. (2020) concluded from their systematic literature review (SLR) that the most used ML models for PV power forecasting are Artificial Neural Networks (ANN) and Support Vector Machines (SVM), which generally produce good results. Krechowicz et al. (2022) also showed the extensive use of Recurrent Neural Networks (RNN) in PV power forecasting.

All the methods mentioned above focus only on single models. Yet, these single models are not usually able to accurately predict when using different datasets or different hyperparameters for the same data, having quite volatile results (Khan et al., 2022). Recent research indicates increasing development of ensemble methods, showing their potential to predict PV power output and overcome these issues. Ensemble methods combine the predictive capabilities of various models, known as weak or base learners, to enhance the accuracy and stability of the overall model (Gaboitaolelwe et al., 2023).

Ensemble models can be classified according to how the base learners are connected. One way to connect these models is the bagging method, which uses bootstrap sampling and trains the weak learners in parallel. The final prediction is accomplished by aggregating the individual results. Sedai et al. (2023) showed that a Random Forest (RF), which is an ensemble model that uses the bagging method, has a good predictive capability to estimate the output power of a PV plant in Texas, U.S.A., for multiple time horizons, ranging from 1 to 15 days.

Alternatively, boosting is a method that trains weak learners sequentially, improving performance by assigning higher weights to observations with more significant errors in each iteration (Rahimi et al., 2023). Kyeremeh et al. (2022) used an ensemble of multiple Long Short-Term Memory networks (LSTM) with the Adaptive Boosting (AdaBoost) method to predict PV power output, showing a significantly better performance than the models used as benchmarks, such as RFs and single LSTMs.

Another type of ensemble method is stacking, which starts by training base-learner models in parallel. Their predictions are then used as training data for the final estimator, called meta-learner. Liu et al. (2023a) demonstrated inconsistencies in single models' ability to accurately predict outcomes under different weather types. In contrast, the proposed stacking ensemble model presented a robust performance. The base learners used were: Generalized Regression Neural Network (GRNN), Extreme Learning Machine (ELM), Elman Neural Network (Elmann), and LSTM. The meta-learner used was Backpropagation Neural Network (BPNN). Some studies have even used various ensemble methodologies combined in their research. Abdellatif et al. (2022a) built a stacking model where the base learners were also ensemble models, such as eXtreme Gradient Boosting (XGBoost), Adaboost, and RF. The meta-learner was the Extra Tree Regressor (ETR). This model outperformed the base learners, suggesting that combining different ensemble models further enhances the predictive potential of these methods.

Other ensemble methods applied in recent studies include the weighted average of single models (Sharma et al., 2021) and Bayesian model averaging (Banik & Biswas, 2023), among others.

1.2. Problem Statement

Despite the growing trend in research showing the capabilities of ensemble models for Solar power forecasting, the great majority of these studies focus on a single PV power plant. Taking this into consideration, plus the increasing integration of national electricity markets into a single European market, further research is needed regarding the ability of ensemble models to forecast short-term PV power output at a national or multi-regional level, namely to support the electricity grid's operational planning and the electricity market as a whole.

Portugal has registered a significant growth in PV installed capacity over the past few years, which naturally led to a strong increase in solar power generation (Figure 1.1.). According to the December 2024 "Estatísticas rápidas das renováveis" publication by *Direção-Geral de Energia e Geologia* (DGEG), from 2015 until 2024, Photovoltaic was the technology that had the highest growth in terms of installed capacity, with an increase of around 5.2 GW. Moreover, it expanded its weight in electricity production, from 2.2% in 2012 to 13.6% in 2022, with a more expressive growth in the last two years (Observatório da Energia, DGEG, & ADENE, 2024).

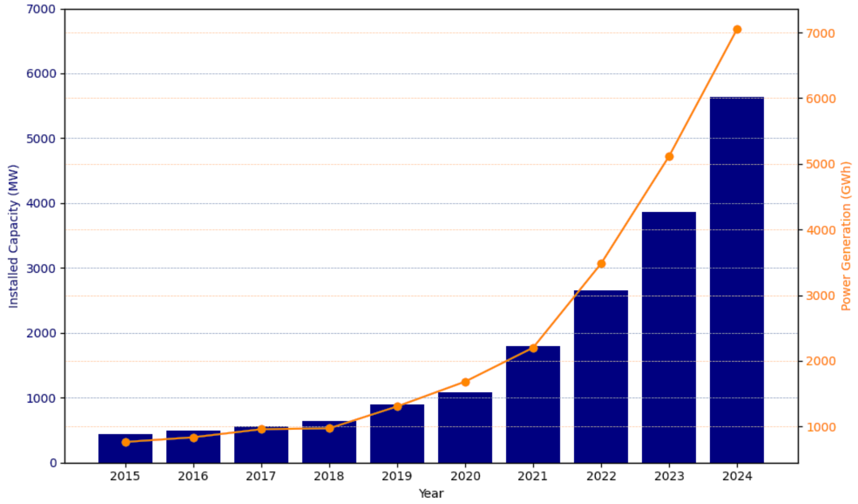


Figure 1.1. PV installed capacity in Portugal's mainland and generated PV power (Data source: DGEG, 2025)

With Portugal's significant growth in PV technology, the need for a reliable short-term output forecast plays an even more crucial role in the country's integration into the European electricity market, as well as in the operational planning and stability of the energy grid.

1.3. Objectives and Contributions

The primary objective of this research is to investigate the application of ensemble models in forecasting short-term day-ahead PV power output at an aggregated, multi-regional level, specifically at Portugal's national scale. This study contains the performance evaluation of state-of-the-art ensemble models, such as XGBoost, RF, and Light Gradient-Boosting Machine (LightGBM). Additionally, a Stacking model will also be tested using as base learners the two best-performing single ensemble models. This evaluation will allow us to assess the capability of ensemble models in predicting PV power output when compared with traditional time series models, and identify the most effective approach. Moreover, various configurations, including different training dataset sizes and different numbers of lagged solar observations used as input features, will be implemented and tested to identify the techniques that yield the best solar power forecasting performance. In addition, a multivariate approach, including weather data at the national level, will also be explored with the goal of identifying possible enhancements when compared to the univariate approach. To achieve these objectives, the study will utilize hourly PV power generated in Portugal from January 1st, 2023, to June 6th, 2024.

Based on these goals, the study's main contributions are: a comparative evaluation of ensemble forecasting models under diverse configurations, the study of a methodology to also include weather variables in PV power forecasting on a multi-regional level, and the gathering of findings of best methods that can support the achievement of a reliable and accurate PV power forecasting in Portugal that can later be used by policymakers.

1.4. Structure of the dissertation

This dissertation presents the following structure: after the introduction, Chapter 2 (Literature Review) describes the systematic literature review conducted on the subject of Ensemble models for PV power forecasting. Chapter 3 (Methodology) outlines the methodology followed, detailing the data preparation and modeling process used. In Chapter 4 (Results and Discussion), the results obtained by the constructed models are presented and discussed, providing a comparison analysis between the different models. Lastly, Chapter 5 (Conclusion and Future Work) summarizes the main conclusions of this study and highlights recommended future work for further research on this subject.

CHAPTER 2

Literature Review

Because of its dynamic nature, the continuous integration of more PV systems in the power grid leads to the increasing necessity of building accurate PV power forecasting. This could help to maintain the stability of the power grid and optimize the utilization of the PV systems (Gaboitaolelwe et al., 2023).

As previously mentioned, Ensemble models have been frequently developed for PV power forecasting in recent years, but have not yet been the subject of a focused review. The SLR described in this section was conducted with the primary purpose of understanding and synthesizing the state-of-the-art ensemble methodologies developed in recent research, published over the last four years, with a main focus on forecasting PV power output. Moreover, it enabled us to identify current research gaps and provide directions for future research.

2.1. Methodology

The approach followed in the present SLR was based on the framework used by Başaran et al. (2020), which begins with the definition of research questions. In this study, two research questions were defined to establish the structure for the systematic review. One of the research questions (**RQ1**) defined as "what dataset was used" in the respective ensemble approach, aims to understand the details of the dataset, such as independent variables used in the forecasting models, the data frequency, the period covered by the data, the forecasting horizon, and the location/country where the PV power was generated. The second research question (**RQ2**) aims to identify and understand "what ensemble models or approaches were implemented for the PV power forecasting" in the reviewed studies.

The next step defines the search strategy by determining the search terms based on the research questions and building the corresponding search queries. In this way, to collect the papers to be reviewed, the digital databases Scopus and Web of Science were utilized. The first set of key terms is related to the "population" of the research, and the terms used were "solar power", "photovoltaic" and its abbreviation "PV". The key terms "output" and "power" were also used to define the domain of the research more clearly, given that during some trial tests, a significant portion of the retrieved articles concerned another topic related to solar energy. The next set of research terms defines the forecasting approaches to investigate (intervention). They are "ensemble model*", "ensemble approach", "ensemble techniq*", "ensemble method*" and lastly "ensemble learning" (the * serves to also include variations of the word, for example, expressions such as ensemble methodology or ensemble methods would be included). Lastly, the outcome of the research is defined by the search terms forecasting and its derivatives ("forecast*"), "prediction", and "estimation".

The search strategy was refined upon a quick review of the retrieved articles. In the first step, all sets of the mentioned key terms were searched in the papers' titles, abstracts, and keywords. However, it was observed that some papers weren't related to PV power forecasting. To solve this issue and narrow the results to studies that focus on the intended problem, the terms "PV", "photovoltaic", and "solar power" were required to be present in the title of the articles. In this way, articles that aimed to predict other variables, such as wind energy production, and would only mention the topic of PV power, were excluded from the search. The key terms related to the research outcome, such as forecasting, prediction, etc., were also required in the title to exclude articles whose purpose was related to the PV power, but not its forecasting, for instance, fault detection in the photovoltaic systems. With this being said, the search scope for both groups of keywords was the title. However, the search scope for the group of words referring to the type of methodology used in power forecasting, as well as the keywords output and power, was the title, abstract, or keywords.

Furthermore, after defining the search queries, it was also necessary to determine other inclusion and exclusion criteria to select which studies should be reviewed in detail. The reviewed studies were those published in the last four years (between 2021 and 2024) included in journal articles or conference proceedings. Review papers, articles that weren't written in English, or duplicates of the same study already retrieved, were excluded from this review. To further narrow the research scope, it was also decided to focus on the deterministic approaches, excluding studies where the main purpose was probabilistic forecasting.

Following this methodology, the research queries in the two online datasets, already considering the year of publication and the type of article, retrieved a total of 172 articles. However, it was possible to observe that a significant portion of them (61 articles) were duplicates, resulting in 111 articles to be reviewed. After removing the duplicates, it was still necessary to execute a manual exclusion, considering some of the criteria defined. Regarding the type of article, it was required to exclude 4 studies that were review papers but weren't previously classified as such. Six of the articles weren't available to be read, and for that reason, they were also excluded, ending up with 101 articles. Moreover, 1 of the articles was published before the selected period of analysis. Furthermore, 10 remaining studies didn't focus on deterministic forecasts and were excluded (90 articles remaining). The remaining studies also went through a quality assessment where, in case the study didn't clearly allow to answer the research questions (understanding what ensemble approaches were followed, the corresponding results, and which datasets were used), they were also excluded. Although the keywords used in the searches were specific, there were still a few studies in which the purpose was not PV power forecasting, such as predicting the optimal panel direction, among others. Besides that, some studies wouldn't focus on exploring ensemble methodologies and would only use them for comparison, ultimately being excluded from the review.

That said, the review ended with 60 articles thoroughly read and included in the research results.

2.2. Literature review results and discussion

To summarize the recent studies on PV power forecasting that used ensemble methods, the information and main characteristics of the 60 selected articles are presented here.

Regarding document type, most studies were articles from journals (37), and the remaining were conference papers (23). Only a small part was published in 2021 (7). The year 2023 had the most publications (19), closely followed by 2024 (17) and 2022 (17), indicating that ensemble methods have been intensively studied over the last three years.

The most important parameter to evaluate the performance of the proposed models in scientific research is precisely the dataset used (Bozyiğit et al., 2019, as cited in Başaran et al., 2020). For this reason, regarding the first research question, the details in the datasets of the reviewed studies, such as data resolution, data period (in days), forecasting horizon, and location, were further analyzed and are present in Appendix A., Table A.1.

The research done on PV power forecasting can be categorized based on its forecasting horizon. Different forecasting horizons serve various purposes. For example, long-term predictions can be used when attention is focused on energy policymaking, whereas short-term or very short-term predictions can be made to prepare power production plans or to evaluate energy contracts (Zhang & Zhu, 2022). Although there isn't an agreement in the research on what are the different categories of the forecast horizon, the current paper considers the existence of four categories: very short-term (less than 1 h), short-term (1h to 1day), mid-term (more than 1 day and less than 1 month), and long-term (more than 1 month). In some cases, the authors mentioned that the forecast was short-term, but didn't mention which steps-ahead were used, so in those cases, for simplification purposes, the author's classification was maintained. It can be seen that the short-term, more specifically, the day-ahead forecasts, are predominant in the reviewed papers. Only 7 articles forecasted more than one day-ahead, being one month the largest forecast horizon. Sharma et al. (2021) built predictions for the short, medium, and long-term, demonstrating the consistency of the proposed ensemble model. Still, some articles (13) didn't mention the forecasting horizon they focused on.

Various data time resolutions are used in the reviewed research, allowing for the observation that the most common was 15 minutes (31%), followed by 1 hour (29%) and 5 minutes (18%). Two studies used two datasets with different data resolutions, so they were counted for both resolutions.

Regarding the location of the PV power systems used in research, it was observed that two of the studies used two datasets from distinct countries, so they were both counted respectively, and three papers didn't mention the location. Australia stands out from the rest of the countries with 13 studies. This might be because the data used in 8 articles were from Alice Springs and 1 from the Yulara solar system; both datasets are easily accessible and downloadable online. PV systems from Central America aren't present in this review, and PV systems from Africa and South America are only present in one study each. Two of the studies use datasets from North America. Europe and Oceania (specifically, Australia) are represented in 10 and 13 studies, respectively. The Asian continent stands out from the other continents with 32 studies.

Moreover, none of these papers focused on a national scale. Only one of the analyzed studies performed PV power forecasting at a regional level (electrical market zones in Italy). Taheri et al. (2024), in the absence of data related to the exact location of solar power plants, made use of geolocated meteorological forecasts from densely populated subregions, in which the Stacking approach consistently demonstrated strong performance. However, not enough details were provided regarding the stacking base and meta-learner models. Kim et al. (2021) used data from three solar power plants in different regions of South Korea to train the model, but the forecast was still done on a plant level, similar to Kyeremeh et al. (2022), who used the normalized data from 21 PV facilities across different regions.

Five of the reviewed works evaluated the generalization capability of the proposed ensemble models by applying them to two different datasets. For example, Mondal et al. (2024) tested the generalization capability of the model by using PV power data from different solar plants within the same geographic region, but more than 150 kms from the original tested plant. Zhang and Zhu (2022), and Dai et al. (2024), utilize two datasets, one from China and another from Australia.

Around 50% of the studies used at most one year of data, with a minimum of 19 days and a maximum dataset of about 7 years. On average, the research studies used approximately 2 years of PV power data.

As mentioned by Ahmed et al. (2020), the input features can influence the model’s performance, so they should be selected with caution to avoid increasing the forecast errors, runtime, and computational complexity. It was observed that some papers applied techniques, such as Correlation analysis or Principal Component Analysis (PCA), to select the exogenous features that mainly affect PV power output and reduce data dimensionality. Appendix A., Table A.2. presents the exogenous variables in the original datasets and the corresponding selected features used in the ensemble models. Zhou et al. (2023a) showed the importance of carefully selecting the independent features with the training of an XGBoost model to determine the significance of each feature, and then performed Sequential Forward Selection (SFS) to choose the optimal set of features, which included Global Horizontal Irradiation (GHI), Diffuse Horizontal Irradiance (DHI), Fixed Tilt Irradiance (FTI), and Tracking Tilt Irradiance (TTI). On the other hand, Song et al. (2022) employed Random Forest (RF) feature importance analysis and Pearson and Spearman correlation, concluding that Solar radiation is the most important feature. As also observed by Natarajan and Singh (2023), solar radiance/irradiance is shown to be the most correlated weather parameter with PV power in various studies. Besides solar radiance/irradiance, the most used input features, including studies that don’t apply feature selection techniques, are ambient temperature, wind speed (WS), and relative humidity (RH). A few studies still didn’t identify the variables from the original dataset that they ended up using as model input.

Regarding RQ2, which aimed to understand which ensemble methodologies were implemented in the proposed models for PV power forecasting, the papers were categorized by the type of ensemble, such as Stacking, Boosting, Bagging, and others. Appendix A., Table A.3 provides information about the proposed model for each study and the corresponding models used for comparison.

Only 11 of the studies proposed the Boosting method. For example, Abdellatif et al. (2022b) compared different tree-based ensemble models and concluded that the Adaboost model had the best performance. Lin (2023) used the Gradient Boosting Regressor (GBR) for the PV power forecasting of a solar plant located in Taiwan, but they focused on studying different hyperparameter optimization algorithms to improve ensemble learning models further. They compare the Differential Evolution Algorithm (DE), Jaya algorithm, Random Search, Particle Swarm Optimization (PSO), and Genetic Algorithm (GA), concluding that DE outperforms the other algorithms. Another article proposing a Boosting model is presented by Petrosian and Zhang (2024), which demonstrates the potential of LightGBM when compared to various other machine learning models.

Additionally, only a small portion (8) of the research articles utilized the Bagging ensemble. Visser et al. (2022) observed that the RF model performed well in terms of the error metrics when compared to the other 12 models tested, considering one or a combination of multiple PV systems. Another group of 8 articles used the weighted average/sum to ensemble the different base models. The weighted average was applied by Siriwardana et al. (2022) to ensemble two Bidirectional LSTM (BiLSTM) models, one using sky images as features and the other using past values of PV power. Sharma et al. (2022) used the same technique to ensemble the results of LSTM models applied to the decomposed series using the Maximal Overlap Discrete Wavelet Transform (MODWT).

Stacking was the most developed ensemble methodology, with almost half of the studies (25) proposing it. It is possible to highlight that two of the most used base learners were XGBoost and RF, models that are already based on ensemble algorithms. These base models were used in stacking models 9 and 8 times, respectively. The LSTM networks were also used 8 times as stacking base learners. The usage of the XGBoost and RF ensemble models as base learners enables us to take advantage of the XGBoost’s ability to capture data details and RF’s superior fitting ability (Abdellatif et al., 2022a). For example, Abdelmoula et al. (2022) used both models and a Multiple Linear Regression (MLR) as the base learners of their stacking model, and the linear Lasso model as meta-learner. The results highlighted the benefits of combining these models. Lateko et al. (2021) also proposed a stacking model, but in this case, they made use of an RNN to incorporate the results of an ANN, Support Vector Regression (SVR), Deep Neural Network (DNN), LSTM, and a Convolutional Neural Network (CNN). This model showed the best performance out of the benchmark model RF, the single learners and the RNN stacking of DNN, SVR, and ANN. Some studies choose the base learners judiciously. For example, Song et al. (2022), made a correlation analysis of the prediction errors of multiple single learners to select the four base models that would have low correlation, those models being XGBoost, Gated Recurrent Unit network (GRU), RF, and Least Squares SVM (LSSVM), which was also used as the meta-learner. Only one of the reviewed studies implemented a second-level stacking method (Zhou et al., 2023a), named Double Nested Stacking (DNS), using an SVR as the final meta-learner, and Gradient Boosting Decision Tree (GBDT), XGBoost, and SVR as base learner models, which outperformed the traditional stacking methodology.

The rest of the articles (8) used a diversity of other ensemble techniques, such as the voting ensemble, applied by Chakraborty et al. (2023) to combine the results of the models Light Gradient Boosting (LGB), ETR, XGBoost, and Histogram-based Gradient Boosting (HGB), showing a slightly better performance than the stacking model that was also tested. Markovics and Mayer (2022) compare various models (Appendix A., Table A.3., column "Comparison Models") and state that the ETR performs the best. They also study the effect of the training period on this model and recommend using at most one year of training data.

It is difficult to compare different studies' performances since the datasets used are different, as well as the forecasting horizon and the scale of the predicted variable. Nevertheless, the reviewed articles still compared the proposed models to other models considered as benchmark or reference (Appendix A., Table A.3., column "Comparison Models"), which already indicates the model's forecasting capability. It was observed that the most used error metrics to evaluate the performance of the models, were the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), the Coefficient of Determination (R^2), and the Mean Absolute Percentage Error (MAPE) (Appendix A., Figure A.1.). The normalized version of RMSE (nRMSE) and the Mean Squared Error (MSE) were also frequently used. A variety of other 17 metrics were applied, including the normalized version of MAE, among others.

As the computational time is a critical constraint in the real-time implementation of ML models (Rai et al., 2022), 10 of the reviewed studies also evaluated the computational time of the proposed models. For example, Massaoudi et al. (2021) compared the testing time of the proposed stacking model of ELM, ETR, K-nearest neighbor (KNN), and Mondrian Forest (MF), with meta-learner Deep Belief Network (DBN), showing it was also very satisfactory compared to the benchmark models. Wang et al. (2022b) studied the model's computational complexity, observing that the proposed linear regression ensemble of XGBoost and TabNet increases significantly, as expected, the training time when compared to the single models, but still has lower computational complexity than traditional deep learning models.

It is worth noting that, although ensemble methods in the reviewed research showed significantly better performance results than other models, in one study (Mohana et al., 2021), the RF and XGBoost models performed worse than models such as Polynomial Regression (PR). This was similar to what occurred in the study performed by Herrera-Casanova et al. (2024) where the RF was outperformed by the Bi-LSTM network. Additionally, Visser et al. (2022) proposed the RF when considering the error metrics, as it was the model with the best performance; however, this conclusion wasn't sustained when evaluating the economic implications of the forecast errors. Considering the economic revenue from forecast errors, the physical model was the one that performed the best. The authors showed in this case that the decision to adopt a specific model to forecast the day-ahead PV power depends on the objective of the operator.

2.3. Conclusions of the literature review

Recent research has been applying ensemble methods, which allow taking advantage of multiple models, showing the great capability of these methods to make a more robust PV power forecasting even with the unstable nature of this variable. It was observed that ensemble models generally performed better than other approaches, yielding more robust and stable models for PV power forecasting. A collection of 60 articles published during the last four years was reviewed in detail, allowing us to conclude that the most used ensemble technique for PV power forecasting is stacking, also highlighting that two of the most used base learners, RF and XGBoost, were also models based on ensemble algorithms.

Most of the selected studies focused on short-term forecasting, concluding that this occurs due to the policies of market operators in the energy sector that mostly trade in the short-term, specifically the day-ahead market for unit commitment (Rai et al., 2022). The use of features related to solar irradiance as input for the ensemble models is a common approach in the reviewed research, due to its high correlation with PV power output. Furthermore, the studies focused on single PV plant-level forecasting. Only a few studies aggregate data from multiple PV plants. Kim et al. (2021) utilized data from three solar power plants in different regions to train the model; however, the forecast is still conducted at the plant level. They followed this approach to increase the data, aiming to improve the model. Only one of the reviewed studies performed a regional forecast, but it doesn’t give enough detail on the ensemble approach followed. The error metrics RMSE and MAE were used in most of the papers to evaluate the performance of the forecasting models.

Rai et al. (2022) also concluded that the reviewed articles didn’t focus on computational time, which is a critical constraint for the real-time implementation of the solutions. It is essential to consider computational efficiency when implementing ensemble approaches, as they are expected to increase computational complexity due to the simultaneous execution of multiple models. Only 10 of the reviewed studies evaluated the runtime of the proposed models.

Additionally, there is a noticeable gap in research focusing on long-term forecasting. It is on this forecasting horizon that governments pay more attention when developing an energy policy (Zhang & Zhu, 2022). There is also a lack of studies in the reviewed collection regarding the generalization capabilities of the models proposed to be used in different locations, through transfer learning techniques or the usage of diverse datasets. The same conclusion was also reached by Ahmed et al. (2020), suggesting the need for a versatile forecasting approach that can be applied in different geo-climatic conditions. Only five of the articles studied the robustness of the models as they applied them to data from two different locations.

Considering these findings, there is a considerable research gap in the capabilities of ensemble models to forecast PV power on a national or multi-regional level. If proven effective, the short-term national ensemble forecast would support the integration of the national market into a single European electricity market by enabling efficient operation and energy planning. While exploring these approaches, the computational efficiency of these models should also be taken into account.

CHAPTER 3

Methodology

3.1. Dataset

The solar power data used as the target variable for the present research was provided by R&D Nester, and represents the hourly PV power generated in Portugal's mainland, measured in megawatts (MW), from January 1st, 2023, to June 6th, 2024, constituting a total of 12,552 records (Figure 3.1.).

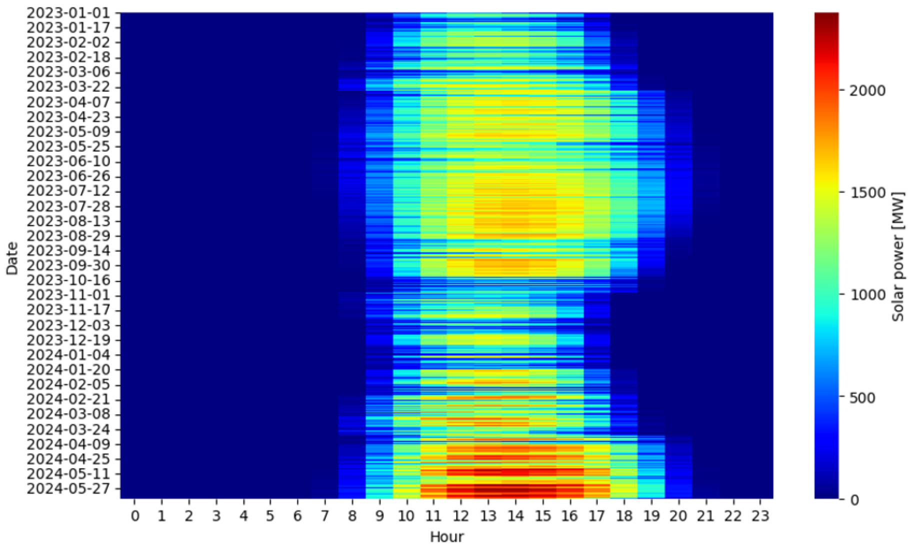


Figure 3.1. Hourly PV Power output in Portugal

3.1.1. Data Preparation

The Data Preparation and subsequent steps were carried out using the Python programming language and multiple of its open-source libraries.

In the first phase, the dataset was analyzed to assess the necessary actions for data preparation. Only small adjustments were required in order to have the data complete for the subsequent analysis and modeling phases. It was necessary to remove one duplicate record and fill in a missing value. None of those actions brought further complexity, since both of these cases occurred at nighttime, when solar power is inherently zero. Similar to what was done by Gaboitaolelwe et al., 2023, feature engineering was performed, where additional features were derived from the existing solar power date-time index. These variables were: Hour, Day of the week, Month, Quarter, and Season of the year (Table 3.1.). The Seasons variable was introduced into the models by using the one-hot encoding technique, where, for example, the variable "season_Spring" assumed the value "1" for the days between March 21st and June 20th.

Table 3.1. Time-Derived features used in the Forecasting Models

Time feature	Details
Hour	Hour of PV power generation (from 0 to 23)
Day of the week	Day of the week of PV power generation (0 represents Monday and 6 represents Sunday)
Month	Month of PV power generation (from 1 to 12)
Quarter	Quarter of PV power generation (from 1 to 4)
Seasons of the year	Season of the year of PV power generation (one-hot encoding variables: Spring, Summer and Autumn)

3.2. Forecasting models

To achieve the primary objective of this research, which is to investigate the potential of ensemble models for day-ahead PV power forecasting at the national level, various models were tested and compared against one another. This section provides a brief explanation of the models chosen for testing in this study, including those used as the baseline.

3.2.1. Baseline Models

Three models not categorized as ensemble models were used as Baseline, meaning they were used as a reference for comparison and to frame the performance of ensemble techniques. One of the baseline models is the traditional time series model SARIMA, the other was the Prophet model, and the last one was the Unobserved Components Model (UCM). The day-ahead forecast estimated by the baseline models was obtained through a direct multi-step approach.

SARIMA:

The ARIMA model is categorized as a Statistical model, as it is based on the statistical analysis of historical data to identify trends and patterns, which are then used to forecast future data. It is often used in solar power forecasting, according to Gaboitaolelwe et al. (2023). The SARIMA model is an extension of the ARIMA model that incorporates a seasonal component, making it suitable for predicting time series that exhibit periodic patterns, which is the case with solar power generation (Chen et al., 2023).

Since classical time series models can only be estimated in stationary data, the PV power series was examined using unit root and stationarity tests. As PV power exhibits a strong daily pattern, daily seasonality was removed by applying a seasonal difference with a period of 24 hours before conducting these tests. The unit root tests Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) (Hamilton, 1994), together with the Kwiatkowski–Phillips–Schmidt–Shin

(KPSS) stationarity test (Kwiatkowski et al., 1992), showed evidence that the seasonal differenced PV power series is stationary. Autocorrelation and Partial Autocorrelation function plots of the original and seasonally-differenced series are present in Appendix B (Figures B.1. and B.2.).

After finding evidence that one seasonal difference was sufficient to achieve stationarity, the *auto_arima* function was used to seek the SARIMA's optimal parameters that would minimize the chosen information criterion, which in this case was the *auto_arima's* default criterion Akaike Information Criterion (AIC). *Auto_arima* identified the model SARIMA(3,0,2)(0,1,1)$_{24}$ as the one best suited to our data, which produced uncorrelated residuals.

Prophet:

The Prophet model, developed by Facebook[1], was also chosen as a baseline. When it was first developed, this model was used to forecast the number of events created daily on Facebook, which constitutes a time series that contains several seasonal effects (Taylor & Letham, 2017). It is based on an additive decomposable time series model with three main components: trend, seasonality, and holidays. The Prophet can be viewed as a Generalized Additive Model (GAM), with the estimation of each component following a Bayesian framework. It has been applied to energy-related topics, such as energy demand forecasting. For example, Chaturvedi et al. (2022), in a comparative study, concluded that the Prophet was the model with the best performance. Its capabilities were also proven by Baloch et al. (2025) in solar irradiance forecasting. It has also been used for solar power forecasting, but not always showing the best results, as it was outperformed by the XGBoost model in monthly and weekly forecasting (Gupta et al., 2022). As this model is robust to outliers and works best on time series with strong seasonal effects, it was selected to be used as a baseline model in this research. A set of different hyperparameter values was tested, and the parameters that produced the best results were obtained, as shown in Table 3.2.

Table 3.2. Prophet model hyperparameter values tested and corresponding best values

Parameter	Values tested	Best Values
changepoint_prior_scale	0.001, 0.005, 0.01, 0.1	0.001
seasonality_mode	Additive, Multiplicative	Multiplicative
weekly_seasonality	True, False	True
fourier_order (applicable when weekly_seasonality is True)	3, 5, 7, 9	9
daily_seasonality	True	True

[1] https://facebook.github.io/prophet/docs/quick_start.html#python-api

UCM:

The UCM model was proposed by Harvey in 1989 as an alternative to the ARIMA model, since it doesn't make assumptions regarding the stationarity or distribution of the time series, as cited by Atwan (2022). The UCM decomposes a time series (which can contain multiple seasonal patterns) into components, such as trend, seasonal, and cycle. Even though both UCM and Prophet decompose a time series into various components, UCM considers them to be stochastic (random) processes within a state-space framework and estimates them through the Kalman Filter technique (Harvey & Koopman, 2009), whereas the Prophet model considers them to be deterministic processes, estimated through predefined functions (Taylor & Letham, 2017).

Since PV power has a strong daily seasonality pattern and the model doesn't require extensive feature engineering, it was also decided to be tested and used as a baseline. Similar to the Prophet model, different hyperparameter values were tested, and the parameters that yielded the best results were obtained (Table 3.3.).

Table 3.3. UCM model hyperparameter values tested and corresponding best values

Parameter	Values tested	Best Values
level	dtrend, llevel, ntrend	llevel
stochastic_freq_seasonal	True, False	True
freq_seasonal harmonics (period=24)	1, 2, 3, 4	4
autoregressive	4, 5	5

3.2.2. Ensemble Models

As the primary focus of this research is to determine the capabilities of ensemble models for day-ahead solar power forecasting at a national level, models with various ensemble methodologies were tested. The ensemble models - XGBoost, RF, LightGBM, and a Stacking model - were implemented in this research and are described in this section.

XGBoost:

The XGBoost model is an improved and scalable implementation of the Gradient Boosting (GB) method (Abdellatif et al., 2022a). The GB technique builds a powerful model by iteratively combining different decision tree models, called weak base learners. The way the XGBoost works is by using the original data as input for the first base learner. After that, in a second iteration, a new model (the second base learner) is fitted on the residuals obtained from the first model to enhance its learning capability. This procedure of residue buildup is continued until the specified requirements are reached. The final predictions are obtained by aggregating the results of all the base models (Abdellatif et al., 2022a). In XGBoost, the decision trees are built using a maximum depth strategy, and then are pruned in reverse order, based on the gain values of each branch.

This model has consistently shown strong potential in diverse subjects, including PV power forecasting, as proven by Zhou et al. (2023b). For this reason, it was chosen in this research to assess its capabilities in predicting an aggregated multi-regional level of PV power generation. This model can be categorized as a Boosting ensemble method, since it attempts to reduce the model's bias by sequentially training numerous models to enhance each previously created model.

LightGBM:

Similar to XGBoost, the LightGBM model is also based on the GB decision tree algorithm (Ke G. et al., 2017). The model incorporates two novel techniques, one of which is Gradient-based One-Side Sampling (GOSS). The GOSS technique performs random sampling on instances with small gradients (namely, those that the model already predicts with high accuracy), while retaining all instances with large gradients. The selected small gradient instances are amplified by a constant, putting more focus on the under-trained (i.e., large gradient) instances (Ke G. et al., 2017). The second novel technique used in LightGBM is Exclusive Feature Bundling (EFB), which aims to effectively reduce the number of features by combining mutually exclusive ones. Another difference between LightGBM and other boosting algorithms is that the decision trees are split leaf-wise instead of level-wise.

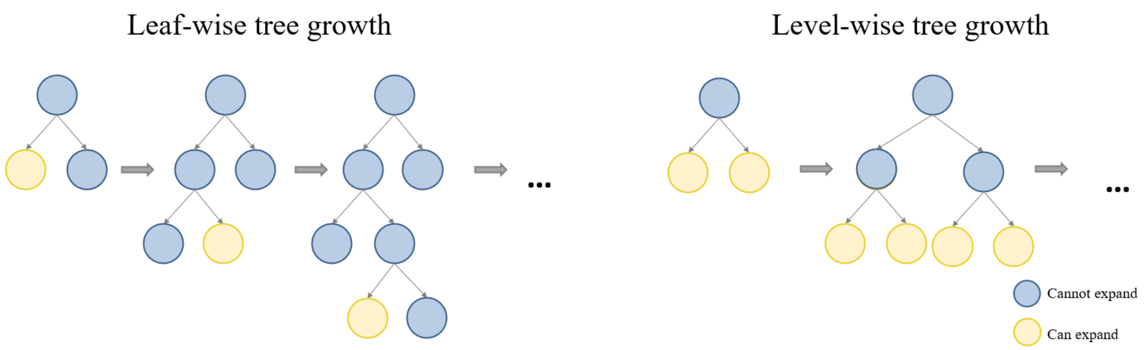


Figure 3.2. LightGBM's leaf-wise learning process vs. level-wise tree growth
(Adapted from Thai, 2022)

This model gets the name "Light" by being the faster and more efficient version of GB, since its techniques focus on computational efficiency, making it appropriate for the current context, given that PV power forecasting must be timely and efficient.

Random Forest:

The Random Forest (RF) model, in contrast to the two ensemble models described above, is a bagging ensemble method, since it trains the different base learners in parallel and simultaneously. In RF, the base models are also decision trees. As explained by Banik & Biswas (2023), the RF model first starts by creating a random subset of samples from the dataset using the bootstrapping technique. Then, as a second step, a random subset of features is chosen to identify the optimal split for each node of the decision tree. The previous two steps are repeated to create multiple distinct decision trees. The final predictions of the RF are the average of all the individual decision tree predictions.

In summary, the RF is based on bootstrap aggregation and random feature selection. These techniques work toward the diversity of the decision trees, reducing variance and correlation between them. This leads the model to an increase in generalization capabilities of the whole model, making it less prone to individual influence and avoiding overfitting. For these reasons and given that they can handle and capture complex non-linear relations, they are ideal for PV power forecasting.

Stacking model:

Stacking is an ensemble learning approach in which the predictions of several base learners are stacked to train a new model, called a meta-learner. The base learners are trained using the original dataset, and their predictions are compiled into a new dataset. This new dataset is used to train the meta-learner, which produces the final predictions (Abdellatif et al., 2022a). The meta-learner aims to correct the errors produced by the base learners to create more accurate final predictions. In this research, the base learners chosen were XGBoost and RF due to their different ensemble algorithms and individual results obtained (present in the next chapter). These were also frequently used as base learners in the stacking models in the related studies. The meta-learner chosen was linear regression, given its simplicity. Figure 3.3. gives an overview of the Stacking model tested in this research.

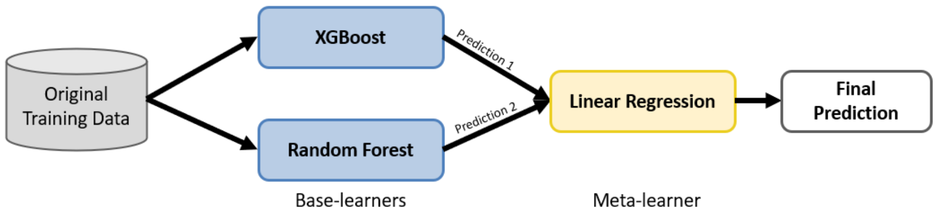


Figure 3.3. Overview of the Stacking model

3.3. Feature Engineering and Data Preprocessing for Ensemble Models

As ensemble models are ML models not inherently designed to handle sequential data, lag features of the PV power time series were created and included in the feature set for model training. Including these features in the input of the ensemble models allows them to capture the temporal patterns in the solar power data. To predict one day ahead of PV power output, two different numbers of lags were tested to evaluate the impact of historical information on the forecasting accuracy: 24 lag features, Solar t-1, Solar t-2, ..., Solar t-24, representing the previous day of solar power generation, and 48 lag features, representing the two last days, with a similar structure.

To assess how the size of the training dataset influences the models' learning abilities and forecasting results, two experiments were conducted using different training sizes: one with 8736 hours, corresponding to around one year's worth of historical data, and another with 12360 hours, corresponding to around one and a half years of historical data. Further details on training and test subset split are provided in section 3.7.

As can be seen in Figure 3.4., the Solar Power production shows a continuous increasing trend starting around December 2023, with the monthly averages of 2024 being higher than those of the previous year. Moreover, the maximum value registered in 2024, present in the dataset (2374 MW, May 27th), is significantly higher than the maximum value registered in the whole year of 2023 (1800 MW, September 23rd). This increase may not just be related to solar energy conditions during the analyzed period, but also dependent on the installed capacity in Portugal's mainland, which, according to information provided by DGEG in the "*Estatísticas rápidas das renováveis*" monthly reports, has registered a substantial and linear increase since May of 2023 (Figure 3.5.). For this reason, and given that solar power output has a maximum theoretical value that changes over time, an adaptation of the min-max scaling technique was tested. The ensemble models explored in this study are generally robust to the scale of input features, as they are based on decision trees, which, as mentioned by Jobayer et al. (2023), do not require normalization or scaling. Nevertheless, a scaling approach was tested to investigate whether the models would still benefit from it. It could also improve convergence speed and ensure that feature importance estimates are not biased by the magnitude of the variables. The solar power was divided by its corresponding monthly installed capacity data.

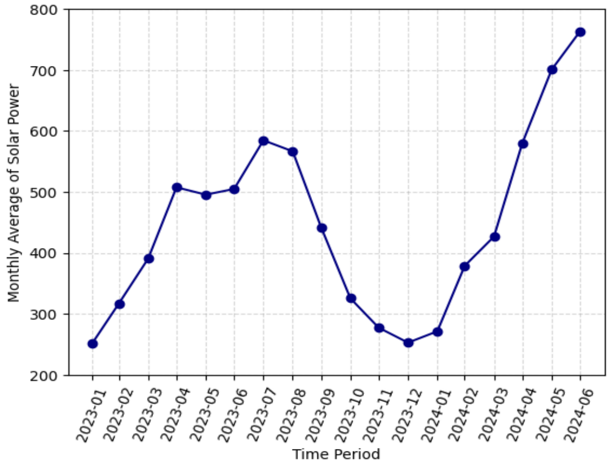


Figure 3.4. Monthly Average of the PV power series

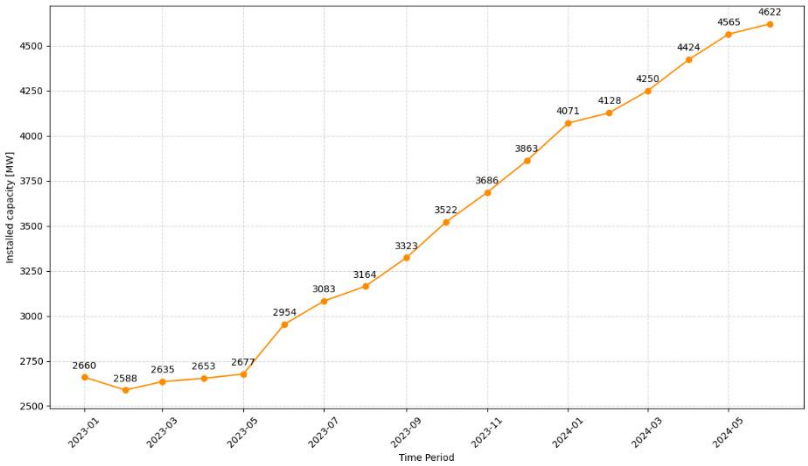


Figure 3.5. Monthly installed Solar Power Capacity in Portugal's mainland (Data Source: DGEG)

A summary of all the tested configurations is presented in Table 3.4.

Table 3.4. Overview of the tested configurations in the Ensemble models

Configuration Parameter	Values
Lags	24, 48
Training size (hours)	8736, 12360
Scaling with monthly maximum installed capacity	No, Yes

3.4. Hyperparameter tuning

The results of the ML model depend on the set of hyperparameters chosen during the training stage. Due to the high variety of possible hyperparameter sets in supervised learning methods, hyperparameter optimization techniques are crucial in choosing values that lead to more accurate models, while simplifying and speeding up the data processing pipelines (Jamieson & Talwalkar, 2015).

This research initially employed one of the most common hyperparameter tuning methods called Randomized Search with cross-validation. However, since hyperparameter tuning was to be performed for each model configuration and due to the extensive number of hyperparameters from every single ensemble model, this tuning method required excessive computational time. This situation led to a search for a more efficient tuning alternative.

To mitigate the computational burden, instead of the traditional Randomized Search, the Halving Randomized Search with cross-validation was implemented because of its improved efficiency. This tuning approach utilizes Successive Halving, proposed by Jamieson and Talwalkar (2015) for the context of hyperparameter optimization. It begins by allocating a small amount of resources, such as the number of training samples or the number of epochs, to train and evaluate the different hyperparameter combinations. Then, a certain portion of the worst-performing configurations of hyperparameters is discarded, and on the following iterations, more resources will be allocated to the remaining configurations. This process is repeated until only one configuration is left or there are no more resources available. For the current research, the Halving Random search was implemented with 200 candidates (i.e., 200 combinations of hyperparameter values) for the first iteration, and 1/3 of the best candidates were selected in each subsequent iteration. The resource used for allocation in each round was the number of training samples. The Halving search was performed using 3-fold cross-validation, with a random state of 42 and mean squared error as the scoring method. In Table 3.5., the different parameter grids for the hyperparameter tuning of each model can be found. For the stacking model, both the XGBoost and RF grids were used.

Table 3.5. Hyperparameter tuning grids for the Ensemble models

Model	Tuning Grid
XGBoost	**Objective**: ‘reg:squarederror’ **Learning_rate**: i * 0.01 for i in range(3, 10) **Max_depth**: i+2 for i in range(5, 17) **N_estimators**: 70, 100, 120, 150, 170, 180, 200 **Subsample**: 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 **Colsample_bytree**: 0.5, 0.6, 0.8, 1.0 **Gamma**: 0.5, 0.7, 1, 1.5, 2.0 **Reg_alpha**: 0.5, 0.7, 1 **Reg_lambda**: 0.5, 1, 1.5, 2.0
LightGBM	**Boosting_type**: ‘gbdt’, 'dart', 'goss' **Learning_rate**: i * 0.01 for i in range(1, 10) **Max_depth**: 3, 4, 5, 6, 8, 10, 12, 15, 17, 20, 22, 25 **Num_leaves**: 5, 10, 20, 30, 40, 50, 60, 70, 80, 100 **Feature_fraction**: 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 **N_estimators**: 10, 30, 40, 50, 60, 70, 80, 100, 120, 150 **Metric**: ‘rmse’

RF	**Max_depth**: i+2 for i in range(5, 17) **N_estimators**: 50, 70, 100, 120, 150, 170 **Min_samples_split**: 2, 5, 10, 15 **Min_samples_leaf**: 1, 2, 4, 6 **Max_features**: 'sqrt', 'log2', 0.2, 0.5, 1, 2 **Bootstrap**: True, False **Accuracy_metric**: ‘mse’

Considering all the combinations of configurations used in this research (present in Table 3.4.), the best parameters were obtained using Halving Random Search for hyperparameter tuning for each one.

3.5. Forecasting Strategies

As previously mentioned, the main objective of this study is to perform one-day-ahead PV power forecasting. This objective can be accomplished by predicting 24 hours simultaneously in a direct multi-step forecast, or by dividing the forecasting horizon into shorter steps, such as 1, 2, 3, 4, 6, 8, or even 12 hours, and building the full-day forecast iteratively. In this research, both approaches were implemented to evaluate whether the models would benefit from dividing the one-day forecast horizon into smaller steps.

The recursive forecasting strategy was implemented in cases where the 24 hours were split into shorter intervals. With this strategy, predictions are generated sequentially, and each prediction steps are then used as input to predict the subsequent time steps until the whole day is forecasted. As an example, Figure 3.6. illustrates the recursive forecast approach applied in the 2-step ahead forecast.

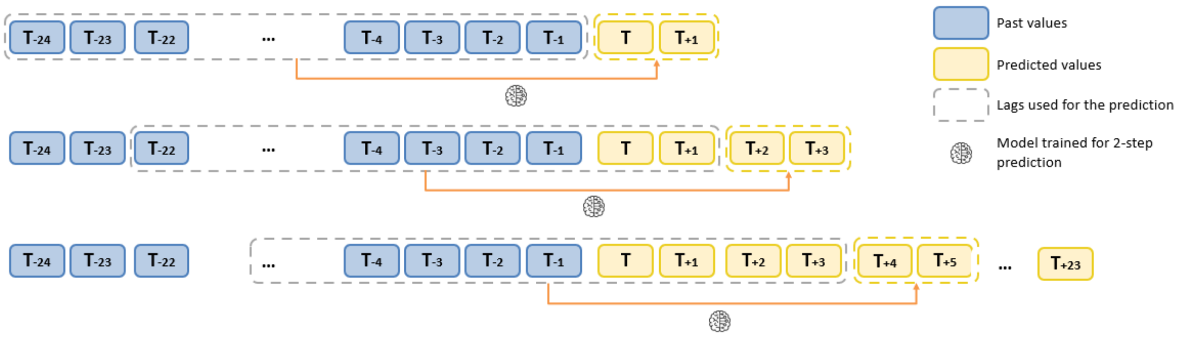


Figure 3.6. Diagram of recursive multi-step forecasting (adapted from Skforecast, “Recursive multi-step forecasting”, n.d.)

As RF and LightGBM models do not inherently support multi-step forecasting, the *MultiOutputRegressor* function from *scikit-learn’s* multioutput package was used to adapt the models for this purpose.

3.6. Tune Stage

During the exploratory data analysis, as expected due to its physical relation with daylight, it was observed that the PV power generated between 11 p.m. and 6 a.m. was consistently zero. The absence of solar power generation during nighttime is a physical constraint that the ML models do not intrinsically account for. This means that even though not present in the data, the models could still produce non-zero or even negative predictions during the mentioned time period, unless properly guided through feature engineering or post-processing. Although the models were trained on clean data and provided with time-related features, namely the hour of the day, they occasionally produced non-zero or even slightly negative predictions during that period, when solar power is known to be zero. This behavior highlights a limitation of purely data-driven models, which may fail to strictly follow domain-specific deterministic constraints. Therefore, similar to what was done by Hokmabad et al. (2023), a tuning stage was incorporated into the pipeline, where all predictions between 11 p.m. and 6 a.m. were set to zero to ensure physical consistency and produce realistic model outputs.

3.7. Ensemble model validation

By evaluating performance results across multiple train-test splits, cross-validation techniques enhance the robustness and reliability of the evaluation results, providing a more accurate estimate of the models' generalization capability. For this reason, a cross-validation approach was implemented.

Taking into account that time series data is sequential, models can only be tested on future data points. Only past information can be used for training purposes. To simplify this research, the testing set used to assess the model results was defined as the last six days of the dataset, namely from the 1^{st} to the 6^{th} of June, 2024. The models were trained using two different training set sizes: the previous year and the previous year and a half. In this second case, all the available data, excluding the testing dataset, was used for model training. Since the objective is one-day-ahead forecasting, a 6-fold cross-validation strategy was used to assess the final model results, where each day of the testing dataset was used as a fold. To maintain consistent learning conditions and a fair comparison, the training set size was the same for each testing day. This allows for a more accurate assessment of the impact of model configurations, since it excludes the influence of varying training data volumes on the model performance. The final results are the simple average of each fold. An illustration of the Cross-validation approach can be visualized in Figure 3.7.

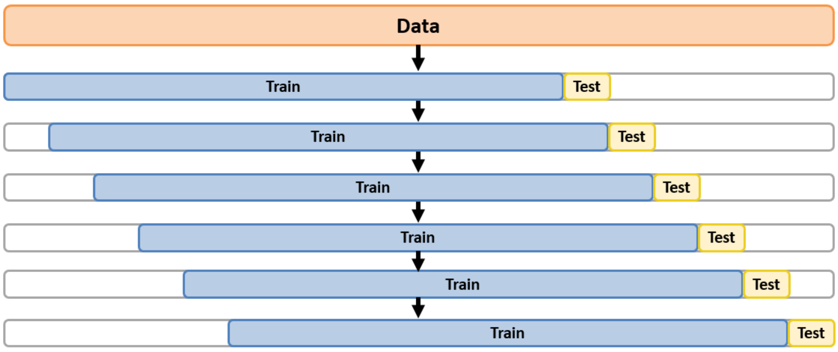


Figure 3.7. Cross-validation method used for the PV power forecasting

3.7.1. Evaluation Metrics

After the training phase, it is necessary to define metrics that enable the evaluation of the model's performance and the assessment of the best-performing one in PV power forecasting. Three of the most commonly used metrics in the related works were employed in this research, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The MAE gives the average of the absolute differences between y_t and $\hat{y}_t$, which are the observed and forecasted values at time *t*, respectively (Equation 3.1). In MAE, equal weight is given to all the differences, making it robust to outliers and suitable for the high variability scenario of solar power.

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t| \qquad (3.1)$$

The MAPE gives the mean of the absolute percentage errors (Equation 3.2), which is scale-independent. Since the dataset includes zero values that were used in the forecasting process, these observations were excluded from the MAPE calculation.

$$MAPE = \frac{100}{T} \sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \qquad (3.2)$$

The RMSE is calculated as the square root of the average squared differences (Equation 3.3), which gives higher weights to larger errors and makes it sensitive to outliers. This means that a smaller RMSE might indicate that the model made fewer larger errors.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2} \qquad (3.3)$$

Both the RMSE and the MAE will also be presented in a normalized version by dividing them by the peak of the Solar PV power observed in the dataset. This leads to better comparability with other analyses by providing scale-independent error metrics.

3.8. Feature exclusion and performance reassessment

3.8.1. Explainable AI

After estimating all the models, an explainable AI (XAI) technique was employed to gain insights into the impact of each feature on the model. The method used was SHAP (SHapley Additive exPlanations), a game-theoretic approach that enables the interpretation of the contribution of each feature in the model's output. Given the multiple configurations involved in this research, for simplification purposes and to not add variability from other model setups, it was decided to perform this feature impact analysis only on the tuned version of the one-step ahead XGBoost model, in which one year of data in original scale was used for model training with 24 lags to predict the day-ahead. This analysis enables the identification of the most and least impactful features, as well as a more accurate interpretation of the model's behavior.

3.8.2. Performance reassessment

The least impactful features obtained through the SHAP analysis were excluded from the feature set. All the models were retrained and reevaluated to assess whether their performance would improve with a simplified and reduced feature set. For a fair comparison with the previous models, trained with the whole feature set, and to isolate the effect of excluding low-impact features, this stage was performed without further hyperparameter tuning.

3.9. Inclusion of Weather Variables

As there is a natural dependency between weather and solar power output, one of the objectives of this study was to explore the inclusion of meteorological data for PV power forecasting, with the aim of possibly enhancing forecast performance. The lack of information at the PV plant level adds complexity to the integration of weather data. Since solar power forecasting is conducted at the national level, there is a need to incorporate weather variables on the same scale. To achieve this objective, the following methodology was implemented: in the first stage, the coordinates of the photovoltaic plant with the highest installed capacity, and the total installed capacity were collected for each district from the Map webpage of Endogenous Energies of Portugal[2]; using the registered coordinates per district in the previous step, the hourly weather data was retrieved from the Weather Query Builder page of VisualCrossing[3]. On the VisualCrossing platform, three weather stations were selected as the maximum number of stations to be used for interpolating the historical observations. The weather data collected

[2] https://e2p.inegi.up.pt/
[3] https://www.visualcrossing.com/weather-query-builder/

includes temperature (ºC), humidity (%), precipitation (mm), wind speed (km/h), wind direction (degrees), pressure (mb), cloud cover (%), and solar radiation (W/m^2). After the data was collected, the total installed capacity per district was used as weights to build the weighted average of each weather variable, which is then fed into the models.

All the ensemble models were evaluated following both the univariate approach, where only the solar power and time-related features were used, and the multivariate approach, where weather data was also included in the modeling process. Following the univariate approach, each model was estimated in a total of 384 settings, considering all combinations of different configurations present in Table 3.4., the number of steps ahead, and 6 different feature exclusion scenarios (specified in section 4.2. of the next chapter) based on the results of the Explainable AI technique. The multivariate approach comprised a total of 96 different estimations per ensemble model, considering the same configurations, but only including the use of 24 lags of each variable (models were not trained using 48 lags) and 4 different feature set scenarios (described in section 4.3.).

3.10. Development Environment

This research was conducted in Google Colab's virtualized cloud environment to make use of its machine resources. For most of the tasks, mainly model estimation and evaluation, the Tesla T4 GPU with 16GB GDDR6 memory and 2.560 CUDA cores was chosen, since it combines relatively low cost with enhanced performance for machine learning training tasks.

CHAPTER 4

Results and Discussion

In this chapter, the results obtained by the ensemble models, following the methodology described previously, are presented and analyzed. The findings are then discussed, considering the study's objectives, highlighting the patterns, strengths, and limitations found during model evaluation. This chapter is divided into two main sections: the univariate approach and the multivariate approach.

4.1. Baseline model results

The results of the ensemble models were validated and compared with the performance of the baseline models, which served as the reference. Table 4.1. summarizes the average performance of the Baseline models in the day-ahead prediction of PV power on the testing dataset (similar to the subsequent tables). SARIMA was the best-performing baseline model, followed by UCM, which was significantly worse than SARIMA only when considering the MAPE metric. The Prophet model showed substantially worse performance than the other baseline models, indicating that it is not suitable for PV power forecasting at Portugal's national level.

Table 4.1. Average performance results of Baseline models for direct day-ahead predictions

Model	RMSE	MAE	MAPE	nRMSE	nMAE
$SARIMA(3,0,2)(0,1,1)_{24}$	103.87	65.93	13.36%	4.38%	2.78%
UCM	104.31	68.41	27.51%	4.39%	2.88%
Prophet	151.58	99.16	21.06%	6.39%	4.18%

Note: each baseline model was estimated for direct day-ahead forecasting, with training on around a year and a half of data preceding the six days of the testing set.

4.2. Univariate approach

In the univariate forecasting approach, the results were estimated using only historical values of solar power and time-related features, such as hour, month, quarter, season of the year, and day of the week. This approach provides reference results regarding the predictive capabilities of each model when no additional input features are included.

As mentioned in the methodology section, the performance of the ensemble models was evaluated under various configurations. The analysis comprised the assessment of the influence of the number of lags utilized, comparing the use of 24 to 48 previous hours as feature input. It also included a comparison of two different training data lengths (around one year and one and a half years) and the use of a scaling technique. The 24-hour forecasting horizon was also predicted using two different approaches: the multi-step direct approach and the division into different numbers of steps ahead. In this last case, the day-ahead forecast was built recursively.

Utilizing the best parameter values obtained with hyperparameter tuning in the one-step ahead XGBoost model with no scaling, one year of training data, and 24 lags included in the input features, the SHAP explainable AI method was applied to assess an estimate of the overall features' impact in Solar Power forecasting (XGBoost hyperparameter values used in the SHAP analysis are listed in Table 4.2.).

Table 4.2. XGBoost hyperparameters used for the SHAP analysis (univariate approach)

Model	Parameter	Value	Parameter	Value
XGBoost (Steps ahead=1 Training_size=8736 Lags=24 Scaling=No)	Subsample	0.50	Gamma	0.50
	Reg_lambda	0.50	Max_depth	8
	Reg_alpha	0.50	Learning_rate	0.06
	Objective	Reg:squarederror	Colsample_bytree	0.80
	N_estimators	170	Seed	42

In Figure 4.1., we can observe the top 20 most impactful features, allowing us to conclude that the lags representing the previous hour of generated power (Solar t-1) and the same hour of the previous day (Solar t-24) were the features with the most significant impact in the PV power output predictions, highlighting the daily cycle that this target variable demonstrates. These variables exhibit the expected behavior, as lower values result in smaller predictions. After the variables Solar t-22, Solar t-2, and Solar t-23, the remaining variables present in the figure have much smaller effects. Moreover, it is possible to conclude that the variables Day of week, Quarter, and Season of the year are the least important, since they do not appear in the SHAP Beeswarm plot.

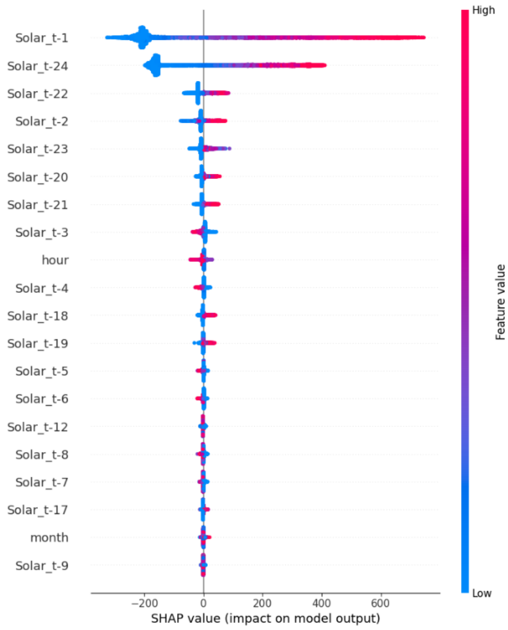


Figure 4.1. SHAP Beeswarm plot (univariate approach)

With these conclusions, all the models were trained again using the best parameters already obtained through hyperparameter tuning, but excluding the least important variables to reassess the performance metrics and see if any benefits would arise from simplifying each model. Several variable-exclusion scenarios were tested, including the individual removal of the variables Day of week, Quarter, Season, a combined exclusion of Seasons and Day of week, and a combined exclusion of Seasons, Day of week, and Quarter.

4.2.1. XGBoost

Regarding the ensemble models, XGBoost was the first to be studied (Appendix C, Tables C.1. to C.6. contain several results that will be mentioned and discussed in the next paragraphs).

When comparing the use of the previous two days (48 lags) instead of one day (24 lags) of PV power output to predict one day-ahead, there doesn't seem to exist evidence that using a higher number of lags significantly improves the forecasting performance of the XGBoost models. It was observed that from all the configurations compared between using 48 instead of 24 lags (192 configurations), it had an improvement in around 42% and 40% of them, considering RMSE and MAE, respectively. The overall impact was adverse, but doesn't seem significant, since it increased on average the RMSE and MAE metrics by 0.36 and 0.15 points, respectively (Table C.4.). Nevertheless, it improved the MAPE metric on average by 1.11 percentual points, meaning it slightly worsened the absolute errors, but improved the relative error metric. This could indicate that the model improved at forecasting smaller values with the use of more lags, as MAPE assigns greater weights to errors in smaller values.

The XGBoost models yielded better results when using variables in their original scale, as using monthly installed capacity to scale solar power worsened the error metrics in the majority of the tested configurations. Overall, only 12% of the results from the compared configurations of XGBoost improved with this scaling technique, considering RMSE (Table C.5).

Regarding the training dataset size, using a higher number of records to train the XGBoost, more specifically, around one and a half years (12360 hours) instead of one year (8736 hours), impacted the model performance, improving the RMSE by around 7.93 points, the MAE by around 5.05 points, and the MAPE by around 2.26 percentual points. More than 80% of the tested configuration error results improved by using a larger training set (Table C.6.).

Table 4.3. presents the best model configurations of XGBoost, when no variable was yet excluded from the model training. It allowed us to conclude that the best-performing XGBoost configurations to forecast one day-ahead without feature exclusion comprise the use of around one and a half years of training data with no scaling, and predicting the 24 hours recursively

with 3 steps in each iteration. The first and second rows of the table include the model configurations that yielded the best RMSE, and the best MAE and MAPE, respectively.

Table 4.3. XGBoost best model configurations without feature exclusion (univariate approach)

Steps ahead	Training size	Lags	Scaling	RMSE	MAE	MAPE	nRMSE	nMAE
3	12360	48	No	84.87	53.49	12.74%	3.58%	2.25%
3	12360	24	No	90.49	52.74	12.23%	3.81%	2.22%

According to the results shown in Table 4.4., excluding certain variables for model training, such as Day of the week or Seasons, led to better forecasting performance of XGBoost. In this way, we can conclude that the best RMSE (first row of the table) was obtained by training the model with 12360 hours, using the previous 48 hours without scaling, excluding the Day of week variable, and forecasting the next day using a recursive forecast with 3-step iterations (hyperparameter values that led to this result are shown in Table 4.5.). This model presented overall good results under the three different metrics. The second and third rows of the table contain the XGBoost configurations with the best MAE and MAPE, respectively, which also exhibited overall good performance results.

Table 4.4. XGBoost best model configurations with feature exclusion (univariate approach)

Steps ahead	Training size	Lags	Scaling	Excluded variables	RMSE	MAE	MAPE	nRMSE	nMAE
3	12360	48	No	Day of week	82.13	49.84	12.10%	3.46%	2.10%
4	12360	24	No	Seasons	82.47	47.70	11.61%	3.47%	2.01%
3	12360	24	No	Seasons + Day of week + Quarter	84.48	49.33	11.51%	3.56%	2.08%

Table 4.5. XGBoost best-performing hyperparameters (univariate approach)

Model	Parameter	Value	Parameter	Value
XGBoost (Steps ahead=3 Training size=12360 Lags=48 Scaling=No)	Subsample	0.5	Gamma	0.5
	Reg_lambda	0.5	Max_depth	11
	Reg_alpha	1.0	Learning_rate	0.03
	Objective	Reg:squarederror	Colsample_bytree	0.6
	N_estimators	200	Seed	42

4.2.2. Random Forest

In the experiments conducted using the RF models (Appendix C, Tables C.7. to C.12.), contrary to what was observed in XGBoost, there seems to be evidence that using 48 instead of 24 lags to predict one day-ahead improves forecasting performance (Table C.10.), improving 58%, 52% and 58% of all the compared model configurations (192), considering RMSE, MAE and MAPE, respectively.

The scaling technique also worsened the model's performance in the majority of cases, only improving around 4% of the RF configurations, taking into account RMSE. It improved around 38% of all configurations when considering MAPE (Table C.11.).

In RF models, it was observed that using a higher training size only improves around 24% and 26% of all tested experiments regarding RMSE and MAE, and improves around 66% regarding MAPE (Table C.12.). This suggests that the model improved at forecasting smaller values with the use of a larger training set. These results diverge from XGBoost, which benefited from a larger training dataset.

The best-performing RF model configurations without feature exclusion are summarized in Table 4.6. The configuration present in the first row represents the most effective across all performance metrics. The second and third configurations, although not the best performing in any of the metrics, are still a robust choice for PV power forecasting, as they showed a balanced performance across the different metrics.

Table 4.6. RF best model configurations without feature exclusion (univariate approach)

Steps ahead	Training size	Lags	Scaling	RMSE	MAE	MAPE	nRMSE	nMAE
24	8736	24	No	91.23	56.13	12.13%	3.84%	2.36%
4	8736	48	No	92.56	56.37	12.47%	3.90%	2.37%
12	8736	24	No	92.18	56.84	13.47%	3.88%	2.39%

Table 4.7. presents the top-performing RF configurations after excluding the previously selected sets of features. The first row contains the RF model with feature exclusion that achieved the best MAE and MAPE, which occurred when excluding the quarter variable (hyperparameter values are presented in Table 4.8.). The second row includes the model configuration with the best RMSE, obtained excluding the seasons variable. Overall, these two configurations yielded the best performance for the Random Forest model.

Table 4.7. RF best model configurations with feature exclusion (univariate approach)

Steps ahead	Training size	Lags	Scaling	Excluded variables	RMSE	MAE	MAPE	nRMSE	nMAE
24	8736	24	No	Quarter	88.95	54.38	11.81%	3.75%	2.29%
12	8736	24	No	Season	87.03	55.23	13.19%	3.67%	2.33%

Table 4.8. RF best-performing hyperparameters (univariate approach)

Model	Parameter	Value	Parameter	Value
RF (Steps ahead=24 Training size=8736 Lags=24 Scaling=No)	N_estimators	50	Max_depth	14
	Min_samples_split	2	Bootstrap	True
	Min_samples_leaf	2	Accuracy_metric	'mse'
	Max_features	0.50	Seed	42

4.2.3. LigthGBM

LightGBM models (Appendix C, Tables C.13. to C.18.), when using 48 lags instead of 24, demonstrated better results in 43% and 50% of all compared configurations (192), regarding RMSE and MAE, which doesn't show strong evidence that using a higher number of lags impacts the model performance (Table C.16.).

Similar to the previous models, the tested scaling technique also didn't enhance the forecasting performance of LightGBM (Table C.17.). Nevertheless, the configurations that are enhanced by data scaling are concentrated in the one-step-ahead models (Table C.14.).

Regarding the training dataset size, although not as obvious as in XGBoost, the LigthGBM benefited from a larger training dataset in 60% and 58% of the compared experiments, considering RMSE and MAE, respectively (Table C.18.). It had on average a reduction of around 2.98 points in RMSE, 1.42 points in MAE, and almost no impact in MAPE (0.04 percentual points), when comparing to a smaller training dataset.

The LightGBM configurations in Table 4.9. are the best performing in terms of RMSE and MAE, and in terms of MAPE, in the first and second rows, respectively, when all variables are included.

Table 4.9. LightGBM best model configurations without feature exclusion (univariate approach)

Steps ahead	Training size	Lags	Scaling	RMSE	MAE	MAPE	nRMSE	nMAE
8	12360	24	No	91.34	56.79	14.22%	3.85%	2.39%
3	12360	48	No	100.51	61.36	13.48%	4.23%	2.58%

In contrast to what was observed in XGBoost and RF, the exclusion of the defined feature subsets didn't appear to impact LightGBM's forecasting ability significantly. Top-performing model configurations with feature exclusion are present in Table 4.10. (Hyperparameter values are present in Table 4.11.).

Table 4.10. LightGBM best model configurations with feature exclusions (univariate approach)

Steps ahead	Training size	Lags	Scaling	Excluded variables	RMSE	MAE	MAPE	nRMSE	nMAE
4	8736	24	No	Seasons + day of week	90.60	55.89	13.64%	3.82%	2.35%
12	8736	24	No	Seasons + day of week	98.44	60.79	13.15%	4.15%	2.56%

Table 4.11. LightGBM best-performing hyperparameters (univariate approach)

Model	Parameter	Value	Parameter	Value
LightGBM	Num_leaves	70	Learning_rate	0.06
	N_estimators	120	Feature_fraction	1.0
	Max_depth	22	Boosting_type	'gbdt'

Note: both best-performing configurations of LightGBM converged to the same hyperparameter values

4.2.4. Stacking Model

Based on the evaluation of the individual model results, XGBoost was the best-performing model across different experimental setups. Random Forest had similar results to LightGBM in the scenario where all the variables were included for model training, but demonstrated superior performance when excluding a defined set of features. To fully utilize model diversity, and in an attempt to enhance predictive performance and stability while reducing the risk of overfitting, XGBoost and RF were combined as base learners in a stacking ensemble, with a Linear Regression as the meta-model to combine their outputs.

The stacking model (Appendix C., Tables C.19. to C.24.) had a similar behavior to XGBoost, given that there does not seem to be evidence that using a higher number of lags significantly improved the model (Table C.22.), with only around 48% of all compared configurations improving, when considering RMSE and MAE. The scaling technique had a negative effect on forecasting performance (Table C.23.), with the configurations that improved by this technique being concentrated on the one and two-steps ahead models (Table C.20.). Lastly, using a larger training dataset size enhanced the majority of the model configurations (around 65% of configurations regarding RMSE) (Table C.24.).

It was observed that the Stacking of XGBoost and RF did not yield better results than the single models in the scenario without feature exclusions. Best-performing configuration across the three metrics without feature exclusion is present in Table 4.12.

Table 4.12. Stacking best model configurations without feature exclusion (univariate approach)

Steps ahead	Training size	Lags	Scaling	RMSE	MAE	MAPE	nRMSE	nMAE
24	8736	24	No	92.38	57.81	12.58%	3.89%	2.44%

With feature exclusion, the results obtained were similar to those from the RF individual model (Table 4.13.), which indicates that the Stacking results still did not overcome the XGBoost's best performance.

Table 4.13. Stacking best model configurations with feature exclusions (univariate approach)

Steps ahead	Training size	Lags	Scaling	Excluded variables	RMSE	MAE	MAPE	nRMSE	nMAE
12	8736	24	No	Day of week	87.23	55.14	14.42%	3.67%	2.32%
12	8736	24	No	Seasons + day of week	87.92	54.84	13.72%	3.70%	2.31%
12	12360	24	No	Seasons + day of week	90.77	57.79	12.05%	3.82%	2.43%

The top-performing Stacking hyperparameter values, in terms of RMSE and MAE, are shown in Table 4.14.

Table 4.14. Stacking best-performing hyperparameters (univariate approach)

Model		Parameter	Value	Parameter	Value
Stacking (Steps ahead=12 Training size=8736 Lags=24 Scaling=No)	**XGBoost**	Subsample	0.7	Gamma	2.00
		Reg_lambda	1.00	Max_depth	17
		Reg_alpha	1.00	Learning_rate	0.08
		Objective	Reg:squarederror	Colsample_bytree	0.80
		N_estimators	120	Seed	42
	RF	N_estimators	100	Max_depth	12
		Min_samples_split	5	Bootstrap	True
		Min_samples_leaf	2	Accuracy_metric	‘mse’
		Max_features	0.50	Seed	42
	Linear Regression	Fit_intercept	False		

4.2.5. Univariate analysis result discussion

After analyzing all ensemble models under the different configurations following the univariate approach, it can be concluded that there does not appear to be a significant impact on using a larger number of lags (48 instead of 24). RF is the model where the effect of using the two previous days to predict the day-ahead PV power output was greater, with more than 50% of all configurations improving when using the two last days instead of one as input, whereas in the other models, this result was lower than 50%.

The scaling strategy based on the monthly installed capacity did not prove effective in enhancing any of the tested models. It aimed to account for operational variability in solar power generation, as this theoretical maximum has continuously increased during the analyzed period. However, it might not have yielded performance improvements due to the inherent robustness of ensemble models in handling feature scaling. Another limitation of this approach is the assumption that the theoretical maximum PV power output is constant for the entire month, which could have led to misrepresentation in the scaled data.

Overall, a larger training dataset improved the forecasting performance across the different models, except RF, which appeared to benefit from a smaller training dataset in most configurations. XGBoost was the model that benefited more from a larger training dataset size.

In the different models, the highest forecasting accuracies were achieved after excluding a set of features and retraining the models, which suggests the introduction of noise from some of the variables. These findings also validate the SHAP analysis, as these variables did not have a high importance and their exclusion did not jeopardize forecasting performance. From the variables excluded, it seems that the Day of the week and the Seasons of the year are the most commonly excluded across the best metric values. The Seasons variable had low importance, which was unexpected given the natural linkage between weather conditions and solar energy across the different seasons. In fact, statistically significant differences in the variance of PV power between different seasons were confirmed by a Levene's test. Nevertheless, this limited contribution to the models could be due to the fact that one and a half years' worth of data was not enough for the models to capture the underlying patterns across different seasons (only one full annual cycle is present in the dataset).

Across the different models, we can conclude that XGBoost is the best performing, with a RMSE of 82.13, an MAE of 49.84, and a MAPE of 12.10%, using 48 lags, a training data size of around a year and a half, predicting 3 hours at a time in a recursive forecast approach, and without the Day of the week variable. From the other models, RF yielded better results than LightGBM when considering feature exclusions. The Stacking model did not improve the forecasting performance of XGBoost, producing similar results to those of Random Forest. This result might be due to the amount of data available not being enough for the further complexity brought by stacking, which limits the generalization gain of this approach. The linear regression meta-learner may have also been too simple to capture nonlinear interactions between both model outputs.

Post-hoc overfitting assessment

XGBoost was considered the best univariate model from the performance metrics obtained on the testing set. A post-hoc overfitting assessment was executed by comparing training and testing errors under the same conditions using a multiple-output model with 3-step ahead predictions. This simplifies the analysis and avoids error-propagation effects that arise from recursive day-ahead forecasting.

The results show that the model maintained a consistent and stable fit to the training data, with RMSE values of around 24. Test errors were generally higher, ranging from 26 to 51, except for June 6th, suggesting it was a day with atypical weather conditions. Overall, this pattern indicates that the model has good generalization capability.

Table 4.15. Overfitting Diagnostic for XGBoost best-performing configuration: Train/Test RMSE

RMSE	01/Jun	02/Jun	03/Jun	04/Jun	05/Jun	06/Jun	Average
Train	24.05	24.07	24.02	23.79	23.76	23.91	23.93
Test	32.10	26.28	50.91	48.48	35.73	138.13	55.27

Note: The Train RMSE is measured on the training window used to predict the corresponding day (in-sample error), and the Test RMSE reports the out-of-sample error.

Figure 4.2. represents the day-ahead predictions obtained with the best-performing XGBoost model configuration (configuration present in the first row of Table 4.4.).

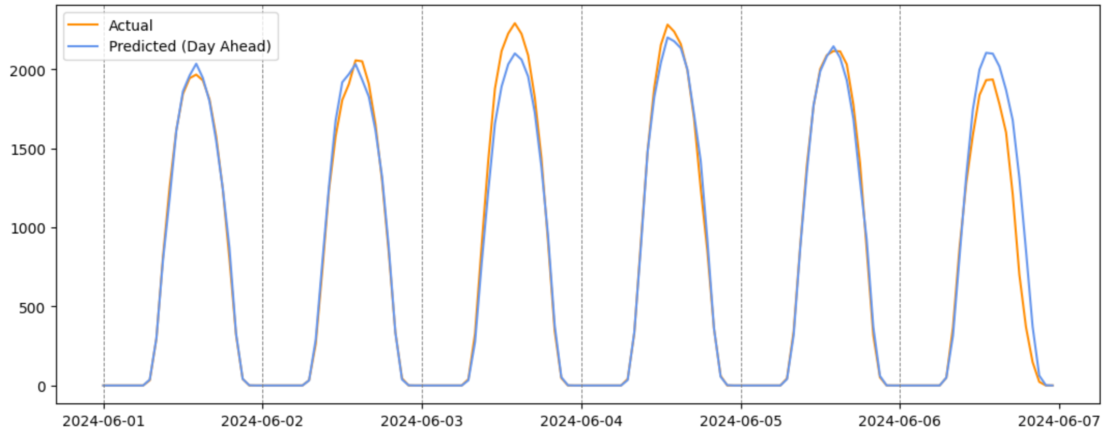


Figure 4.2. XGBoost best model configuration for day-ahead predictions in the testing set

4.3. Multivariate approach

One of the objectives of this research was to assess the inclusion of meteorological variables in PV power output forecasting at the national level and evaluate whether it would yield better results than the univariate approach.

In the multivariate approach, the number of lags used as model input was limited to 24. The inclusion of 48 lags would excessively increase the dimensionality of the feature space, and consequently, the computational burden, which could potentially outweigh the positive effects on predictive performance. Additionally, in contrast to the univariate approach, where hyperparameter tuning was conducted before feature exclusion, the multivariate approach executed the tuning stage only after exclusion of low-impact features. This strategy was also adopted to reduce the feature set dimensionality and computational complexity, by focusing the hyperparameter search on the most informative subset of variables, potentially leading to an enhanced forecasting performance.

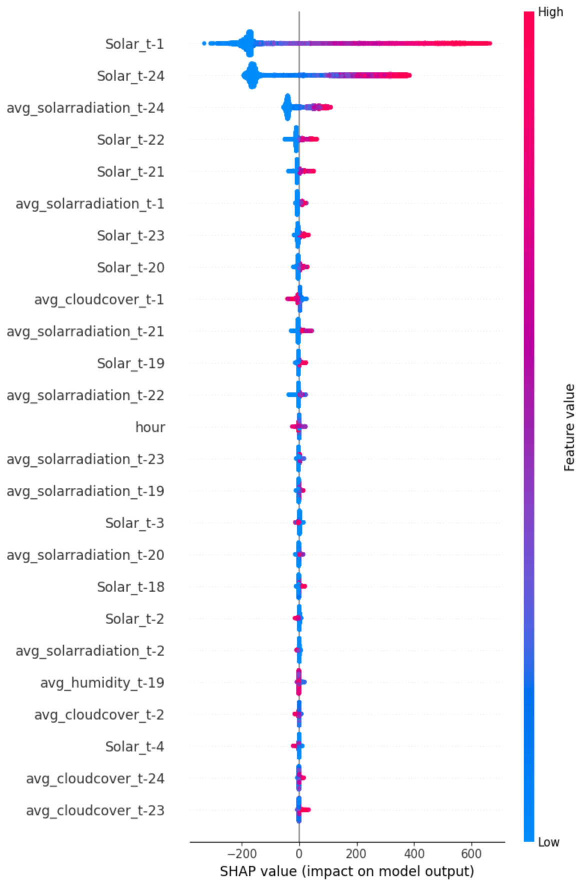
The analysis performed on feature impact was similar to the univariate approach, making use of the previously mentioned XGBoost model configuration, with the inclusion of the weather features temperature, humidity, precipitation, wind speed, wind direction, pressure, cloud cover, and solar radiation (hyperparameters used in the SHAP method are listed in Table 4.16.). In the remainder of the text, these weather features can be denoted using the prefix "avg_" to indicate that their values are from the weighted average as described in the methodology section 3.9. (e.g., avg_temperature).

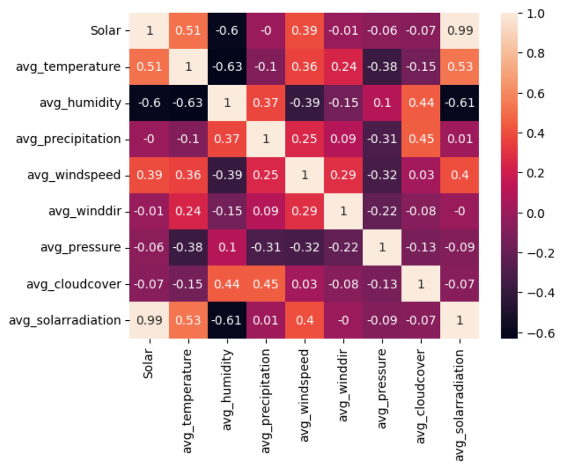
Table 4.16. XGBoost hyperparameters used for the SHAP analysis (multivariate approach)

Model	Parameter	Value	Parameter	Value
XGBoost (Steps ahead=1 Training_size=8736 Lags=24 Scaling=No)	Subsample	0.60	Gamma	1.50
	Reg_lambda	1.00	Max_depth	17
	Reg_alpha	0.70	Learning_rate	0.07
	Objective	Reg:squarederror	Colsample_bytree	1.00
	N_estimators	200	Seed	42

Regarding the time-related features, based on the results obtained from the univariate feature impact analysis, the features quarter, season, and day of week will be excluded from further model tuning and estimation. Regarding the weather variables, the choice of which to exclude for the remainder of the modeling process was made using the SHAP feature importance plot, in conjunction with the Spearman correlation matrix analysis. As a remark, please note that all lags from the same variable were kept, even if only a few would show higher importance in the SHAP plot, with the assumption that there are temporal dependencies and patterns that the model should learn (i.e., night-time zeros, morning increases), and also because they are needed for the recursive forecasting strategy.

As expected, due to the inherent relation with PV power, and the results presented in related studies, the variable Solar radiation shows a high importance in the SHAP Beeswarm plot (Figure 4.3.) and a very strong correlation in the Spearman matrix (Figure 4.4.). It also exhibits the expected effect on model output as higher values lead to positive impacts. For the reasons mentioned, Solar radiation was not excluded from the models. Cloud cover has a high impact on the SHAP analysis, but a very weak correlation with PV power. This suggests that the importance of cloud cover might emerge in combination with other features or through non-monotonic effects. Although not as pronounced as the impact of solar radiation, it can be observed in Figure 4.3. (lag avg_cloudcover t-1) that the higher the cloud cover, the smaller the impact on the prediction. The humidity variable shows high correlation with PV power, and one lag with high impact on the model (Figure 4.3.), and for these reasons, it was not discarded for the remainder of the multivariate approach.





	Solar	avg_temperature	avg_humidity	avg_precipitation	avg_windspeed	avg_winddir	avg_pressure	avg_cloudcover	avg_solarradiation
Solar	1	0.51	-0.6	-0	0.39	-0.01	-0.06	-0.07	0.99
avg_temperature	0.51	1	-0.63	-0.1	0.36	0.24	-0.38	-0.15	0.53
avg_humidity	-0.6	-0.63	1	0.37	-0.39	-0.15	0.1	0.44	-0.61
avg_precipitation	-0	-0.1	0.37	1	0.25	0.09	-0.31	0.45	0.01
avg_windspeed	0.39	0.36	-0.39	0.25	1	0.29	-0.32	0.03	0.4
avg_winddir	-0.01	0.24	-0.15	0.09	0.29	1	-0.22	-0.08	-0
avg_pressure	-0.06	-0.38	0.1	-0.31	-0.32	-0.22	1	-0.13	-0.09
avg_cloudcover	-0.07	-0.15	0.44	0.45	0.03	-0.08	-0.13	1	-0.07
avg_solarradiation	0.99	0.53	-0.61	0.01	0.4	-0	-0.09	-0.07	1

Figure 4.3. SHAP Beeswarm plot (multivariate approach)

Figure 4.4. Spearman correlation matrix

Regarding the remaining weather variables, wind direction, pressure, wind speed, temperature, and precipitation were excluded, as they exhibited very weak or moderate correlations with PV power, and none of their lags appeared in the features with the highest impact on model output.

With the selection of which time-related and weather variables to maintain in the multivariate approach in addition to the PV power lags, the following three scenarios of feature input variables for modeling were tested:

- Scenario 1: month, hour, avg_solarradiation
- Scenario 2: month, hour, avg_solarradiation, avg_cloudcover
- Scenario 3: month, hour, avg_solarradiation, avg_cloudcover, and avg_humidity

4.3.1. XGBoost

The multivariate approach in the XGBoost model (Appendix D, Tables D.1. to D.4.) revealed similar conclusions to the univariate approach. This method did not show evidence that the tested scaling technique improved forecasting performance. It improved around 48% and 44%

of the 48 compared configurations, when considering the RMSE and MAE metrics, respectively (Table D.3.). Moreover, a larger training dataset improved the great majority of the configurations (100% and 96%, taking into account RMSE and MAE, respectively) (Table D.4.). Table 4.17. shows the best model configurations for XGBoost when incorporating weather data (hyperparameters used for the best RMSE result are present in Table 4.18.). It allowed us to conclude that none of these best-performing configurations surpassed the univariate approach results.

Table 4.17. XGBoost best model configurations (multivariate approach)

Steps ahead	Training size	Scaling	Features included	RMSE	MAE	MAPE	nRMSE	nMAE
6	12360	No	Scenario 3	88.08	52.06	15.25%	3.71%	2.19%
3	12360	No	Scenario 1	91.73	54.63	11.57%	3.86%	2.30%

Table 4.18. XGBoost best-performing hyperparameters (multivariate approach)

Model	Parameter	Value	Parameter	Value
XGBoost (Steps ahead=6 Training size=12360 Lags=24 Scaling=No Features=Scenario 3)	Subsample	0.50	Gamma	1.00
	Reg_lambda	1.50	Max_depth	9
	Reg_alpha	0.70	Learning_rate	0.07
	Objective	Reg:squarederror	Colsample_bytree	0.50
	N_estimators	180	Seed	42

4.3.2. Random Forest

In RF model configurations (Appendix D, Tables D.5. to D.8.), both the use of the scaling technique and the use of a larger training dataset size showed slight evidence that they would improve model performance, as they improved around half of the compared configurations, considering RMSE and MAE (Tables D.7. and D.8.). As a result, the best-performing RF results were obtained with scaling and around one year and a half of training data (Table 4.19.). Nevertheless, these results did not surpass the previous univariate results. (Both multivariate best-performing configurations converged to the hyperparameter values present in Table 4.20.).

Table 4.19. RF best model configurations (multivariate approach)

Steps ahead	Training size	Scaling	Features included	RMSE	MAE	MAPE	nRMSE	nMAE
24	12360	Yes	Scenario 2	91.61	54.28	13.11%	3.86%	2.29%
24	12360	Yes	Scenario 3	91.83	53.95	12.02%	3.87%	2.27%

Table 4.20. RF best-performing hyperparameters (multivariate approach)

Model	Parameter	Value	Parameter	Value
RF (Steps ahead=24 Training size=12360 Lags=24 Scaling=Yes)	N_estimators	150	Max_depth	10
	Min_samples_split	2	Bootstrap	True
	Min_samples_leaf	1	Accuracy_metric	'mse'
	Max_features	0.50	Seed	42

4.3.3. LightGBM

The conclusions obtained regarding LightGBM (Appendix D, Tables D.9. to D.12.) are similar to those from XGBoost in the sense that the scaling technique only enhanced around 50% of the 48 compared configurations (Table D.11.). Also, considering RMSE, around 81% of the compared configurations improved with a larger training dataset size (Table D.12.), having on average a reduction of 7.16 points in RMSE, 3.95 points in MAE, and limited influence in MAPE (-0.73 percentual points), when considering a smaller training dataset. In this case, the best-performing configuration under the three error metrics (Table 4.21.) surpassed the results obtained in the univariate approach (hyperparameter values present in Table 4.22).

Table 4.21. LightGBM best model configurations (multivariate approach)

Steps ahead	Training size	Scaling	Features included	RMSE	MAE	MAPE	nRMSE	nMAE
24	12360	Yes	Scenario 3	86.74	52.54	13.24%	3.65%	2.21%

Table 4.22. LightGBM best-performing hyperparameters (multivariate approach)

Model	Parameter	Value	Parameter	Value
LightGBM (Steps ahead=24 Training size=12360 Scaling=Yes Features=Scenario 3)	Num_leaves	70	Learning_rate	0.06
	N_estimators	120	Feature_fraction	1.0
	Max_depth	22	Boosting_type	'gbdt'

4.3.4. Stacking

Lastly, in the Stacking model (Appendix D, Tables D.13. to D.16.) the benefits of the multivariate approach are not obvious, as the best-performing configuration (Table 4.23.) had a better MAPE than those from the univariate approach; however, both RMSE and MAE showed worse results (Hyperparameter values are present in Table 4.24.).

Table 4.23. Stacking best model configurations (multivariate approach)

Steps ahead	Training size	Scaling	Features included	RMSE	MAE	MAPE	nRMSE	nMAE
24	12360	Yes	Scenario 3	96.07	56.56	12.01%	4.05%	2.38%
4	12360	No	Scenario 2	92.83	57.45	18.06%	3.91%	2.42%

Table 4.24. Stacking best-performing hyperparameters (multivariate approach)

Model		Parameter	Value	Parameter	Value
Stacking (Steps ahead=24 Training size=12360 Scaling=Yes Features=Scenario 3)	**XGBoost**	Subsample	0.6	Gamma	1.5
		Reg_lambda	1.0	Max_depth	9
		Reg_alpha	1.0	Learning_rate	0.08
		Objective	Reg:squarederror	Colsample_bytree	0.5
		N_estimators	200	Seed	42
	RF	N_estimators	150	Max_depth	11
		Min_samples_split	5	Bootstrap	False
		Min_samples_leaf	2	Accuracy_metric	'mse'
		Max_features	0.2	Seed	42
	Linear Regression	Fit_intercept	False		

The scaling technique only improved around 40% of the configurations (considering RMSE and MAE), whereas larger training set sizes improved more than 60% of them (Figures D.15. and D.16., respectively).

4.3.5. Multivariate approach discussion

Following the multivariate approach with the use of a weighted average of each weather variable, it was possible to conclude that it did not produce the desired outcomes, as the obtained performance metric values were worse than those from the univariate approach. Only the LightGBM, which did not achieve particularly satisfactory results in the univariate, has now shown a slightly better performance in terms of RMSE and MAE on the best configuration.

In terms of the overall effects of the different setups on model performance, the results were similar to those from the univariate approach, as the tested scaling strategy did not exhibit evidence that it would enhance the ensemble models. Regarding the size of the training dataset, the positive effect of using 12360 hours instead of 8736 hours is noticeable in XGBoost, LightGBM, and Stacking, but there is no appreciable effect on the RF.

From the SHAP analysis, it was possible to conclude that, at first instance, the weather variables that showed a higher impact on model prediction were solar radiation, cloud cover, and humidity. The models were then tested under three different input variable scenarios. However, based on the best-performing results of each model, it is not possible to conclude which scenario is more recommendable than the others.

4.4. Business interpretation of the best-performing model results

Reliable day-ahead PV power forecasting supports the power grid's good functioning, contributes to effective operational planning, and strengthens the country's position and contribution in energy markets.

Considering the proposed XGBoost model, we obtained an RMSE of 82.13 MW, meaning that on average, the forecasted solar power is around 82 MW away from the actual daily PV power output. This value represents 3.46% of the peak PV power produced during the analyzed period. The forecast deviation represents system imbalances that will require the system operators to apply balancing measures, which will impose extra costs on the PV production. Nevertheless, because it outperforms all alternatives, it is the most advantageous choice for operational use.

The superior performance of XGBoost lowered the forecasting error by 20.9% when compared to the SARIMA baseline model[4]. Moreover, two naïve forecast models called Persistence and the Hourly average of training data were also computed to frame the business value of the proposed model (RMSE per Testing Day and Average of the testing days are present in Figure 4.5.). In the Persistence model, the previous day is used as forecast for the day-ahead horizon. Overall, XGBoost outperforms the benchmarks, obtaining, on average, less 21.2% of RMSE than Persistence. The Hourly Average Naïve forecast had very poor performance, showing not to be suitable as a reference model.

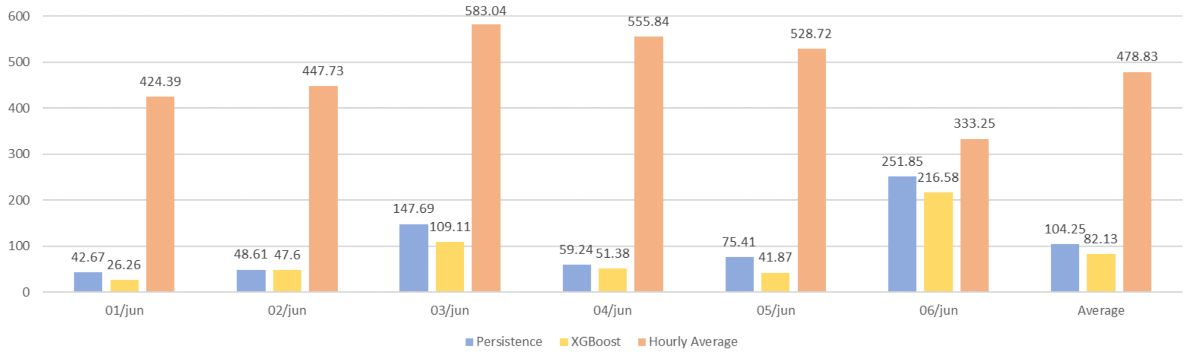


Figure 4.5. Comparison of day-ahead RMSE on test subset for proposed XGBoost and Naïve forecasts

Ensemble modeling approaches are expected to increase computational complexity as a result of training and combining multiple models. This could be a critical constraint for real-

[4] The Prophet model is extensively used by organizations due to its low parametrization requirements, easy implementation, and interpretability. Nevertheless, in this research, the Prophet showed lower performance than the other baseline models (SARIMA and UCM). For this reason, it was not included in the further business interpretation and framing of the best model results.

time implementation of ensemble solutions. However, following the defined methodology, the recommended XGBoost model demonstrated low computational time, thus being suitable for real-time day-ahead Portugal’s PV power forecasting. On average, the model took around 8 seconds to train on around a year and a half of historical data, and 0.3 seconds to predict the next 24 hours.

The results indicate that deploying the XGBoost ensemble model for Portugal’s day-ahead context would deliver tangible accuracy gains, potentially leading to cost reductions and improved grid management. In addition, the execution times the model requires make it compatible with daily operations. Lastly, even though XGBoost is more difficult to interpret than other models, techniques like SHAP support operational trust and governance.

CHAPTER 5

Conclusions and Future Work

The present research focused on the performance comparison of different ensemble models for Portugal’s national-scale PV power forecasting. It included the evaluation of XGBoost, Random Forest, LightGBM, and a Stacking model. Those models were evaluated under various setups, such as the number of previous lags used as input, the availability of training data, and the implementation of a scaling technique, which allowed the gathering of insights into the most effective configurations for completing the purpose of national-level solar power prediction.

The study's results suggest that the best-performing ensemble model was XGBoost, reaching the following error metric results in the best-performing setup without the usage of the adjusted weather data: RMSE of 82.13, an MAE of 49.84, and a MAPE of 12.10%. Based on the performance analysis of the different model setups, it is recommended to use XGBoost with approximately one and a half years’ worth of training data, as the model benefited more from a larger training dataset compared to using one year. From the remaining ensemble models, RF yielded slightly better results than LightGBM when considering feature exclusions. Moreover, the estimated Stacking model, with base learners XGBoost and RF, could not enhance the results of the single models. It did not improve the forecasting performance of XGBoost and produced similar results to RF.

A multivariate approach, incorporating weighted averages of the individual weather variables, was tested to investigate potential forecasting enhancements. However, it did not produce the desired outcomes, as the obtained performance metrics were worse than those from the approach that did not use weather data, leaving open the challenge of implementing an effective way of incorporating weather data into national-scale PV power forecasting.

In the different models, following the univariate approach, the highest forecasting accuracies were achieved after excluding a set of features and retraining the models, which suggests the introduction of noise from some of the variables. Nonetheless, XGBoost also produced satisfactory results without feature exclusion. After the exclusion of the least impactful variables, the models were retrained, but no further hyperparameter tuning was performed to isolate the effects of the exclusion of those features. Still, it is recommended to re-execute the hyperparameter tuning stage after excluding those variables to seek optimized parameters for the reduced input feature set.

The method of using a fixed time window to set the nighttime predictions to zero simplifies its deployment, but it does not account for the different seasonal changes in sunrise/sunset times, which could lead to overstating certain periods, as for example, the winter early night hours. It is recommended to adapt this strategy to ensure the time window that is considered to be nighttime is according to the data from each season.

As mentioned by Yang et al. (2024), the prevalence of black-box AI models is a concern for users as they do not provide information on how the variables influence the predictions. To overcome this issue, the present research implemented an explainable AI analysis and incorporated its results for variable selection. It was found that in both the univariate and multivariate approaches, the most impactful variables in predicting output were the previous hour and the same hour of the previous day. Future research could evaluate the models using only these two lags to assess whether a simpler model would yield similar performance. In the multivariate approach, Solar radiation, more specifically, the same hour from the previous day, was also revealed to have a meaningful influence on the model predictions. The Feature Selection based on the SHAP value results had the limitation of being executed having in consideration a single model's optimized configuration, leading to conclusions that may not apply to all the tested models.

Another limitation of the current research is the fact that overfitting was assessed only after model selection and focused on a single set of XGBoost hyperparameters, using train-test error comparisons. While the model exhibited a small generalization gap, this ex-post procedure may understate variance. Future work should integrate overfitting studies earlier in the model development, for example, by applying early stopping and building learning curves.

All the ensemble models proposed yielded better results than the baseline models and were proven effective for PV power forecasting at a national scale. Nevertheless, they weren't tested against machine learning models like SVM, ANN, or RNN, which are frequently applied in PV power forecasting, as shown in literature reviews (Başaran et al., 2020; Krechowicz et al., 2022). As future work, these models should also be incorporated in the model comparisons for a more concrete framing of the ensemble model results. If proven effective, those models could also be experimented with in Stacking model approaches.

The model results presented in this study are specific to Portugal's PV power production, which is closely related to its geographic and grid characteristics, making it difficult to generalize the conclusions obtained. As also observed in the related works, there is a lack of studies that test the generalization capabilities of the proposed models using data from different locations. As this was not executed in this research, it is proposed as future work to test this methodology using data from different countries, which would likely further consolidate the validity of the obtained results.

References

Abdellatif, A., Mubarak, H., Ahmad, S., Ahmed, T., Shafiullah, G. M., Hammoudeh, A., Abdellatef, H., Rahman, M. & Gheni, H. M. (2022a). Forecasting photovoltaic power generation with a stacking ensemble model. Sustainability, 14(17), 11083. https://doi.org/10.3390/su141711083

Abdellatif, A., Mubarak, H., Ahmad, S., Hammoudeh, A., Mekhilef, S., & Mokhlis, H. (2022b). Forecast of Solar Photovoltaic Power Output Based on Polycrystalline Panel-based Employing Various Ensemble Machine Learning Methods. Proceedings of the 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT) (pp. 1-6). IEEE. https://doi.org/10.1109/GlobConPT57482.2022.9938348

Abdelmoula, I. A., Elhamaoui, S., Elalani, O., Ghennioui, A., & Aroussi, M. E. (2022). A photovoltaic power prediction approach enhanced by feature engineering and stacked machine learning model. Energy Reports, 8, 1288–1300. https://doi.org/10.1016/j.egyr.2022.07.082

Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. D. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renewable & Sustainable Energy Reviews, 124, 109792. https://doi.org/10.1016/j.rser.2020.109792

Atwan, T. A. (2022). Time Series Analysis with Python Cookbook: Practical Recipes for Exploratory Data Analysis, Data Preparation, Forecasting, and Model Evaluation. Packt Publishing.

Baloch, M., Honnurvali, M. S., Kabbani, A., Ahmed, T., Chauhdary, S. T., & Saeed, M. S. (2025). Solar Energy Forecasting Framework using Prophet Based Machine Learning Model: An opportunity to explore solar energy potential in Muscat Oman. Energies, 18(1), 205. https://doi.org/10.3390/en18010205

Banik, R., & Biswas, A. (2023). Improving Solar PV Prediction Performance with RF-CatBoost Ensemble: A Robust and Complementary Approach. Renewable Energy Focus, 46, 207–221. https://doi.org/10.1016/j.ref.2023.06.009

Başaran, K., Bozyiğit, F., Siano, P., Taşer, P. Y., & Kılınç, D. (2020). Systematic literature review of photovoltaic output power forecasting. Iet Renewable Power Generation, 14(19), 3961–3973. https://doi.org/10.1049/iet-rpg.2020.0351

Boutahir, M. K., Farhaoui, Y., Azrour, M., Sedik, A., & Nasralla, M. M. (2024). Advancing solar power forecasting: integrating boosting cascade forest and Multi-Class-Grained scanning for enhanced precision. Sustainability, 16(17), 7462.

Chakraborty, D., Mondal, J., Barua, H. B., & Bhattacharjee, A. (2023). Computational solar energy – Ensemble learning methods for prediction of solar power generation based on meteorological parameters in Eastern India. Renewable Energy Focus, 44, 277–294. https://doi.org/10.1016/j.ref.2023.01.006

Chen, C., Chai, L., & Wang, Q. (2024). Research on stacking ensemble method for day-ahead ultra-short-term prediction of photovoltaic power. Renewable Energy, 238, 121853. https://doi.org/10.1016/j.renene.2024.121853

Chen, X., Xie, B., Zhang, P., & Qiu, X. (2023). Research on Wind and Solar Power Generation Forecasting Based on SARIMA-LSTM Model. Proceedings of the 2023 3rd International Conference on New Energy and Power Engineering (ICNEPE) (pp. 695-699). IEEE. https://doi.org/10.1109/ICNEPE60694.2023.10429158

Cicek, Z. (2024). A stacking ensemble model for forecasting solar photovoltaic power generation. Proceedings of the 2024 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 01-05). IEEE. https://doi.org/10.1109/ASYU62119.2024.10757016

Córtez, J., Cumbicos, J., Terada, L., Lopez, J., Giesbrecht, M., Fraidenraich, G., & Rider, M. (2024). Fuzzy Ensemble Algorithm for Day-ahead Photovoltaic Power Forecasting. Proceedings of the 2024 International Conference on Smart Energy Systems and Technologies (SEST) (pp. 01-06). IEEE. https://doi.org/10.1109/SEST61601.2024.10694514

Dai, Y., Yu, W., & Leng, M. (2024). A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting. Energy, 299, 131458. https://doi.org/10.1016/j.energy.2024.131458

Direção Geral de Energia e Geologia. (2023-2024). Estatísticas rápidas das renováveis. https://www.dgeg.gov.pt/pt/estatistica/energia/publicacoes/estatisticas-rapidas-das-renovaveis/

Gaboitaolelwe, J., Zungeru, A. M., Yahya, A., Lebekwe, C. K., Vinod, D. N., & Salau, A. O. (2023). Machine learning based solar photovoltaic power forecasting: A review and comparison. IEEE Access: Practical Innovations, Open Solutions, 11, 40820–40845. https://doi.org/10.1109/access.2023.3270041

Guo, X., Wang, X., Ao, Y., Dai, W., & Gao, Y. (2022). Short-term photovoltaic power forecasting with adaptive stochastic configuration network ensemble. WIREs Data Mining and Knowledge Discovery, 12(6). https://doi.org/10.1002/widm.1477

Gupta, R., Yadav, A., Jha, S., Pathak, P. (2022) Time series forecasting of solar power generation using Facebook Prophet and XG Boost. Proceedings of the 2022 IEEE Delhi section conference (DELCON) (pp 1–5). IEEE. https://doi.org/10.1109/DELCON54057.2022.9752916

Hamilton, J. D., (1994). Time Series Analysis. Princeton University Press.

Harvey, A. & Koopman, S. J. (2009). Unobserved components models in economics and finance. IEEE Control Systems Magazine (pp. 71-81). IEEE. https://doi.org/10.1109/MCS.2009.934465

Herrera-Casanova, R., Conde, A., & Santos-Pérez, C. (2024). Hour-Ahead Photovoltaic Power Prediction Combining BiLSTM and Bayesian Optimization Algorithm, with Bootstrap Resampling for Interval Predictions. Sensors, 24(3), 882. https://doi.org/10.3390/s24030882

Hokmabad. H., Belikov, J., & Petlenkov, E. (2023). Day-ahead PV Output Power Forecasting Utilizing Boosting Recursive LightGBM-LSTM Framework. Proceedings of the 2023 IEEE PES Innovative

Smart Grid Technologies Europe (ISGT EUROPE) (pp. 01-05). IEEE. https://doi.org/10.1109/ISGTEUROPE56780.2023.10408090

Huang., Y., Wang, A., Jiao, J., Xie, J., & Chen, H. (2023). Short-Term PV Power Forecasting Based on CEEMDAN and Ensemble DeepTCN. IEEE Transactions on Instrumentation and Measurement, 72, 1–12. https://doi.org/10.1109/tim.2023.3310090

Jamieson, K., & Talwalkar, A. (2015). Non-stochastic best arm identification and hyperparameter optimization. https://doi.org/10.48550/arXiv.1502.07943

Jobayer, M., Shaikat, M. a. H., Rashid, M. N., & Hasan, M. R. (2023). A systematic review on predicting PV system parameters using machine learning. Heliyon, 9(6), e16815. https://doi.org/10.1016/j.heliyon.2023.e16815

Kane, S., Rathore, A., Abhijeet, Anees, S., & Kane, A. (2024). Comparative Analysis of Machine Learning Models for Solar Power Prediction Using Time Series Weather Data. Proceedings of the 2024 IEEE Third International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES) (pp. 779-784). IEEE. https://doi.org/10.1109/ICPEICES62430.2024.10719133

Karazor, A., & Kavaz, A. (2023). Machine Learning Ensembles for Solar Power Forecasting. Proceedings of the 2023 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES) (pp. 01-05). IEEE. https://doi.org/10.1109/CIEES58940.2023.10378818

Karthikeyan, G., & Jagadeeshwaran, A. (2023). Enhancing solar Energy Generation: A comprehensive Machine Learning-Based PV prediction and fault analysis system for Real-Time tracking and Forecasting. Electric Power Components and Systems, 52(9), 1497–1512. https://doi.org/10.1080/15325008.2023.2293947

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: a highly efficient gradient boosting decision tree. Proceedings of the 2017 31st International Conference on Neural Information Processing Systems (NIPS17) (pp. 3149–3157).

Khan, W. A., Walker, S., & Zeiler, W. W. (2022). Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. Energy, 240, 122812. https://doi.org/10.1016/j.energy.2021.122812

Kim, B., Suh, D., Otto, M., & Huh, J. (2021). A novel hybrid Spatio-Temporal forecasting of multisite solar photovoltaic generation. Remote Sensing, 13(13), 2605. https://doi.org/10.3390/rs13132605

Kyeremeh, F., Zhi, F., Yi, Y., Gyamfi, E., & Nti, I. K. (2022). Solar PV Power Forecasting with a Hybrid LSTM-AdaBoost Ensemble. Proceedings of the 2022 IEEE/IET International Utility Conference and Exposition (IUCE) (pp. 1-7). IEEE. https://doi.org/10.1109/IUCE55902.2022.10079424

Krechowicz, A., Krechowicz, M., & Poczęta, K. (2022). Machine Learning Approaches to Predict Electricity Production from Renewable Energy Sources. Energies, 15(23), 9146. https://doi.org/10.3390/en15239146

Komar, M., Lipianina-Honcharenko, K., Domanskyi, V., & Melnyk, N. (2024). Regression-based method for real-time solar power plant efficiency forecasting. Proceedings of the Modern Machine Learning Technologies Workshop (MoMLeT 2024) (pp. 235-245). (CEUR-WS)

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54(1–3), 159–178. https://doi.org/10.1016/0304-4076(92)90104-y

Lateko, A. A. H., Yang, H., Huang, C., Aprillia, H., Hsu, C., Zhong, J., & Phương, N. H. (2021). Stacking Ensemble Method with the RNN Meta-Learner for Short-Term PV Power Forecasting. Energies, 14(16), 4733. https://doi.org/10.3390/en14164733

Lateko, A. A. H.; Yang, H., & Huang, C. (2022). Short-Term PV power forecasting using a Regression-Based ensemble method. Energies, 15(11), 4171. https://doi.org/10.3390/en15114171

Liu, L., Sun, Q., Wennersten, R., & Chen, Z. (2023a). Day-ahead Forecast of Photovoltaic Power Based on A Novel Stacking Ensemble Method. IEEE Access, 11, 113593–113604. https://doi.org/10.1109/access.2023.3323526

Liu, L., Guo, K., Chen, J., Lin, G., Ke, C., Liang, J., & He, D. (2023b). A photovoltaic power prediction approach based on data decomposition and stacked deep learning model. Electronics, 12(13), 2764. https://doi.org/10.3390/electronics12132764

Lin, H. H. (2023). Applying Evolutionary Algorithms to Optimize Hyperparameters for Prediction Model of Solar Power Generation. Proceedings of the 2023 Sixth International Symposium on Computer, Consumer and Control (IS3C) (pp. 64-67). IEEE. https://doi.org/10.1109/IS3C57901.2023.00025

Loh, C., Chen, Y., Su, C., & Su, H. (2024). Establishing lightweight and robust prediction models for solar power forecasting using Numerical–Categorical Radial Basis Function deep neural networks. Applied Sciences, 14(22), 10625. https://doi.org/10.3390/app142210625

Markovics, D., & Mayer, M. J. (2022). Investigating the effect of training time for machine learning based photovoltaic power forecasting. Proceedings of the 2022 8th International Youth Conference on Energy (IYCE) (pp. 1-7). IEEE. https://doi.org/10.1109/IYCE54153.2022.9857544

Massaoudi, M., Abu-Rub, H., Refaat, S. S., Trabelsi, M., Chihi, I., & Oueslati, F. S. (2021). Enhanced Deep Belief Network Based on Ensemble Learning and Tree-Structured of Parzen Estimators: An Optimal Photovoltaic Power Forecasting Method. IEEE Access, 9, 150330–150344. https://doi.org/10.1109/access.2021.3125895

Mohana, M., Saidi, A. S., Alelyani, S., Alshayeb, M., Basha, S., & Anqi, A. E. (2021). Small-Scale solar photovoltaic power prediction for residential load in Saudi Arabia using machine learning. Energies, 14(20), 6759. https://doi.org/10.3390/en14206759

Mondal, R., Roy, S. K., & Giri, C. (2024). Solar power forecasting using domain knowledge. Energy, 302, 131709. https://doi.org/10.1016/j.energy.2024.131709

Natarajan, K. P., & Singh, J. G. (2023). Solar Power Forecasting using Stacking Ensemble Models with Bayesian Meta-Learning. Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE. https://doi.org/10.1109/ICCCNT56998.2023.10308202

Nespoli, A., Leva, S., Mussetta, M., & Ogliari, E. G. C. (2022). A Selective Ensemble Approach for Accuracy Improvement and Computational Load Reduction in ANN-Based PV Power Forecasting. 10, 32900–32911. https://doi.org/10.1109/access.2022.3158364

Observatório da Energia, Direção-Geral de Energia e Geologia (DGEG), & Agência para a Energia (ADENE). (2024). Energia em números – Edição 2024. https://www.dgeg.gov.pt/media/e1eb3n0l/dgeg-aen-2024e.pdf

Peng, Y., Wang, S., Chen, W., Ma, J., Wang, C., & Chen, J. (2023). LightGBM-Integrated PV Power Prediction based on Multi-Resolution Similarity. Processes, 11(4), 1141. https://doi.org/10.3390/pr11041141

Petrosian, O., & Zhang, Y. (2024). Solar power generation forecasting in smart cities and explanation based on explainable AI. Smart Cities, 7(6), 3388–3411. https://doi.org/10.3390/smartcities7060132

Rai, A., Shrivastava, A., & Jana, K. C. (2022). An Empirical analysis of machine learning algorithms for solar power forecasting in a high dimensional uncertain environment. IETE Technical Review, 40(4), 558–573. https://doi.org/10.1080/02564602.2022.2136270

Rahimi, N., Park, S., Choi, W., Oh, B., Kim, S., Cho, Y., Ahn, S., Chong, C., Kim, D., Jin, C., & Lee, D. (2023). A comprehensive review on ensemble solar power forecasting Algorithms. Journal of Electrical Engineering & Technology, 18(2), 719–733. https://doi.org/10.1007/s42835-023-01378-2

Rusina, A. G., Osgonbaatar, T., Stepanova, A. I., & Matrenin, P. V. (2023). Ensemble Machine Learning Model for Day Ahead Solar Power Forecasting for Mongolia Power System. Proceedings of the 2023 Belarusian-Ural-Siberian Smart Energy Conference (BUSSEC) (pp. 84-87). IEEE. https://doi.org/10.1109/BUSSEC59406.2023.10296344

Sedai, A. K., Dhakal, R., Gautam, S., Dhamala, A., Bilbao, A., Wang, Q., Wigington, A., & Pol, S. (2023). Performance analysis of statistical, machine learning and deep learning models in Long-Term forecasting of solar power production. Forecasting, 5(1), 256–284. https://doi.org/10.3390/forecast5010014

Sharma, N., Mangla, M., Yadav, S., Goyal, N., Singh, A., Verma, S., & Saber, T. (2021). A sequential ensemble model for photovoltaic power forecasting. Computers & Electrical Engineering, 96, 107484. https://doi.org/10.1016/j.compeleceng.2021.107484

Siriwardana, S., Nishshanka, T., Peiris, A., Boralessa, M. A. K. S., Hemapala, K. T. M. U., & Saravanan, V. (2022). Solar Photovoltaic Energy Forecasting Using Improved Ensemble Method For Micro-grid Energy Management. Proceedings of the 2022 IEEE 2nd International Symposium on

Sustainable Energy, Signal Processing and Cyber Security (iSSSC) (pp. 1-6). IEEE. https://doi.org/10.1109/iSSSC56467.2022.10051275

Skforecast. (n.d.). Recursive multi-step forecasting. https://skforecast.org/0.9.1/user_guides/autoregresive-forecaster

Song, Y., Zhao, T., Niu, Z., Du, J., Jiang, F., & Zhai, F. (2022). Short-Term PV Output Prediction Method Based on Feature Analysis and Multi-model Fusion. Proceedings of the 2022 3rd Information Communication Technologies Conference (ICTC) (pp. 232-237). IEEE. https://doi.org/10.1109/ICTC55111.2022.9778479

Taheri, N., Piazzi, A., & Tucci, M. (2024). Analyzing the Impact of Training Data Availability on Machine Learning Models Accuracy for Regional Photovoltaic Production Forecast. Proceedings of the 2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI) (pp. 636-641). IEEE. https://doi.org/10.1109/RTSI61910.2024.10761794

Thai, H. (2022). Machine learning for structural engineering: A state-of-the-art review. Structures, 38, 448–491. https://doi.org/10.1016/j.istruc.2022.02.003

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Tseng, Y., Chiu, C., Huang, W., & Chang, G. (2024). A Hybrid Clustering and Neural Network-Based Ensemble Method for Day-Ahead PV Output Forecast. Proceedings of the 2024 IEEE Texas Power and Energy Conference (TPEC) (pp. 01-06). IEEE. https://doi.org/10.1109/TPEC60005.2024.10472241

Tung, C. (2024). IoT data collection and Short-Term Solar Power Forecasting using Stacked Generalization Ensemble model. Sensors and Materials, 36(1), 67. https://doi.org/10.18494/sam4530

Visser, L., AlSkaif, T., & Van Sark, W. (2022). Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. Renewable Energy, 183, 267–282. https://doi.org/10.1016/j.renene.2021.10.102

Wang, Q., Chai, S., Liu, Y., & Wang, G. (2022b). GTFD-XTNet: A tabular learning-based ensemble approach for short-term prediction of photovoltaic power. Iet Renewable Power Generation, 16(12), 2682–2693. https://doi.org/10.1049/rpg2.12494

Wang, H., Yan, S., Ju, D., Ma, N., Fang, J., Wang, S., Li, H., Zhang, T., Xie, Y., & Wang, J. (2023). Short-Term photovoltaic power forecasting based on a feature Rise-Dimensional Two-Layer Ensemble learning model. Sustainability, 15(21), 15594. https://doi.org/10.3390/su152115594

Wang, M., & Wang, P. (2021). Ensemble Multilayer Perceptron Model for Day-ahead Photovoltaic Forecasting. Proceedings of the 4th International Conference on Control and Computer Vision (pp. 189-194). https://doi.org/10.1145/3484274.3484304

Wang, M., Wang, P., & Zhang, T. (2022a). Evidential Extreme Learning Machine Algorithm-Based Day-Ahead Photovoltaic Power Forecasting. Energies, 15(11), 3882. https://doi.org/10.3390/en15113882

Wu, J., Tang, Z., Gao, Y., Wei, L., Zhou, J., Han, F., & Liu, J. (2022). Regional Distributed Photovoltaic Short Term Power Prediction Method Based on Cluster Analysis and Stacking Ensemble Learning. Proceedings of the 2022 Asian Conference on Frontiers of Power and Energy (ACFPE) (pp. 361-365). IEEE. https://doi.org/10.1109/ACFPE56003.2022.9952286

Xia, L., Long, H., & Wang, J. (2023). Photovoltaic Power Prediction Based on Cluster Analysis and Fitting Residual Random Forest Ensemble Model. Proceedings of the 2023 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 850-855). IEEE. https://doi.org/10.1109/AEECA59734.2023.00156

Xu, W., Wang, Z., Wang, W., Zhao, J., Wang, M., & Wang, Q. (2024). Short-Term Photovoltaic Output Prediction Based on Decomposition and Reconstruction and XGBoost under Two Base Learners. Energies, 17(4), 906. https://doi.org/10.3390/en17040906

Yang, X., Ye, S., Jiang, K., Sun, C., Lu, Z., Liu, L., Han, Y., & Zhang, B. (2024). Medium- and Long-Term distributed photovoltaic power prediction based on multiple Time Series feature and Multiple-Model Fusion. Sustainability, 16(20), 8806. https://doi.org/10.3390/su16208806

Zhang, H., & Zhu, T. (2022). Stacking model for Photovoltaic-Power-Generation prediction. Sustainability, 14(9), 5669. https://doi.org/10.3390/su14095669

Zhang, Y., Li, B., & Han, Z. (2024). A new method for short-term photovoltaic power generation forecast based on ensemble model. AIP Advances, 14(9). https://doi.org/10.1063/5.0226761

Zhang, Y., & Li, X. (2023). A PV Power Prediction Method Based On Optimized VMD and Combination Prediction Model Selection Strategy. Proceedings of the 2023 8th International Conference on Power and Renewable Energy (ICPRE) (pp. 1788-1793). IEEE. https://doi.org/10.1109/ICPRE59655.2023.10353756

Zheng, W., Xiao, H., Zhou, H., Wang, N., Guan, Y., & Teng, W. (2024). Distributed Photovoltaic Power Prediction With Limited Data: A Combined Ensemble Learning and Transfer Learning Approach. Proceedings of the 2024 9th Asia Conference on Power and Electrical Engineering (ACPEE) (pp. 1373-1378). IEEE. https://doi.org/10.1109/ACPEE60788.2024.10532490

Zhou, B., Chen, X., Li, G., Gu, P., Huang, J., & Yang, B. (2023a). XGBoost–SFS and Double Nested Stacking Ensemble Model for Photovoltaic Power Forecasting under Variable Weather Conditions. Sustainability, 15(17), 13146. https://doi.org/10.3390/su151713146

Zhou, Q., Tang, X., Lv, Q., Li, Z., Shen, J., Wang, J., & Yong, B. (2023b). Ensemble Machine Learning Method for Photovoltaic Power Forecasting. Proceedings of the 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 332-337). IEEE. https://doi.org/10.1109/CSCWD57460.2023.10152827

Zhou, Y., Xue, W., Liu, J., & Li, K. (2022). Photovoltaic power prediction based on SVMD-PCA-EL model. Proceedings of the 2022 41st Chinese Control Conference (CCC) (pp. 5621-5626). IEEE. https://doi.org/10.23919/CCC55666.2022.9902119

Appendix

Appendix A. Literature Review

Table A.1. Information from the used datasets in the reviewed articles

Authors (Year)	Data resolution	Data period (days)	Forecasting horizon	Location / Country
Abdellatif et al. (2022a)	daily	1461	Short-term (day-ahead)	Malaysia
Abdellatif et al. (2022b)	5min	365	Short-term (hour-ahead)	Malaysia
Abdelmoula et al. (2022)	2min	182	Short-term (hour-ahead)	Benguerir, Morocco
Banik & Biswas (2023)	1h	2557	Long-term (month-ahead)	Tripura, India
Boutahir et al. (2024)	15min	unknown	unknwon	Nashik, India Gandikota, India
Chakraborty et al. (2023)	1min	213	Unknown	West Bengal, India
Chen et al. (2024)	15min	349	Very Short-term (15min)	China
Cicek (2024)	15min	89	Unknown	Istanbul, Turkey
Córtez et al. (2024)	15min	1219	Short-term (day-ahead)	Campinas, Brazil
Dai et al. (2024)	undetermined	2557 2192	Short-term (unknown)	Shanghai, China Alice Springs, Australia
Guo et al. (2022)	7.5min	91	Short-term (unknown)	unknown
Herrera-Casanova et al. (2024)	30min	730	Short-term (hour-ahead)	Alcala, Spain
Hokmabad et al. (2023)	1h	730	Short-term (day-ahead)	Tallin, Estonia
Huang et al. (2023)	15min	366	Short-term (unknown)	Alice Springs, Australia
Kane et al. (2024)	15sec	Approximately 61	unknown	Maharashtra, India
Karazor & Kavaz (2023)	1h	457	Short-term (day ahead)	Denmark
Karthikeyan & Jagadeeshwaran (2023)	5min	1096	Short-term (unknown)	Alice Springs, Australia
Khan et al. (2021)	15min (Bunnik) 1h (Breda)	Bunnik: 1096 Breda: 1461	15min ahead (Bunnik) 1h ahead (Breda) Day-ahead (both cities)	Bunnik, Netherlands Breda, Netherlands
Kim et al. (2021)	1h	1461	Hour-ahead	South Korea
Kyeremeh et al. (2022)	3h	990	Short-term (day-ahead)	Germany
Komar et al. (2024)	1h	1461	unknown	Ternopil, Ukraine
Lateko et al. (2021)	1h	365	Day-ahead and Three days-ahead	Taiwan
Lateko et al. (2022)	1h	366	Short-term (day-ahead)	Taiwan
Lin (2023)	undetermined	426	Unknown	Taiwan
Liu et al. (2023a)	15min	730	Short-term (day-ahead)	Oregon, U.S.A.
Liu et al. (2023b)	15min	365	Very-short-term (one and two steps ahead)	Alice Springs, Australia
Loh et al. (2024)	1h	792	unknown	Australia
Markovics & Mayer (2022)	15min	731	Day-ahead and intraday (unknown)	Hungary
Massaoudi et al. (2021)	5min	1371	Short-term (daily)	Alice Springs, Australia
Mohana et al. (2021)	5min	91	unknown	Asir, Saudi Arabia
Mondal et al. (2024)	15min	365	Very/Short-term (15min-1h) Mid-term (1-7 days)	Kolkata, India

Natarajan & Singh (2023)	5min	1096	unknown	Thailand
Nespoli et al. (2022)	1h	269	Short-term (day-ahead)	Milan, Italy
Peng et al. (2023)	15min	335	Unknown	China
Petrosian & Zhang (2024)	3h	365	Mid-term (7, 15 days) Long-term (30 days)	Berkeley, U.S.A.
Rai et al. (2022)	undetermined	731	6h ahead	Ghaziabad, India
Rusina et al. (2023)	1h	1096	Short-term (day-ahead)	Mongolia
Sedai et al. (2023)	1h	365	1,3, 5, 15 days ahead	Texas, USA
Sharma et al. (2021)	5min	1258	Hour-ahead; day-ahead;10 days-ahead; month-ahead	Yulara, Australia
Siriwardana et al. (2022)	30sec	29	Very short-term (10min ahead)	Gampaha, Sri Lanka
Song et al. (2022)	5min	28	short-term (unknown)	Australia
Taheri et al. (2024)	1h	365	Short-term (1h)	Italy
Tseng et al. (2024)	10min	731	Short-term (day ahead)	Taiwan
Tung (2024)	15min	334	Short-term (15 min ahead)	Thailand
Visser et al. (2022)	1h	1155	Day-ahead	Utrecht, Netherlands
Wang & Wang (2021)	1h	365	Day-ahead	Macau
Wang et al. (2022a)	1h	365	Short-term (day-ahead)	Macau
Wang et al. (2022b)	15min	731	Short-term (day-ahead)	China
Wang et al. (2023)	15min	730	Short-term (day-ahead)	Alice Springs, Australia
Wu et al. (2022)	5min	365	Short-term (daily)	Alice Springs, Australia
Xia et al. (2023)	1h	731	Short-term (day ahead)	China
Xu et al. (2024)	15min	92	Short-term (unknown)	Hebei, China
Yang et al. (2024)	15min	730	Long-term (month-ahead)	Shanxi, China Sichuan, China
Zhang & Li (2023)	1h	36	Short-term (unknown)	Yangzhou, China
Zhang et al. (2024)	15min	Approximately 70	Short-term (unknown)	Shandong, China
Zheng et al. (2024)	15min	883	unknown	unknown
Zhang & Zhu (2022)	5min (Australia) undetermined (China)	Australia: 1349 China: unknown (17295 records)	Short-term (unknown)	Australia; China
Zhou et al. (2022)	5min	50	Short-term (unknown)	Australia
Zhou et al. (2023a)	15min	31	unknown	unknown
Zhou et al. (2023b)	5min	19	unknown	Gansu Providence, China

Table A.2. Exogenous variables and their corresponding selected variables for the models

Authors (Year)	Exogenous variables	Selected variables
Abdellatif et al. (2022a)	solar irradiation; WS; Temperature (module and ambient)	solar irradiation; WS; Temperature (module and ambient)
Abdellatif et al. (2022b)	Solar radiation; WS	Solar radiation; WS
Abdelmoula et al. (2022)	Temperature (module and ambient); Global Tilted irradiance (GTI); GHI; RH; WS	Temperature (module and ambient); GHI; GTI; WS; RH;
Banik & Biswas (2023)	15 meteorological features	Undetermined (based on Pearson Correlation)
Boutahir et al. (2024)	Ambient and Module Temperature; Irradiation	Ambient and Module Temperature; Irradiation
Chakraborty et al. (2023)	18 meteorological features	WS; Wind Direction (WD); global radiation; diffuse radiation; direct radiation; inclined solar radiation; azimuth
Chen et al. (2024)	Not specified	Global Irradiance
Cicek (2024)	None	None
Córtez et al. (2024)	GHI; Temperature	GHI; Temperature
Dai et al. (2024)	31 weather attributes	11 weather attributes. Performed Pearson Correlation and RF Feature Importance
Guo et al. (2022)	Panel temperature; light intensity; WS; conversion efficiency; WD; etc.	light intensity
Herrera-Casanova et al. (2024)	GHI; WS; WD; RH; Air and Panel Temperature	GHI; RH; Air and Panel Temperature
Hokmabad et al. (2023)	44 weather features	11 weather features (Pearson Correlation; Feature Importance Ranking; Recursive Feature Elimination)
Huang et al. (2023)	WS; WD; GHR; DHR; Temperature; Humidity; daily rainfall; etc.	Undetermined
Kane et al. (2024)	Irradiance; temperature; WS; WD; precipitation; Humidity	Irradiance; temperature; WS; precipitation
Karazor & Kavaz (2023)	GHI; Temperature; Humidity; Mean and Std of WS previous horizon; WS; WD; Plane of array irradiance; Pressure; Power produced from the wind turbine	GHI; Temperature; Humidity; Mean and Std of WS previous horizon; WS; WD; Plane of array irradiance; Pressure; Power produced from the wind turbine
Karthikeyan & Jagadeeshwaran (2023)	Not specified	Not specified
Khan et al. (2021)	GHI; temperature; RH	GHI; temperature; RH
Kim et al. (2021)	ambient temperature; humidity; precipitation; air pollutants; WS; WD; amount and thickness of cloud; amount of irradiance; amount and thickness of Particulate matter (PM)	Ambient temperature; humidity; precipitation; air pollutants; WS; WD; amount and thickness of cloud; amount of irradiance; amount and thickness of PM
Komar et al. (2024)	Air temperature; Humidity; Pressure; WS; Cloudiness; Cloud transparency; Direct Radiation; Scattered Radiation	Air temperature; Humidity; Pressure; WS; Cloudiness; Cloud transparency; Direct Radiation; Scattered Radiation

Kyeremeh et al. (2022)	15 weather attributes (Temperature; sun position; solar radiation diffused; solar radiation direct; RH; WS; etc.)	Undetermined. PCA performed
Lateko et al. (2021)	air temperature; RH; WS	air temperature; RH; WS
Lateko et al. (2022)	GHI; air temperature; precipitation; RH; WS	GHI; air temperature; precipitation; RH; WS
Lin (2023)	sun meter; air temperature; RH; WS; WD; cloud coverage; UV indicator; rain probability	sun meter; air temperature; RH; WS; WD; cloud coverage; UV indicator; rain probability
Liu et al. (2023a)	GHI; Air temperature	GHI; Air temperature
Liu et al. (2023b)	GHI; DHI; Temperature; RH; WS	Undetermined
Loh et al. (2024)	12 weather attributes	Not specified
Markovics & Mayer (2022)	GHI; Ambient temperature; WS; solar azimuth, elevation angle; zenith angle	GHI; Ambient temperature; WS; solar azimuth, elevation angle; zenith angle
Massaoudi et al. (2021)	GHI; DHI; RH; WD; temperature	GHI; DHI; RH; WD; temperature
Mohana et al. (2021)	ambient temperature; RH; WS; WD, solar radiation; rain rate; etc.	Undetermined. Various combinations tested
Mondal et al. (2024)	Solar Radiation; Average temperature; RH; WS; WD; Air Pressure; Module temperature	Solar Radiation; Average temperature; RH; WS; WD; Air Pressure; Module temperature
Natarajan & Singh (2023)	ambient temperature; solar radiation; status (day or night); panel temperature	ambient temperature; solar radiation; status (day or night); panel temperature
Nespoli et al. (2022)	GHI; Global plane of array irradiance (GPOA); Ambient temperature	Undetermined
Peng et al. (2023)	Air pressure; WD; WS; temperature; Humidity; Irradiance	Irradiance; Humidity; temperature
Petrosian & Zhang (2024)	Sky cover; Distance to Solar noon; avg_temperature; avg_WD; avg_WS; RH; Pressure	Sky cover; Distance to Solar noon; Avg_temperature; avg_WD; avg_WS; RH; Pressure
Rai et al. (2022)	reflected radiation; direct radiation; diffused radiation; sun height; temperature; WS	Undetermined
Rusina et al. (2023)	OLAR temperature; cloud cover; precipitation	temperature; cloud cover; precipitation
Sedai et al. (2023)	DNI; GHI; DHI; Temperature; WS	DNI; GHI; DHI; Temperature; WS
Sharma et al. (2021)	Ambient temperature; rainfall; WS; WD; RH; GHI; DHI	WS; ambient temperature; RH; GHI; WD
Siriwardana et al. (2022)	sky images	sky images
Song et al. (2022)	solar radiation; scattered radiation; temperature; RH; WS; WD; rainfall	solar radiation; scattered radiation; temperature; RH; WS;
Taheri et al. (2024)	beam horizontal irradiance (BHI); DHI; GHI; Temperature	DHI; BHI; GHI; Temperature
Tseng et al. (2024)	daily total irradiance; daily average temperature; amplitude between avg. irradiance and maximum irradiance; irradiance std	daily total irradiance, daily average temperature, amplitude between avg. irradiance and maximum irradiance; irradiance std
Tung (2024)	solar radiation; ambient and module temperature;	solar radiation; ambient and module temperature;
Visser et al. (2022)	air pressure; air temperature; total precipitation; total cloud cover; low cloud cover; clear sky irradiance; zenith and azimuth angle; etc.	Undetermined

Wang & Wang (2021)	temperature; humidity; WD; WS; precipitation cloud amount; atmospheric pressure	temperature; humidity; WD; WS; precipitation cloud amount; atmospheric pressure
Wang et al. (2022a)	solar irradiation; atmospheric temperature; WS; WD; wind angle; atmospheric pressure; RH; precipitation; cloud amount; similar day power	solar irradiation; atmospheric temperature; similar day power; RH and cloud amount
Wang et al. (2022b)	22 numerical weather predictions (Air temperature; cloud cover; WS; WD; etc.)	Undetermined. Selected by GBDT
Wang et al. (2023)	solar irradiance; ambient temperature; RH; WS	solar irradiance; ambient temperature; RH
Wu et al. (2022)	temperature; humidity; WD; rainfall; GHI; DHI; RGT; RDT	GHI; DHI; humidity; temperature; RGT and RDT
Xia et al. (2023)	Not specified (mention irradiance)	Not specified (mention irradiance)
Xu et al. (2024)	Irradiance; Temperature; Pressure; Humidity; WD; WS	Irradiance; Temperature; Humidity
Yang et al. (2024)	Not specified	Downward solar short-wave radiation; Total sky direct solar radiation; Direct solar radiation
Zhang & Li (2023)	Temperature; Radiation	Temperature; Radiation
Zhang & Zhu (2022)	Australia: Temperature; Humidity; Global Horizontal Radiation (GHR); Global Horizontal Radiation (DHR); WD; daily rainfall; Radiation Diffuse tilted (RDT); Radiation Global Tilted (RGT) China: temperature of the panels; temperature; light intensity; WS; WD; Transfer efficiency	Undetermined
Zhang et al. (2024)	Ambient and module temperature; Irradiance; Precipitation; Pressure; Humidity	Ambient and module temperature; Irradiance; Precipitation; Pressure; Humidity
Zheng et al. (2024)	Irradiation; Temperature; Pressure; Cloudiness; Precipitation; WS; etc.	Irradiation; Temperature; Pressure; Cloudiness; Precipitation; WS; etc.
Zhou et al. (2022)	WS; air temperature; RH; global radiation; diffuse radiation; WD and rainfall	Undetermined. Performed successive variational mode decomposition and PCA
Zhou et al. (2023a)	Temperature; Snowfall depth; Azimuth; Ground pressure; Dew point temperature; Zenith angle; DHI; FTI; TTI; Direct normal irradiance (DNI); Atmospheric precipitable water; WD; RH; WS; GHI; Similar day power	GHI; FTI; TTI; DHI; zenith angle; Similar day power
Zhou et al. (2023b)	global radiation; direct radiation; diffuser radiation; air and cell temperature; WS; WD; pressure; RH	global radiation; direct radiation; diffuser radiation; air and cell temperature; WS

Table A.3. Proposed ensemble models and models used for comparison of the results

Authors (Year)	Proposed Ensemble Models	Comparison Models
Abdellatif et al. (2022a)	XGBoost-Adaboost-RF→ETR	RF; XGBoost; Decision Tree Regressor (DTR); Adaboost; ETR; Stack-RF; Stack-Adaboost; Stack-XGBoost
Abdellatif et al. (2022b)	AdaBoost	ETR; RF; GBR; DTR
Abdelmoula et al. (2022)	MLR-RF-XGBoost→Linear Lasso	XGBoost; RF; MLR; PVWatts (physical model)
Banik & Biswas (2023)	RF-Categorical Boosting with Bayesian model averaging	Base learners; XGBoost; LGB; DTR
Boutahir et al. (2024)	Boosting Cascade Forest	SARIMAX; Prophet; VMD-SVM; NN; Auto Encoder LSTM
Chakraborty et al. (2023)	Voting regressor (LGB-ETR-XGBoost-HGB) Stacking (LGB-ETR-XGBoost-HGB)	LR; Elastic net; Bayesian ridge; KNN; Base learners; RF; GB; AdaBoost
Chen et al. (2024)	TimeGAN-(SVM-ELMNN-LSTM) →POS_BPNN	SCM; ELM; Stacking (without TimeGAN)
Cicek (2024)	SVR-LSTM→Linear Regression	SVR; LSTM
Córtez et al. (2024)	Fuzzy Ensemble Algorithm (based on statistical and LSTM models)	Single models
Dai et al. (2024)	Optimized BiGRU model with bagging ensemble	SVM; XGBoost; TCN; BiLSTM; BiGRU; ensemble optimized BiLSTM
Guo et al. (2022)	Adaptive ensemble (AE)-Stochastic Configuration networks (SCN)	LR; RF; ARMA; SCN; Ensemble SCN; Random Vector Functional-link Network
Herrera-Casanova et al. (2024)	RF	MLP; BiLSTM
Hokmabad et al. (2023)	LightGBM-LSTM→GBR	Persistence; LSTM (1-step and multi-step); LightGBM; XGBoost
Huang et al. (2023)	10 DeepTCN with Multiverse Optimizer based on no-negative constraint theory weight integration called Complete Ensemble Empirical Mode Decomposition with Adaptive Noises (CEEMDAN) DeepTCN	ARIMA; ELM; GRU; CNN; LSTM; DeepTCN; GRU-CNN; CNN-LSTM; Empirical Modal Decomposition (EMD)-DeepTCN; Variational Model Decomposition (VMD)-DeepTCN; CEEMDAN-CNN; CEEMDAN-LSTM; CEEMDAN-DeepTCN); EMD-DeepTCN-WOA; VMD-DeepTCN-PSO
Kane et al. (2024)	LR-FT-TCN-SVM→LR	LR; NN; SVM; XGBoost; FT; TCN; LSTM
Karazor & Kavaz (2023)	XGBoost-LSTM→Ridge Regression	CNN; LSTM; XGBoost; LightGBM; SVR; XGBoost-LightGBM -> Ridge Regression XGBoost-CNN -> Ridge Regression
Karthikeyan & Jagadeeshwaran (2023)	Wavelet Packet decomposition – LSTM-Weighted sum	CNN; DNN; SVM; LR; DT
Khan et al. (2021)	ANN-LSTM→XGBoost	Single-learners; Bagging

Kim et al. (2021)	SARIMAX→LSTM	SARIMAX; SVR; LSTM; DNN; RF
Komar et al. (2024)	Gradient Boosting	LightGBM; XGBoost; CatBoost; RF; AdaBoost; SVM
Kyeremeh et al. (2022)	LSTM-AdaBoost	LSTM; MLP; SVR; RF; Radial Basis Function Neural; Ensembles with these single-models
Lateko et al. (2021)	ANN-DNN-SVR-LSTM-CNN→RNN	Single-learners; RF; RNN Stacking of DNN, SVR and ANN
Lateko et al. (2022)	5 RFs ensembled by Linear Regression (LR) or SVR depending on K-means weather clustering, and SVM classification	Individual RFs; Stacking-RNN
Lin (2023)	Differential Evolution for hyperparameter tuning of GBR	Jaya; PSO; GA; Random search (all using GBR)
Liu et al. (2023a)	GRNN-ELM-ElmanNN-LSTM→BPNN	Base models; Traditional Stacking;
Liu et al. (2023b)	CNN-BiLSTM-LSTM→ELM	BPNN; LSTM; ELM; SVM; VMD-LSTM, EEMD-LSTM; VMD-CNN-BiLSTM; Ensemble Empirical Modal Decomposition-LSTM
Loh et al. (2024)	Numerical-Categorical-RBF-DNN ensemble	RF; DNN
Markovics & Mayer (2022)	ETR	LR; XGBoost; LGB
Massaoudi et al. (2021)	ELM-ETR-KNN-MF→DBN	ELM; KNN; ETR; DBN; MF; MLP; GRU; BiLSTM; LSTM; Ensemble models from literature
Mondal et al. (2024)	Two Bi-LSTM averaging	LSTM; GRU; RNN; ELM; SVR
Mohana et al. (2021)	RF; XGBoost	Lasso; LR; PR; SVM; Deep Learning
Natarajan & Singh (2023)	MLP-SVR-RF-XGB→Bayesian Regression	SVR-MLP; DBN-SVM-RF; XGB-CategoricalBoost-LGB-RF
Nespoli et al. (2022)	Selective ensemble of Multilayer perceptron (MLP)	CNN; LSTM
Peng et al. (2023)	LGB to incorporate two Deep CNN-LSTM	LSTM; CNN; BPNN
Petrosian & Zhang (2024)	LightGBM	XGB; CatBoost; ANN; RNN; LSTM; GRU; Bi-RNN; Bi-LSTM; Bi-GRU
Rai et al. (2022)	ETR	LGB; XGBoost; RF; BR; GBR; LASSO; Ridge Regression; KNN; DTR; AdaBoost: Elastic net; Orthogonal Matching Pursuit; Least Angle Regression; SVM with linear kernel
Rusina et al. (2023)	AdaBoost	RF; XGBoost
Sedai et al. (2023)	RF	ARIMA-LSTM; ARIMA; GRU; SVR; Multiple GRU; Multiple LSTM; Encoder-Decoder LSTM; BiLSTM
Sharma et al. (2021)	MODWT-LSTM weighted average	DWT-LSTM; LSTM; RNN; GRU; Adaptive network-based fuzzy inference system
Siriwardana et al. (2022)	Weighted average of sky imagery model and statistical model both developed using BiLSTM	sky imagery model; statistical model

Song et al. (2022)	XGBoost-GRU-RF-LSSVM→LSSVM	XGBoost; LSTM
Taheri et al. (2024)	Stacking Ensemble (not specified)	SVR; RF; KR; Knn; GBR
Tseng et al. (2024)	(SOM and K-means Weather clustering)→(BPNN, LSTM, Bi-LSTM, GRU)-> ANN	Bi-LSTM; LSTM; BPNN; GRU_attention based
Tung (2024e)	SVR-LSSVR-LASSO-RIDGE→GBR	SVR; LSSVR; LASSO; RIDGE
Visser et al. (2022)	RF	MLR; SARIMAX; LASSO; Linear and Nonlinear kernel variant of SVR; Gradient Boosting; ANN; LSTM; Persistance Models; Physical Model
Wang & Wang (2021)	Weighted sum of 3 MLP independent models	MLP
Wang et al. (2022a)	ELM→Evidential regression	MLP; ELM; Evidential regression; SVM
Wang et al. (2022b)	TabNet and XGBoost combined by linear regression	TabNet; XGBoost; SVM; MLP; CNN; CNN-LSTM
Wang et al. (2023)	XGBoost-RF-CategoricalBoosting→LSTM	BPNN; XGBoost; RF; CategoricalBoosting; LSTM
Wu et al. (2022)	XGB-LGB-RF-CategoricalBoost→LGB	XGBoost; LGB; SVR; RF
Xia et al. (2023)	Fiting residual random forest ensemble	RF; Fitted residual RF w/o ensemble
Xu et al. (2024)	ICEEMDAN-Bagging-XGBoost	XGBoost; MLR; Transformer; LSTM; BPNN
Zhang & Li (2023)	LSTM-LSSVM-KELM with optimized VMD	LSTM; KELM; LSSVM; ensemble of the single-models using OVMD
Zhang & Zhu (2022)	XGBoost-LGB-RF→GBDT	LR; RF; GBDT; LGB; XGBoost; STACKING-GBDT; STACKING-LGB; STACKING-GB; STACKING-RF
Zhang et al. (2024)	BiLSTM-XGBoost	RF; CNN-LSTM; XGBoost; BiLSTM
Zheng et al. (2024)	MLP-LightGBM-LSTM with improved error fusion	MLP; LightGBM; LSTM
Zhou et al. (2022)	Linear combination of Back Propagation Adaboost; Relevance Vector Machine; ELM and RF	Base learners
Zhou et al. (2023a)	XGBoost-SFS-DNS	Traditional Stacking; SVR; GBDT; XGBoost
Zhou et al. (2023b)	XGBoost	AR; ARIMA; LSTM; GRU; Temporal Convolutional Networks (TCN); GBDT

Note: in the “Proposed Ensemble Models” column, the right side of the arrow “→” denotes the meta-learner of stacking models and the left side denotes the base learners separated by a dash “ - “

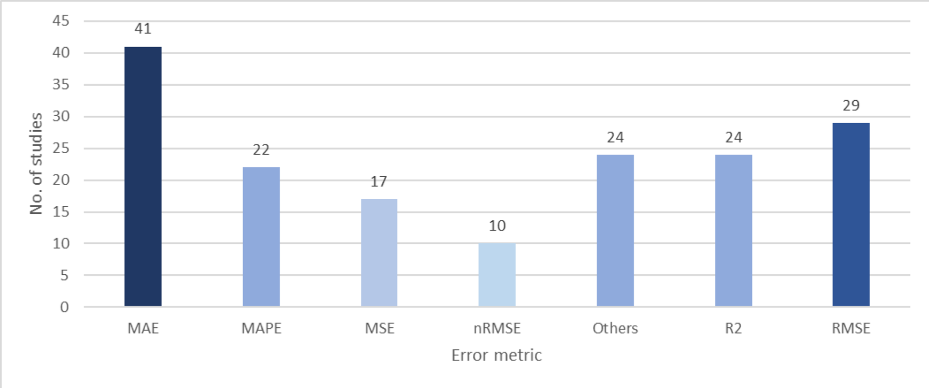


Figure A.1. Error metrics used across the reviewed studies

Appendix B. Methodology: Stationarity study

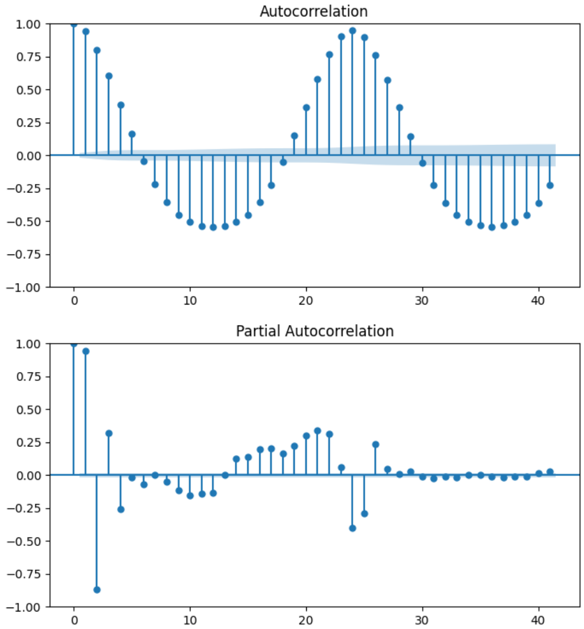


Figure B.1. ACF and PACF plots of the original PV Power series

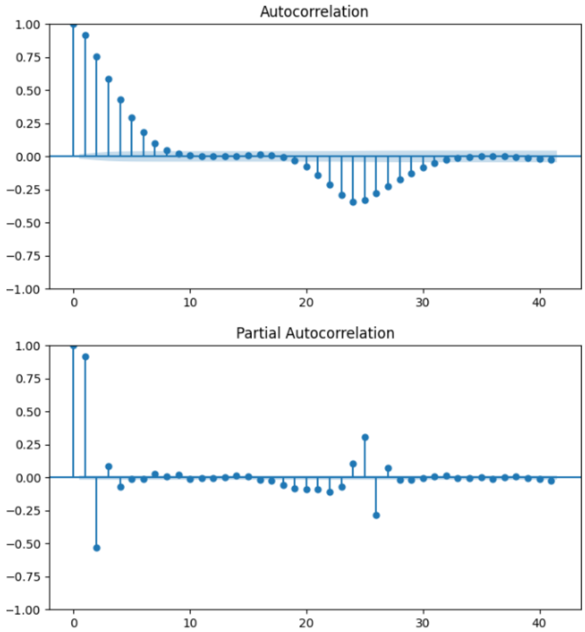


Figure B.2. ACF and PACF plots of the seasonal differenced PV Power series

Appendix C. Model results analysis: Univariate Approach

As a remark, in the different tables, the column “All vars” indicates that none of the variables were excluded from the model training.

XGBoost

Table C.1. Effect of using 48 instead of 24 lags in XGBoost across the performance metrics

Steps ahead	Training size	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	No	29.98	19.31	3.10	-5.01	-3.70	0.20	-16.82	-13.50	-2.13	1.10	1.14	-0.12	1.76	-5.34	0.14	4.14	3.81	0.38
1	8736	Yes	-6.77	-6.56	-7.05	5.65	-1.05	-2.33	5.65	-1.05	-2.33	7.24	1.73	-3.22	-1.23	-4.92	-7.95	-6.04	-6.25	-3.46
1	12360	No	-11.38	-6.49	0.78	0.96	1.24	0.26	-15.82	-11.07	-0.18	-3.33	0.20	1.59	18.21	13.56	2.57	-15.18	-6.84	0.44
1	12360	Yes	2.12	0.70	-0.92	3.04	-0.92	-2.75	3.04	-0.92	-2.75	-0.99	-2.49	-3.99	3.53	-3.08	-10.11	-8.63	-7.67	-6.86
2	8736	No	-10.20	-7.97	-2.17	-4.25	-4.74	0.04	-0.12	-2.04	-0.24	-0.97	0.41	-1.93	11.47	7.80	-0.52	2.52	0.66	-2.12
2	8736	Yes	1.88	-0.91	-8.47	6.59	1.81	-1.69	6.59	1.81	-1.69	-1.45	-2.57	-4.26	3.68	1.92	-6.26	-3.07	-3.37	-4.88
2	12360	No	6.08	4.03	0.41	-14.45	-10.58	0.03	2.80	-1.28	0.70	-8.17	-9.42	-0.01	7.84	4.78	1.79	0.05	-0.85	0.92
2	12360	Yes	-2.93	-1.45	-1.01	2.51	-1.23	-6.34	2.51	-1.23	-6.34	8.46	5.55	-4.45	2.98	-1.66	-4.78	0.07	0.83	-3.16
3	8736	No	8.05	7.34	3.02	-9.38	-7.23	0.16	-12.72	-6.12	0.62	1.39	2.44	-0.68	-22.02	-15.52	-0.16	-10.96	-6.70	0.78
3	8736	Yes	1.05	-0.36	-2.35	5.03	0.22	-5.03	5.03	0.22	-5.03	1.76	-0.51	-2.52	3.04	0.52	-4.84	-0.41	-2.56	-4.46
3	12360	No	-4.16	0.33	-0.18	-13.40	-11.28	-0.81	-4.46	-4.52	0.12	-3.30	-4.48	0.14	-10.18	-7.26	-0.53	-13.64	-7.50	0.23
3	12360	Yes	6.02	5.34	1.01	7.66	2.25	-4.22	7.66	2.25	-4.22	5.68	1.89	-7.49	5.88	0.98	-10.03	5.70	1.92	-6.78
4	8736	No	2.55	2.65	1.82	-8.37	-5.85	-1.78	-23.06	-12.35	-1.78	-12.44	-6.99	0.46	6.83	5.28	2.38	2.01	0.00	0.87
4	8736	Yes	4.69	2.18	-2.56	5.12	1.94	-2.56	5.12	1.94	-2.56	3.44	1.00	-3.79	6.42	2.13	-1.87	3.71	1.31	-5.16
4	12360	No	-2.41	-1.40	0.08	7.03	2.40	1.74	5.82	3.80	0.40	-9.71	-7.76	0.31	-5.12	-3.98	0.34	8.62	6.05	2.04
4	12360	Yes	4.74	3.12	0.09	6.30	3.47	-4.68	6.30	3.47	-4.68	2.87	0.97	-6.42	-0.06	-0.92	-5.93	0.28	0.95	-2.53
6	8736	No	-5.75	-3.59	-0.15	-1.59	-1.77	2.27	-2.70	-1.10	1.80	-2.06	-0.74	1.56	15.33	12.18	1.88	4.01	-1.63	0.58
6	8736	Yes	4.92	3.76	-3.71	3.31	2.61	-3.89	2.84	2.25	-3.95	4.54	3.20	-5.17	5.41	2.34	-7.62	5.20	3.33	-5.97
6	12360	No	-2.14	-4.99	0.20	-4.88	-1.92	-0.17	-15.41	-10.61	-0.84	2.76	2.16	0.62	-4.92	-2.72	-0.36	-2.98	-3.68	-0.13
6	12360	Yes	4.34	2.46	0.83	8.41	5.98	-6.34	8.41	5.98	-6.34	1.90	0.20	-4.12	3.86	-0.76	-8.40	8.85	5.76	-5.03
8	8736	No	-0.20	0.97	0.29	-4.04	-3.99	-0.85	0.79	-0.74	-0.66	5.92	2.99	-2.23	2.87	0.26	0.60	2.30	0.58	0.35
8	8736	Yes	5.58	1.77	-7.24	4.23	1.91	-2.88	4.20	1.87	-3.00	3.16	0.24	-7.00	5.38	1.58	-8.35	7.29	3.81	-2.02
8	12360	No	-1.74	-1.55	0.75	-10.27	-5.18	0.31	-2.35	-2.19	1.12	4.09	1.52	0.75	9.35	4.54	1.00	10.60	4.79	0.97
8	12360	Yes	1.79	2.10	-2.34	4.95	2.97	-4.50	4.95	2.97	-4.50	-0.18	-2.10	-13.73	0.28	-2.56	-9.42	8.80	4.16	-10.01
12	8736	No	4.35	2.64	0.56	4.13	4.87	1.74	-0.99	0.38	1.11	-0.65	0.92	0.84	6.97	5.19	1.58	0.16	1.11	1.08
12	8736	Yes	3.17	2.66	-5.51	-0.14	0.91	-3.55	-0.15	0.85	-3.51	7.71	4.77	-4.86	4.96	2.76	-7.56	6.15	3.81	-4.50
12	12360	No	-2.94	-2.07	-0.18	5.69	6.31	0.64	-2.03	-3.14	0.33	6.85	3.58	1.79	8.67	5.35	1.13	-2.28	0.78	1.11
12	12360	Yes	5.27	4.02	-0.95	4.11	2.19	-9.25	4.11	2.19	-9.25	8.74	3.13	-5.24	6.39	4.11	-3.68	8.12	2.80	-4.38
24	8736	No	-0.67	-0.76	-0.39	0.81	-0.06	1.59	4.10	1.26	1.49	1.19	0.26	1.29	1.35	1.68	1.45	0.72	2.67	1.41
24	8736	Yes	-2.89	-0.66	-4.03	-1.59	-0.27	-3.08	-1.41	-0.21	-2.57	-1.07	-1.62	-3.44	-9.43	-5.37	-4.12	-3.64	-1.89	-3.19
24	12360	No	-1.68	-0.42	0.21	13.17	8.79	0.71	-1.20	-0.74	0.16	9.19	5.08	0.05	1.63	3.07	0.24	2.85	3.41	0.98
24	12360	Yes	4.68	2.97	-2.43	9.38	3.99	-11.79	9.38	3.99	-11.79	5.43	0.66	-10.36	1.23	2.03	-5.77	-0.08	-3.11	-11.39

Table C.2. Effect of using installed capacity scaling in XGBoost across the performance metrics

Steps ahead	Training size	Lags	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	24	46.29	30.75	11.66	18.68	12.03	6.42	3.02	0.93	4.92	22.43	16.45	4.74	23.70	14.88	10.40	6.28	2.49	3.39
1	8736	48	9.55	4.88	1.51	29.34	14.69	3.89	25.49	13.38	4.73	28.56	17.05	1.64	20.70	15.30	2.30	-3.90	-7.57	-0.45
1	12360	24	4.03	2.26	3.33	42.98	28.76	15.21	33.52	23.33	14.86	36.51	28.25	15.03	50.11	36.29	19.12	36.68	25.04	13.45
1	12360	48	17.53	9.45	1.64	45.06	26.61	12.21	52.38	33.48	12.29	38.85	25.56	9.45	35.43	19.65	6.45	43.23	24.20	6.14
2	8736	24	-10.80	-5.87	7.60	-6.55	-3.97	3.23	-8.73	-5.32	2.85	-5.69	-3.04	1.59	14.25	9.34	6.60	16.50	9.56	6.38
2	8736	48	1.27	1.19	1.30	4.28	2.58	1.51	-2.02	-1.47	1.41	-6.16	-6.02	-0.75	6.46	3.46	0.86	10.91	5.52	3.62
2	12360	24	16.49	10.19	2.36	8.64	7.61	16.10	18.74	13.15	16.42	17.68	10.15	13.73	41.06	27.35	14.63	30.39	17.20	14.68
2	12360	48	7.47	4.71	0.93	25.60	16.96	9.73	18.45	13.20	9.37	34.32	25.12	9.29	36.20	20.91	8.06	30.42	18.88	10.60
3	8736	24	18.19	13.24	4.21	8.81	6.00	5.75	-5.13	-1.46	4.60	6.97	5.47	1.67	-1.47	-1.39	6.11	16.96	11.28	5.70
3	8736	48	11.20	5.54	-1.16	23.22	13.45	0.57	12.62	4.87	-1.04	7.35	2.52	-0.17	23.58	14.64	1.43	27.51	15.42	0.46
3	12360	24	10.93	7.51	0.85	8.21	4.86	10.39	19.14	13.20	11.71	26.67	18.63	16.50	14.97	11.39	16.41	8.70	7.93	13.43
3	12360	48	21.11	12.52	2.04	29.28	18.39	6.98	31.26	19.97	7.37	35.65	25.00	8.87	31.02	19.62	6.91	28.04	17.35	6.42
4	8736	24	2.35	0.88	4.38	3.07	2.67	5.12	-10.74	-4.69	4.96	-9.79	-5.27	4.58	8.38	5.82	5.34	2.28	2.22	6.22
4	8736	48	4.49	0.41	0.00	16.56	10.46	4.33	17.44	9.60	4.18	6.09	2.73	0.33	7.97	2.66	1.10	3.98	3.54	0.19
4	12360	24	7.08	5.53	0.66	30.44	19.78	16.00	31.40	21.45	15.84	17.66	13.24	15.53	19.72	12.09	13.80	23.97	16.00	12.90
4	12360	48	14.23	10.05	0.67	29.72	20.85	9.59	31.88	21.13	10.77	30.24	21.97	8.80	24.78	15.15	7.53	15.63	10.90	8.34
6	8736	24	-3.30	-3.63	3.60	5.97	2.17	3.73	-3.89	-4.66	4.03	0.95	-0.73	6.27	10.46	7.12	8.48	-3.07	-3.17	5.89
6	8736	48	7.37	3.71	0.04	10.87	6.55	-2.43	1.65	-1.31	-1.71	7.54	3.21	-0.46	0.55	-2.73	-1.02	-1.88	1.79	-0.66
6	12360	24	7.14	4.98	0.84	20.91	13.72	14.66	17.16	10.55	14.90	28.34	22.26	14.63	23.53	19.52	16.37	21.65	15.22	13.61
6	12360	48	13.62	12.44	1.46	34.20	21.61	8.49	40.98	27.14	9.40	27.48	20.31	9.89	32.32	21.48	8.32	33.49	24.66	8.71
8	8736	24	6.82	4.31	7.55	5.76	1.14	4.26	9.36	3.81	4.76	2.08	1.42	5.60	7.27	2.60	9.46	7.37	2.06	3.80
8	8736	48	12.60	5.11	0.03	14.03	7.03	2.23	12.77	6.41	2.42	-0.68	-1.33	0.82	9.77	3.92	0.51	12.37	5.29	1.43
8	12360	24	13.21	5.41	3.57	13.78	10.42	15.83	17.93	12.07	15.99	24.68	18.62	24.53	21.35	15.04	20.04	21.99	15.42	18.99
8	12360	48	16.74	9.06	0.47	29.01	18.57	11.01	25.23	17.23	10.37	20.42	15.00	10.04	12.28	7.93	9.62	20.18	14.80	8.01
12	8736	24	1.69	-1.12	6.42	1.48	2.31	7.41	-4.19	-2.68	7.20	-5.96	-4.35	6.12	0.94	2.02	10.95	-4.65	-2.58	6.56
12	8736	48	0.51	-1.10	0.35	-2.80	-1.66	2.12	-3.35	-2.20	2.58	2.39	-0.50	0.43	-1.07	-0.41	1.80	1.34	0.13	0.99
12	12360	24	9.69	6.91	2.69	27.79	21.47	17.62	21.16	15.66	17.79	28.87	23.24	16.98	30.13	19.90	15.46	21.40	18.50	15.23
12	12360	48	17.90	13.00	1.91	26.21	17.34	7.73	27.31	20.99	8.21	30.76	22.79	9.94	27.85	18.67	10.65	31.80	20.52	9.73
24	8736	24	14.51	6.18	7.01	18.10	9.53	5.37	17.10	8.69	5.24	15.21	8.34	5.80	16.66	9.75	7.54	12.86	8.16	6.20
24	8736	48	12.30	6.28	3.37	15.70	9.31	0.70	11.59	7.22	1.18	12.95	6.46	1.08	5.88	2.69	1.97	8.50	3.60	1.61
24	12360	24	12.27	8.39	3.04	25.74	19.74	20.26	22.82	16.61	19.74	24.71	21.34	20.98	23.54	19.90	16.03	23.19	21.30	19.73
24	12360	48	18.62	11.78	0.40	21.95	14.94	7.75	33.40	21.35	7.80	20.95	16.92	10.56	23.13	18.87	10.01	20.26	14.77	7.36

Table C.3. Effect of using 12360 instead of 8376 hours for XGBoost training across the performance metrics

Steps ahead	Lags	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
1	24	No	12.83	8.34	-0.12	-12.07	-8.19	-2.30	-18.26	-13.86	-3.44	-5.60	-4.76	-3.65	-24.40	-18.33	-2.55	-17.29	-12.66	-2.77
1	24	Yes	-29.43	-20.15	-8.45	12.23	8.54	6.49	12.23	8.54	6.49	8.49	7.04	6.63	2.01	3.08	6.18	13.11	9.90	7.29
1	48	No	-28.53	-17.46	-2.44	-6.09	-3.25	-2.25	-17.27	-11.42	-1.49	-10.03	-5.69	-1.94	-7.95	0.57	-0.12	-36.61	-23.30	-2.70
1	48	Yes	-20.54	-12.89	-2.32	9.62	8.67	6.07	9.62	8.67	6.07	0.26	2.81	5.86	6.78	4.92	4.02	10.52	8.47	3.90
2	24	No	-26.54	-16.20	-3.90	-1.25	-1.46	-3.17	-13.53	-8.36	-3.86	-13.66	-7.19	-5.04	-6.80	-2.43	-3.35	-2.17	-1.38	-3.58
2	24	Yes	0.74	-0.15	-9.15	13.94	10.12	9.70	13.94	10.12	9.70	9.71	5.99	7.10	20.01	15.58	4.69	11.72	6.27	4.72
2	48	No	-10.26	-4.20	-1.31	-11.45	-7.30	-3.18	-10.61	-7.60	-2.92	-20.86	-17.03	-3.12	-10.44	-5.45	-1.04	-4.64	-2.89	-0.54
2	48	Yes	-4.06	-0.68	-1.68	9.86	7.07	5.04	9.86	7.07	5.04	19.62	14.11	6.91	19.31	12.00	6.16	14.87	10.47	6.44
3	24	No	-1.45	0.43	0.69	9.85	6.96	-0.11	-15.02	-8.84	-2.58	-17.45	-8.94	-3.79	-9.71	-5.36	-0.43	7.28	4.35	-0.31
3	24	Yes	-8.71	-5.30	-2.67	9.25	5.82	4.53	9.25	5.82	4.53	2.25	4.22	11.04	6.74	7.41	9.87	-0.98	1.00	7.42
3	48	No	-13.66	-6.59	-2.52	5.83	2.91	-1.08	-6.75	-7.24	-3.08	-22.14	-15.86	-2.96	2.14	2.90	-0.80	4.59	3.55	-0.86
3	48	Yes	-3.74	0.39	0.69	11.89	7.86	5.33	11.89	7.86	5.33	6.17	6.62	6.07	9.58	7.88	4.69	5.13	5.48	5.09
4	24	No	-8.88	-8.29	-0.97	-16.29	-9.78	-2.86	-31.06	-18.82	-2.85	-13.47	-8.50	-2.71	0.43	1.77	-0.55	-11.69	-6.96	-2.05
4	24	Yes	-4.15	-3.64	-4.69	11.07	7.32	8.02	11.07	7.32	8.02	13.98	10.00	8.24	11.76	8.05	7.90	10.00	6.83	4.63
4	48	No	-13.85	-12.34	-2.71	-0.90	-1.54	0.65	-2.18	-2.67	-0.67	-10.74	-9.27	-2.86	-11.53	-7.49	-2.59	-5.08	-0.90	-0.88
4	48	Yes	-4.10	-2.70	-2.04	12.26	8.85	5.91	12.26	8.85	5.91	13.41	9.97	5.61	5.28	5.00	3.84	6.57	6.47	7.27
6	24	No	-17.54	-10.43	-1.90	-8.89	-7.05	-1.06	-15.47	-11.07	-1.05	-18.53	-13.72	-0.74	-1.94	-2.55	0.04	-18.80	-13.24	-1.41
6	24	Yes	-7.10	-1.82	-4.66	6.05	4.50	9.87	5.58	4.14	9.82	8.87	9.28	7.62	11.13	9.85	7.93	5.92	5.15	6.31
6	48	No	-13.93	-11.84	-1.55	-12.18	-7.20	-3.50	-28.18	-20.59	-3.69	-13.71	-10.81	-1.68	-22.19	-17.46	-2.19	-25.80	-15.30	-2.12
6	48	Yes	-7.67	-3.11	-0.12	11.15	7.86	7.42	11.15	7.86	7.42	6.23	6.28	8.67	9.58	6.75	7.15	9.57	7.58	7.25
8	24	No	-0.36	1.01	-1.15	-1.12	-2.54	-2.61	-1.70	-1.56	-2.38	-9.68	-5.98	-4.42	-4.47	-3.58	-1.12	-6.65	-4.49	-1.49
8	24	Yes	6.03	2.11	-5.14	6.90	6.74	8.97	6.87	6.71	8.85	12.92	11.22	14.51	9.61	8.85	9.46	7.97	8.88	13.71
8	48	No	-1.90	-1.51	-0.69	-7.34	-3.74	-1.44	-4.83	-3.01	-0.60	-11.51	-7.44	-1.44	2.01	0.70	-0.72	1.65	-0.28	-0.86
8	48	Yes	2.24	2.44	-0.24	7.63	7.81	7.35	7.63	7.81	7.35	9.58	8.88	7.78	4.52	4.71	8.39	9.47	9.23	5.72
12	24	No	-9.28	-7.50	-1.39	-14.46	-11.35	-1.12	-13.40	-10.49	-1.45	-19.33	-15.61	-2.18	-17.50	-11.38	0.17	-15.22	-11.53	-1.58
12	24	Yes	-1.28	0.52	-5.11	11.85	7.81	9.09	11.96	7.84	9.14	15.50	11.98	8.67	11.69	6.51	4.68	10.84	9.55	7.09
12	48	No	-16.57	-12.21	-2.12	-12.90	-9.90	-2.22	-14.44	-14.01	-2.23	-11.84	-12.95	-1.22	-15.80	-11.23	-0.28	-17.65	-11.85	-1.54
12	48	Yes	0.82	1.89	-0.56	16.10	9.09	3.39	16.21	9.19	3.40	16.54	10.34	8.29	13.12	7.86	8.56	12.81	8.53	7.21
24	24	No	-9.16	-7.08	-0.45	-10.93	-7.96	-0.65	-8.87	-5.64	0.06	-10.97	-8.05	1.24	-6.71	-7.09	0.26	-9.41	-6.93	0.03
24	24	Yes	-11.41	-4.87	-4.42	-3.29	2.26	14.23	-3.15	2.29	14.56	-1.47	4.95	16.41	0.16	3.07	8.75	0.92	6.20	13.55
24	48	No	-10.17	-6.74	0.15	1.44	0.89	-1.53	-14.18	-7.65	-1.27	-2.96	-3.23	0.00	-6.44	-5.71	-0.95	-7.28	-6.19	-0.40
24	48	Yes	-3.85	-1.24	-2.82	7.68	6.52	5.52	7.63	6.48	5.34	5.04	7.23	9.49	10.82	10.47	7.10	4.48	4.99	5.35
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table C.4. Summary of the effects of using 48 instead of 24 lags of PV power in XGBoost

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	0.36	-4.80	80	42%
MAE	0.15	-3.50	77	40%
MAPE	-1.11	-2.50	115	60%
Total	**-0.20**	**-3.46**	**272**	**47%**

Table C.5. Summary of the effects of using the maximum installed capacity scaling in XGBoost

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	10.80	-4.98	23	12%
MAE	6.42	-3.01	29	15%
MAPE	2.76	-0.72	14	7%
Total	**6.66**	**-3.21**	**66**	**11%**

Table C.6. Summary of the effects of using 12360 instead of 8736 hours to train XGBoost

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-7.93	-10.21	160	83%
MAE	-5.05	-6.69	155	81%
MAPE	-2.26	-2.53	176	92%
Total	**-5.08**	**-6.35**	**491**	**85%**

Random Forest

Table C.7. Effect of using 48 instead of 24 lags in RF across the performance metrics

Steps ahead	Training size	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	No	-4.49	0.43	0.49	0.54	1.43	-0.40	2.46	5.40	0.36	-7.67	-4.48	-1.77	-19.71	-12.41	-1.18	-5.83	-1.60	-1.55
1	8736	Yes	-10.30	-5.65	-0.08	-14.86	-10.47	-0.77	-3.10	-2.96	0.58	-9.38	-7.04	-0.31	-2.39	-2.95	0.11	-2.37	-2.84	-0.04
1	12360	No	4.65	2.21	0.35	1.23	0.68	1.18	8.71	6.23	1.02	7.51	5.07	0.88	2.80	1.82	0.50	1.12	1.16	-0.80
1	12360	Yes	-0.42	-2.19	-0.54	-1.52	-2.55	-0.16	-2.24	-1.90	-0.06	-12.45	-9.63	-0.65	-5.95	-5.93	-0.52	-8.51	-6.00	-0.54
2	8736	No	-9.99	-5.34	-2.16	-12.86	-6.73	-1.49	-14.01	-8.04	-1.87	-18.62	-11.30	-2.61	-7.70	-4.02	-1.63	-16.54	-9.19	-3.35
2	8736	Yes	4.90	2.41	-0.16	9.03	5.03	0.33	7.54	3.71	0.19	2.11	0.69	-0.19	-9.91	-6.99	-0.85	1.97	1.54	-0.51
2	12360	No	-3.09	-0.03	0.27	-5.11	-3.60	-0.48	-1.33	-0.79	-0.15	-2.86	-1.19	0.37	-3.87	-3.53	-0.84	-6.38	-3.17	-0.90
2	12360	Yes	-1.42	-2.92	-1.09	-3.61	-3.94	-0.61	-6.55	-6.01	-0.98	-12.15	-9.39	-1.97	-11.87	-10.56	-1.45	-2.99	-3.96	-0.79
3	8736	No	-0.70	0.50	0.21	-0.62	-1.58	0.48	1.39	0.60	0.56	3.65	1.48	0.12	0.49	-0.54	0.68	6.21	5.32	0.79
3	8736	Yes	-5.80	-1.89	-0.57	-3.03	-0.55	-0.09	-7.25	-2.53	-0.30	-4.97	-2.39	-0.36	-4.28	-4.05	-1.10	-10.05	-6.43	-1.01
3	12360	No	2.74	0.64	0.52	3.11	0.61	0.22	8.74	4.08	0.25	7.94	3.00	0.72	5.58	1.06	0.27	-1.59	-2.00	0.74
3	12360	Yes	0.53	0.99	0.44	-5.74	-4.47	-0.63	-3.46	-4.42	-0.57	-5.78	-3.92	-0.10	-0.05	0.84	0.22	-7.91	-6.33	-0.49
4	8736	No	-4.00	-3.44	-1.50	3.11	2.85	0.33	-0.01	0.85	0.18	2.46	1.35	-0.31	4.77	2.94	-1.36	-1.37	-0.51	-0.91
4	8736	Yes	-4.38	-3.93	-0.66	-3.85	-2.26	-0.05	-5.89	-4.80	-0.35	-2.52	-0.90	0.23	-4.14	-2.41	0.32	0.19	0.72	0.25
4	12360	No	6.61	3.46	0.86	2.21	1.45	0.54	-5.40	-4.01	-0.08	0.90	0.88	0.37	2.91	1.82	0.39	1.86	1.32	0.26
4	12360	Yes	-3.07	-1.87	0.12	-4.26	-3.08	-0.27	-7.31	-4.50	-0.41	-2.85	-1.19	-0.05	-6.03	-3.32	-0.13	-6.81	-4.72	-0.58
6	8736	No	1.63	1.70	0.17	1.43	2.35	0.42	0.21	2.02	0.47	2.73	3.88	0.00	3.12	3.92	-0.13	-0.57	1.08	-0.32
6	8736	Yes	-1.52	-0.70	-0.13	3.21	2.12	-0.09	6.65	2.57	0.05	3.26	2.47	0.08	3.05	2.15	-0.18	-1.78	-2.39	-0.50
6	12360	No	7.27	3.64	1.07	3.19	0.30	0.69	0.59	-1.06	0.97	10.37	5.23	0.98	9.71	4.44	0.85	5.33	2.25	0.99
6	12360	Yes	-6.13	-4.45	-0.54	-5.69	-1.88	-0.34	-4.20	-1.60	-0.53	-8.70	-4.62	-0.94	-16.19	-9.99	-1.51	-9.58	-5.03	-0.71
8	8736	No	0.07	0.47	-0.39	2.16	0.85	0.03	-0.35	0.10	-0.51	-0.89	0.57	-0.16	4.12	2.62	-0.31	3.45	2.91	-0.32
8	8736	Yes	2.50	1.71	-0.36	-0.69	1.74	0.06	-1.12	2.29	0.16	0.74	1.47	-0.21	1.97	1.87	-0.07	4.84	3.54	0.32
8	12360	No	-3.78	-2.79	0.42	-2.66	-2.06	0.36	-0.96	-0.68	0.67	-0.38	0.72	0.43	-1.88	-2.14	-0.05	-3.17	-2.40	0.50
8	12360	Yes	-18.92	-11.95	-1.49	-16.55	-12.18	-1.32	-26.12	-17.04	-1.88	-13.48	-7.72	-0.83	-14.51	-9.41	-1.34	-11.71	-6.70	-0.78
12	8736	No	4.38	4.31	0.78	9.50	5.90	-0.42	8.50	5.54	-0.11	6.61	2.97	0.32	6.13	4.19	0.41	5.37	2.71	0.20
12	8736	Yes	4.51	3.31	0.01	2.94	1.98	-0.31	3.02	1.02	-0.48	0.82	0.44	-0.70	0.68	0.32	-0.48	-0.35	0.58	-0.58
12	12360	No	1.92	2.25	0.67	6.74	3.43	0.76	2.12	1.29	0.78	6.98	5.22	0.93	-2.21	-1.57	0.74	6.00	2.98	1.02
12	12360	Yes	-7.46	-5.19	-1.34	-5.00	-3.57	-0.90	-8.55	-4.65	-1.04	-8.50	-4.79	-0.83	-10.22	-4.08	-0.69	-13.72	-9.30	-1.11
24	8736	No	3.48	3.15	0.49	0.61	2.74	1.40	8.26	6.72	1.36	3.66	4.38	0.97	1.04	1.92	0.59	0.41	2.20	0.79
24	8736	Yes	-1.10	0.22	-0.39	-5.08	-2.08	-0.65	-3.71	-1.53	-0.72	3.96	2.48	0.01	0.34	1.61	-0.33	2.37	1.99	-0.48
24	12360	No	-5.08	-1.62	-0.28	-3.24	-0.59	0.22	-3.13	-0.58	0.28	-3.77	-1.16	-0.70	-6.89	-3.22	-0.11	-2.32	0.43	0.36
24	12360	Yes	-6.44	-3.28	-0.35	-2.25	-0.11	-0.06	-1.89	-1.04	-0.28	-3.73	-2.88	-0.57	-1.22	0.15	-0.31	4.08	3.09	0.12

Table C.8. Effect of using installed capacity scaling in RF across the performance metrics

Steps ahead	Training size	Lags	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	24	12.80	10.33	0.22	19.37	12.82	-0.25	14.00	13.07	0.23	15.73	10.90	-0.58	2.09	2.25	-1.52	15.44	12.01	-0.61
1	8736	48	6.99	4.25	-0.34	3.98	0.92	-0.62	8.44	4.71	0.44	14.03	8.35	0.88	19.40	11.70	-0.23	18.89	10.77	0.89
1	12360	24	28.10	20.88	0.71	5.65	6.55	-0.72	11.32	9.94	-1.87	30.61	21.63	0.76	23.74	16.58	-0.38	19.84	15.48	-1.88
1	12360	48	23.03	16.48	-0.18	2.90	3.32	-2.06	0.37	1.81	-2.96	10.65	6.94	-0.77	14.99	8.84	-1.40	10.21	8.32	-1.62
2	8736	24	-6.10	-4.26	-1.65	-3.50	-2.17	-1.26	-1.59	-0.77	-1.11	3.81	2.65	-0.77	10.65	6.62	-0.24	-5.79	-3.65	-2.13
2	8736	48	8.79	3.49	0.35	18.39	9.59	0.56	19.96	10.97	0.95	24.54	14.63	1.65	8.45	3.65	0.54	12.73	7.09	0.71
2	12360	24	21.77	16.77	1.03	20.70	14.39	-0.15	28.19	19.65	0.64	32.05	21.98	1.99	29.44	20.19	0.72	23.19	16.49	0.12
2	12360	48	23.44	13.88	-0.32	22.21	14.05	-0.28	22.98	14.44	-0.19	22.77	13.78	-0.36	21.44	13.15	0.10	26.58	15.70	0.23
3	8736	24	14.19	7.93	1.06	3.54	0.48	0.10	11.82	4.83	0.48	19.65	11.60	1.23	13.94	8.11	1.10	22.47	13.57	1.22
3	8736	48	9.09	5.54	0.29	1.13	1.50	-0.47	3.18	1.70	-0.38	11.03	7.72	0.75	9.17	4.61	-0.68	6.21	1.83	-0.59
3	12360	24	19.85	9.70	0.02	19.17	10.72	0.30	25.06	15.78	0.48	29.13	16.53	0.59	19.26	9.36	-0.40	21.90	12.54	0.39
3	12360	48	17.65	10.05	-0.05	10.33	5.63	-0.55	12.86	7.28	-0.34	15.41	9.62	-0.23	13.63	9.15	-0.45	15.58	8.21	-0.84
4	8736	24	14.93	10.25	0.57	14.77	8.54	0.37	15.36	10.60	0.67	19.91	11.27	0.31	22.16	12.99	0.02	16.92	9.38	0.07
4	8736	48	14.56	9.76	1.41	7.82	3.43	-0.01	9.48	4.96	0.14	14.93	9.02	0.85	13.25	7.63	1.69	18.48	10.61	1.23
4	12360	24	16.44	10.71	0.52	12.54	8.54	0.56	8.15	5.32	0.32	11.46	7.05	0.15	18.21	11.62	0.70	15.19	10.09	0.57
4	12360	48	6.76	5.37	-0.21	6.07	4.02	-0.26	6.23	4.83	-0.01	7.71	4.99	-0.27	9.27	6.48	0.18	6.52	4.04	-0.26
6	8736	24	5.44	4.50	0.77	1.35	2.96	0.29	4.88	6.30	0.81	7.20	6.08	0.33	11.96	8.55	0.59	11.13	9.03	0.76
6	8736	48	2.29	2.09	0.47	3.14	2.72	-0.22	11.33	6.86	0.39	7.72	4.67	0.41	11.89	6.77	0.54	9.91	5.56	0.57
6	12360	24	11.56	8.18	0.76	10.16	5.94	0.62	5.94	3.94	0.77	16.74	11.27	1.06	27.02	18.33	2.24	14.95	9.06	1.23
6	12360	48	-1.84	0.08	-0.84	1.28	3.75	-0.41	1.15	3.40	-0.73	-2.33	1.42	-0.85	1.12	3.91	-0.12	0.04	1.78	-0.47
8	8736	24	4.37	2.67	0.15	6.02	2.17	-0.52	8.24	4.00	-0.38	9.56	5.93	0.60	7.53	4.26	0.12	10.41	6.58	-0.04
8	8736	48	6.80	3.91	0.19	3.17	3.05	-0.49	7.47	6.18	0.29	11.19	6.82	0.56	5.38	3.52	0.35	11.80	7.22	0.60
8	12360	24	17.59	11.22	1.01	15.60	11.45	1.07	27.54	18.16	1.70	15.79	10.73	1.18	14.56	10.06	1.14	14.28	8.92	0.94
8	12360	48	2.46	2.05	-0.90	1.70	1.34	-0.60	2.38	1.79	-0.85	2.69	2.29	-0.07	1.92	2.79	-0.14	5.75	4.61	-0.34
12	8736	24	6.05	4.75	0.47	11.53	6.20	0.26	9.13	5.84	0.47	16.90	9.28	1.38	15.78	8.75	1.25	15.57	8.67	1.28
12	8736	48	6.19	3.75	-0.31	4.97	2.28	0.37	3.65	1.32	0.10	11.11	6.76	0.37	10.33	4.88	0.35	9.85	6.54	0.50
12	12360	24	16.43	12.23	2.30	16.27	10.68	1.58	16.47	10.36	1.86	24.40	15.27	1.96	15.13	8.46	1.70	24.88	15.62	2.33
12	12360	48	7.05	4.78	0.28	4.53	3.68	-0.08	5.80	4.41	0.03	8.92	5.26	0.19	7.12	5.95	0.27	5.16	3.34	0.20
24	8736	24	11.49	7.43	1.86	4.13	4.27	1.99	16.03	10.64	2.40	12.48	8.38	2.18	8.11	4.30	1.84	7.64	6.34	1.94
24	8736	48	6.91	4.51	0.98	-1.57	-0.54	-0.06	4.07	2.39	0.32	12.78	6.48	1.22	7.41	3.99	0.92	9.61	6.12	0.67
24	12360	24	6.61	6.06	-0.11	0.74	2.07	-0.06	2.02	2.75	-0.02	4.96	5.82	0.14	0.76	2.98	0.23	0.04	3.25	0.17
24	12360	48	5.24	4.40	-0.18	1.73	2.55	-0.34	3.27	2.30	-0.58	5.01	4.10	0.27	6.43	6.35	0.03	6.45	5.91	-0.07

Table C.9. Effect of using 12360 instead of 8376 hours for RF training across the performance metrics

Steps ahead	Lags	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
1	24	No	-9.83	-4.28	-0.02	10.10	4.83	-0.02	9.35	7.22	2.09	-9.79	-7.59	-1.55	-9.71	-7.71	-1.09	4.54	2.01	1.27
1	24	Yes	5.47	6.27	0.46	-3.63	-1.44	-0.48	6.67	4.09	-0.01	5.09	3.14	-0.21	11.94	6.62	0.05	8.94	5.48	0.00
1	48	No	-0.69	-2.50	-0.16	10.79	4.08	1.57	15.60	8.05	2.76	5.39	1.96	1.10	12.80	6.51	0.59	11.48	4.77	2.01
1	48	Yes	15.35	9.73	-0.01	9.71	6.48	0.13	7.52	5.14	-0.65	2.01	0.56	-0.55	8.39	3.65	-0.57	2.80	2.32	-0.49
2	24	No	-15.55	-10.97	-1.96	-7.56	-4.17	-0.31	-11.89	-7.47	-0.73	-12.05	-7.53	-1.56	-8.42	-4.48	-0.58	-13.98	-8.49	-1.95
2	24	Yes	12.32	10.06	0.72	16.65	12.39	0.80	17.89	12.96	1.03	16.19	11.80	1.20	10.37	9.09	0.38	14.99	11.64	0.30
2	48	No	-8.65	-5.66	0.47	0.19	-1.04	0.71	0.78	-0.22	1.00	3.71	2.58	1.43	-4.59	-3.98	0.21	-3.82	-2.48	0.50
2	48	Yes	6.00	4.74	-0.21	4.01	3.41	-0.14	3.80	3.24	-0.14	1.93	1.73	-0.58	8.41	5.52	-0.22	10.03	6.14	0.02
3	24	No	-1.15	0.29	0.11	-2.61	-2.50	-0.09	-4.30	-2.97	0.12	-0.57	0.49	0.22	-2.43	-1.88	0.20	4.76	3.38	0.30
3	24	Yes	4.51	2.07	-0.93	13.02	7.74	0.11	8.95	7.98	0.12	8.91	5.42	-0.43	2.89	-0.63	-1.31	4.19	2.35	-0.52
3	48	No	2.28	0.44	0.43	1.12	-0.31	-0.35	3.05	0.51	-0.19	3.72	2.01	0.82	2.66	-0.28	-0.22	-3.04	-3.94	0.25
3	48	Yes	10.84	4.95	0.09	10.32	3.82	-0.43	12.74	6.09	-0.16	8.10	3.90	-0.16	7.12	4.27	0.01	6.33	2.45	0.00
4	24	No	-2.91	-2.51	-1.62	3.35	1.46	-0.86	7.38	4.89	-0.49	3.90	1.66	-0.88	1.78	0.16	-1.39	0.99	-0.52	-1.38
4	24	Yes	-1.41	-2.05	-1.67	1.12	1.46	-0.68	0.17	-0.39	-0.85	-4.55	-2.56	-1.03	-2.17	-1.21	-0.71	-0.74	0.19	-0.87
4	48	No	7.70	4.40	0.74	2.45	0.05	-0.65	1.99	0.03	-0.76	2.34	1.19	-0.20	-0.07	-0.96	0.36	4.21	1.31	-0.21
4	48	Yes	-0.10	0.01	-0.89	0.70	0.64	-0.90	-1.25	-0.10	-0.91	-4.88	-2.85	-1.31	-4.06	-2.11	-1.15	-7.74	-5.25	-1.70
6	24	No	3.83	2.69	0.00	5.32	4.26	-0.30	7.83	6.47	0.05	-0.71	0.32	-0.57	3.43	2.99	-0.80	4.40	3.77	-0.84
6	24	Yes	9.95	6.37	-0.01	14.13	7.24	0.03	8.89	4.12	0.02	8.82	5.51	0.16	18.49	12.78	0.85	8.22	3.80	-0.37
6	48	No	9.47	4.63	0.89	7.08	2.21	-0.03	8.21	3.40	0.56	6.92	1.67	0.40	10.02	3.51	0.18	10.30	4.94	0.47
6	48	Yes	5.34	2.62	-0.42	5.23	3.24	-0.22	-1.97	-0.05	-0.56	-3.13	-1.58	-0.86	-0.75	0.64	-0.48	0.43	1.16	-0.58
8	24	No	2.95	1.63	-0.88	4.54	2.82	-1.09	0.85	0.37	-1.34	4.25	1.90	-1.06	4.66	2.78	-1.01	9.01	5.62	-1.15
8	24	Yes	16.16	10.18	-0.02	14.12	12.11	0.49	20.16	14.53	0.74	10.48	6.70	-0.49	11.69	8.59	0.01	12.88	7.95	-0.16
8	48	No	-0.90	-1.62	-0.07	-0.27	-0.09	-0.77	0.24	-0.41	-0.16	4.76	2.04	-0.47	-1.33	-1.97	-0.76	2.38	0.31	-0.33
8	48	Yes	-5.25	-3.48	-1.16	-1.74	-1.81	-0.88	-4.84	-4.80	-1.31	-3.74	-2.49	-1.11	-4.79	-2.70	-1.26	-3.67	-2.29	-1.27
12	24	No	6.24	3.21	-1.27	5.31	2.78	-1.38	6.87	3.74	-1.31	7.27	2.93	-0.87	9.49	5.83	-1.08	3.64	2.34	-1.31
12	24	Yes	16.62	10.70	0.56	10.05	7.26	-0.05	14.21	8.25	0.08	14.76	8.91	-0.30	8.84	5.55	-0.63	12.94	9.30	-0.26
12	48	No	3.78	1.16	-1.38	2.54	0.31	-0.19	0.49	-0.51	-0.42	7.65	5.18	-0.25	1.16	0.07	-0.75	4.27	2.62	-0.49
12	48	Yes	4.64	2.19	-0.79	2.11	1.71	-0.64	2.64	2.59	-0.49	5.45	3.68	-0.43	-2.06	1.14	-0.83	-0.42	-0.58	-0.79
24	24	No	11.04	6.11	1.31	8.17	5.57	1.27	15.02	9.20	1.43	12.23	7.53	1.61	11.95	6.12	1.05	8.58	5.06	0.79
24	24	Yes	6.16	4.73	-0.66	4.78	3.37	-0.78	1.01	1.31	-0.99	4.71	4.96	-0.43	4.60	4.79	-0.56	0.98	1.98	-0.98
24	48	No	2.49	1.34	0.54	4.31	2.24	0.09	3.63	1.90	0.35	4.80	1.99	-0.07	4.02	0.97	0.35	5.85	3.29	0.36
24	48	Yes	0.82	1.23	-0.62	7.61	5.33	-0.19	2.83	1.80	-0.55	-2.97	-0.39	-1.02	3.04	3.34	-0.54	2.69	3.08	-0.38
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table C.10. Summary of the effects of using 48 instead of 24 lags of PV power in RF

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-1.86	-5.97	111	58%
MAE	-1.06	-4.24	99	52%
MAPE	-0.19	-0.68	111	58%
Total	**-1.04**	**-3.61**	**321**	**56%**

Table C.11. Summary of the effects of using the maximum installed capacity scaling in RF

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	11.32	-3.25	7	4%
MAE	7.43	-2.28	5	3%
MAPE	0.26	-0.59	72	38%
Total	**6.34**	**-0.91**	**84**	**15%**

Table C.12. Summary of the effects of using 12360 instead of 8736 hours to train RF

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	4.17	-4.31	47	24%
MAE	2.45	-2.81	49	26%
MAPE	-0.25	-0.68	127	66%
Total	**2.12**	**-1.91**	**223**	**39%**

LightGBM

Table C.13. Effect of using 48 instead of 24 lags in LightGBM across the performance metrics

Steps ahead	Training size	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	No	-30.14	-19.84	-1.46	-43.50	-25.99	-2.81	-29.72	-19.09	-1.41	-12.99	-7.35	0.55	-3.34	-1.04	0.95	-3.34	-1.03	0.95
1	8736	Yes	11.24	8.08	1.33	11.03	8.37	1.41	21.21	15.32	2.06	-12.77	-7.99	-0.38	3.32	4.04	1.02	5.31	5.42	1.25
1	12360	No	31.06	23.81	2.88	26.51	19.00	2.59	30.96	23.77	2.88	28.89	21.12	2.50	29.34	22.20	2.21	22.64	17.26	0.87
1	12360	Yes	-18.04	-11.46	-0.41	-3.78	-1.64	-0.15	-18.05	-11.46	-0.41	-17.60	-11.04	-0.58	-2.90	-1.45	0.09	-4.53	-2.61	-0.02
2	8736	No	5.94	5.79	-0.07	12.95	6.76	0.27	5.93	5.79	-0.08	-1.52	-1.39	-1.29	9.92	3.99	0.10	9.82	3.91	0.09
2	8736	Yes	-17.05	-15.93	-1.20	-12.64	-8.74	-0.68	-16.93	-15.90	-1.19	-12.04	-9.96	-1.11	-15.17	-12.40	-1.29	-16.49	-13.11	-1.38
2	12360	No	-17.73	-12.17	-2.05	-0.02	-0.53	-1.15	-17.77	-12.10	-2.04	1.27	-0.14	-0.65	-3.13	-4.29	-1.64	-3.05	-4.16	-1.64
2	12360	Yes	-25.29	-18.23	-1.73	-16.00	-11.18	-1.42	-25.20	-17.94	-1.70	-17.68	-11.33	-1.32	-7.33	-5.89	-0.63	-8.60	-6.44	-0.67
3	8736	No	1.73	-1.79	-0.44	-1.59	-3.00	-0.17	3.72	-1.24	-0.48	1.35	-2.49	-0.30	-2.39	-3.79	-0.46	-2.56	-3.86	-0.47
3	8736	Yes	3.19	2.10	-0.58	-4.37	-2.36	-0.58	3.13	2.01	-0.63	-1.81	-1.71	-0.07	-3.32	-4.06	-0.65	-2.55	-3.63	-0.61
3	12360	No	0.83	-1.02	-0.37	-2.34	-3.06	-0.86	0.80	-1.07	-0.37	1.89	0.32	-0.08	1.57	0.45	-0.08	1.68	0.51	-0.07
3	12360	Yes	-13.38	-10.13	-0.40	-13.27	-9.26	-0.59	-13.40	-10.14	-0.40	-10.34	-9.79	-0.74	-8.19	-6.82	0.08	-10.30	-8.48	-0.11
4	8736	No	-8.86	-7.14	-0.24	2.61	1.26	1.11	-8.96	-7.25	-0.25	14.68	8.06	2.35	11.88	4.10	0.63	11.88	4.10	0.51
4	8736	Yes	7.65	3.77	-0.33	8.00	5.47	0.18	8.48	4.34	-0.30	12.56	7.53	1.08	-4.74	-1.09	-0.04	-5.84	-1.93	-0.11
4	12360	No	-1.10	-0.34	0.33	12.71	6.42	0.41	-1.24	-0.52	0.30	6.09	3.61	0.85	4.41	2.58	0.39	5.18	2.42	0.38
4	12360	Yes	10.87	5.84	0.60	15.47	10.57	1.40	10.78	5.86	0.61	1.24	-0.98	-0.06	7.21	3.75	0.62	7.24	3.82	0.62
6	8736	No	1.79	0.77	0.37	0.26	-2.39	-1.08	1.96	0.71	0.36	9.60	4.06	0.56	5.33	3.68	-0.01	5.39	3.70	-0.01
6	8736	Yes	18.37	11.46	0.73	10.36	7.30	1.04	17.18	10.89	0.69	11.06	6.56	0.51	7.56	4.44	0.31	7.40	4.38	0.30
6	12360	No	0.63	0.21	0.11	-2.89	-4.03	-1.06	0.36	0.06	0.01	-4.01	-2.63	-1.18	0.55	-0.37	-0.79	0.24	-0.50	-0.81
6	12360	Yes	0.76	0.76	-0.16	9.06	5.81	0.46	0.74	0.74	-0.17	1.38	-1.52	-0.65	-6.86	-4.59	-0.73	-6.83	-4.58	-0.74
8	8736	No	8.02	4.13	0.93	6.83	3.93	2.31	7.90	4.04	0.92	2.98	1.66	0.49	10.24	5.99	1.35	10.02	6.03	1.35
8	8736	Yes	-4.12	-1.70	-0.42	11.72	7.42	0.75	-4.10	-1.74	-0.43	-6.30	-6.12	-1.51	7.54	4.83	-0.11	7.53	4.85	-0.11
8	12360	No	8.31	3.45	0.10	7.42	2.82	0.81	9.29	3.88	0.12	6.64	3.09	0.74	8.56	3.60	1.63	8.51	3.62	1.63
8	12360	Yes	-4.48	-5.37	0.15	-2.00	-2.64	-0.02	-5.30	-5.95	0.10	-9.54	-8.14	-0.85	6.63	2.59	0.60	6.91	2.99	0.62
12	8736	No	1.25	1.15	2.96	4.57	1.22	2.11	2.90	2.15	2.89	17.41	11.14	3.55	2.60	2.38	2.44	2.54	2.41	2.44
12	8736	Yes	2.28	2.45	-0.04	4.49	6.95	0.64	2.65	2.68	-0.01	2.26	2.75	0.70	4.27	3.26	0.21	4.33	3.28	0.20
12	12360	No	2.11	1.38	0.77	3.15	1.09	1.20	2.10	1.45	0.75	-4.51	-2.43	0.74	-3.56	-2.83	0.74	-3.95	-2.96	0.78
12	12360	Yes	-4.81	-2.40	0.43	1.45	-1.01	0.61	-5.01	-2.53	0.42	-17.13	-12.04	-0.42	1.53	1.29	0.55	1.55	1.33	0.56
24	8736	No	-0.34	-0.79	1.25	8.89	3.79	2.04	-0.19	-0.70	1.25	3.21	1.10	1.29	5.19	3.05	1.32	4.23	2.54	1.26
24	8736	Yes	-11.29	-6.99	-0.32	0.43	-0.70	0.55	-11.08	-7.13	-0.34	-2.52	-1.93	0.00	-6.29	-2.90	-0.09	-7.03	-3.38	-0.14
24	12360	No	-0.86	-1.04	0.40	3.23	0.41	0.42	-0.60	-1.09	0.39	3.28	2.24	0.33	6.64	2.17	0.20	7.12	2.25	0.22
24	12360	Yes	-14.43	-8.14	-0.23	-12.21	-9.43	-0.14	-14.68	-8.26	-0.24	-11.35	-5.93	-0.03	-5.65	-4.67	0.23	-5.94	-4.67	0.23

Table C.14. Effect of using installed capacity scaling in LightGBM across the performance metrics

Steps ahead	Training size	Lags	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	24	-43.52	-29.61	-3.34	-55.21	-37.04	-4.32	-43.20	-28.92	-3.29	2.77	2.70	0.62	-16.24	-10.33	-1.84	-15.61	-9.81	-1.78
1	8736	48	-2.15	-1.70	-0.54	-0.68	-2.69	-0.10	7.74	5.49	0.18	2.99	2.06	-0.30	-9.57	-5.25	-1.77	-6.95	-3.36	-1.49
1	12360	24	26.61	17.93	-0.38	12.62	9.68	-0.59	26.59	17.94	-0.38	21.18	13.90	-0.55	11.85	9.52	-1.54	13.59	10.37	-1.46
1	12360	48	-22.48	-17.35	-3.68	-17.67	-10.96	-3.33	-22.41	-17.30	-3.67	-25.32	-18.26	-3.62	-20.39	-14.13	-3.66	-13.59	-9.49	-2.35
2	8736	24	18.91	16.28	1.53	28.22	19.76	1.83	18.94	16.31	1.53	37.42	26.91	2.69	36.58	26.51	3.21	36.46	26.44	3.21
2	8736	48	-4.08	-5.43	0.40	2.63	4.25	0.88	-3.91	-5.37	0.42	26.90	18.33	2.87	11.49	10.12	1.82	10.15	9.43	1.73
2	12360	24	20.63	14.52	0.28	11.72	9.02	-0.07	21.05	14.50	0.27	24.30	17.27	0.63	3.25	2.74	-1.23	4.66	3.42	-1.18
2	12360	48	13.07	8.46	0.60	-4.26	-1.64	-0.34	13.61	8.66	0.61	5.35	6.08	-0.04	-0.95	1.14	-0.22	-0.89	1.13	-0.21
3	8736	24	20.25	11.90	1.28	20.03	12.70	1.80	22.44	12.53	1.25	21.74	13.98	1.08	23.44	16.25	1.70	22.74	15.90	1.67
3	8736	48	21.71	15.78	1.14	17.25	13.34	1.39	21.85	15.78	1.10	18.58	14.76	1.31	22.51	15.97	1.51	22.75	16.13	1.52
3	12360	24	24.30	15.85	1.63	15.44	10.07	0.88	24.33	15.85	1.63	24.32	16.61	1.81	16.01	11.42	0.76	16.12	11.49	0.77
3	12360	48	10.10	6.74	1.59	4.51	3.87	1.15	10.13	6.78	1.59	12.09	6.50	1.15	6.26	4.15	0.92	4.14	2.49	0.73
4	8736	24	-2.30	-1.40	-0.25	20.33	12.55	1.66	-2.05	-1.26	-0.23	12.34	10.12	1.24	32.97	18.59	1.44	32.97	18.58	1.32
4	8736	48	14.21	9.51	-0.33	25.72	16.76	0.73	15.40	10.33	-0.27	10.22	9.60	-0.04	16.35	13.40	0.77	15.25	12.54	0.69
4	12360	24	16.49	12.10	1.12	15.18	8.72	0.50	16.47	12.07	1.12	27.00	18.89	1.29	19.40	13.15	0.90	20.19	12.91	0.89
4	12360	48	28.46	18.28	1.40	17.94	12.86	1.49	28.49	18.45	1.43	22.16	14.31	0.39	22.20	14.32	1.12	22.25	14.31	1.13
6	8736	24	11.29	8.75	0.85	22.58	13.96	0.49	11.22	8.65	0.85	21.36	14.34	1.57	29.70	20.16	1.81	29.76	20.18	1.81
6	8736	48	27.87	19.45	1.21	32.68	23.66	2.61	26.44	18.83	1.17	22.82	16.85	1.52	31.93	20.92	2.12	31.77	20.87	2.12
6	12360	24	12.64	7.61	0.89	15.56	8.14	0.29	12.62	7.64	0.90	17.03	10.67	0.74	26.48	15.23	0.85	26.45	15.18	0.85
6	12360	48	12.76	8.16	0.62	27.51	17.98	1.80	13.00	8.33	0.72	22.42	11.78	1.27	19.07	11.01	0.91	19.38	11.10	0.93
8	8736	24	26.40	15.44	1.57	18.82	11.36	0.59	26.35	15.46	1.57	28.72	20.55	2.38	25.08	15.39	1.61	25.04	15.48	1.62
8	8736	48	14.26	9.61	0.21	23.71	14.85	-0.98	14.34	9.69	0.22	19.45	12.77	0.38	22.38	14.23	0.15	22.54	14.31	0.15
8	12360	24	31.30	19.53	0.94	23.42	15.20	1.49	32.13	20.06	1.00	27.32	19.78	1.50	13.12	9.11	0.37	13.20	9.01	0.34
8	12360	48	18.51	10.72	0.99	14.00	9.74	0.66	17.54	10.24	0.99	11.15	8.55	-0.09	11.19	8.09	-0.66	11.61	8.37	-0.67
12	8736	24	18.68	9.59	1.29	22.55	9.13	0.92	18.84	9.80	1.31	28.70	15.09	1.74	27.46	14.72	1.91	27.73	14.89	1.93
12	8736	48	19.71	10.90	-1.71	22.47	14.86	-0.54	18.59	10.33	-1.60	13.55	6.70	-1.10	29.14	15.60	-0.32	29.52	15.76	-0.31
12	12360	24	18.43	10.35	0.55	7.54	4.88	0.41	18.40	10.33	0.53	23.50	14.55	1.28	14.63	8.56	1.15	14.24	8.26	1.13
12	12360	48	11.52	6.57	0.21	5.83	2.77	-0.18	11.29	6.35	0.19	10.88	4.93	0.12	19.72	12.69	0.97	19.74	12.56	0.90
24	8736	24	23.95	11.78	0.45	27.59	16.03	0.82	23.82	11.82	0.44	28.35	15.22	1.01	27.77	15.62	1.32	28.03	15.93	1.34
24	8736	48	13.01	5.58	-1.11	19.13	11.54	-0.66	12.93	5.39	-1.15	22.62	12.18	-0.28	16.29	9.67	-0.09	16.78	10.00	-0.07
24	12360	24	16.62	9.48	0.30	17.39	11.32	0.75	16.93	9.52	0.31	22.55	12.68	0.79	18.53	12.61	0.30	18.42	12.41	0.27
24	12360	48	3.06	2.38	-0.32	1.96	1.48	0.18	2.85	2.35	-0.31	7.92	4.51	0.43	6.23	5.77	0.33	5.35	5.49	0.28

Table C.15. Effect of using 12360 instead of 8376 hours for LightGBM training across the performance metrics

Steps ahead	Lags	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
1	24	No	-44.93	-30.20	-1.50	-59.48	-39.92	-3.07	-44.58	-29.52	-1.46	-22.78	-13.48	0.65	-24.15	-16.25	0.21	-24.19	-15.91	0.26
1	24	Yes	25.20	17.34	1.45	8.35	6.80	0.66	25.21	17.34	1.45	-4.37	-2.28	-0.52	3.95	3.60	0.52	5.01	4.27	0.58
1	48	No	16.26	13.45	2.84	10.54	5.07	2.33	16.10	13.34	2.83	19.10	15.00	2.60	8.54	6.99	1.48	1.79	2.37	0.17
1	48	Yes	-4.07	-2.19	-0.29	-6.46	-3.21	-0.90	-14.04	-9.44	-1.02	-9.20	-5.32	-0.72	-2.27	-1.90	-0.41	-4.84	-3.76	-0.69
2	24	No	1.47	3.98	1.37	7.55	6.15	1.61	1.55	3.99	1.37	-3.89	-2.09	0.47	13.54	10.69	2.69	13.28	10.50	2.68
2	24	Yes	3.19	2.22	0.12	-8.96	-4.59	-0.29	3.66	2.18	0.11	-17.01	-11.73	-1.58	-19.80	-13.07	-1.75	-18.52	-12.52	-1.71
2	48	No	-22.20	-13.98	-0.60	-5.42	-1.14	0.19	-22.14	-13.89	-0.60	-1.10	-0.85	1.12	0.49	2.42	0.95	0.41	2.44	0.95
2	48	Yes	-5.05	-0.08	-0.41	-12.32	-7.04	-1.03	-4.62	0.14	-0.41	-22.65	-13.10	-1.79	-11.95	-6.56	-1.09	-10.63	-5.86	-1.00
3	24	No	-6.01	-4.57	-0.66	2.31	2.07	0.41	-3.88	-3.96	-0.69	-0.38	-0.14	-0.10	3.24	1.99	0.20	3.24	2.00	0.20
3	24	Yes	-1.96	-0.62	-0.32	-2.28	-0.55	-0.51	-1.99	-0.64	-0.32	2.20	2.49	0.63	-4.18	-2.84	-0.74	-3.39	-2.41	-0.71
3	48	No	-6.91	-3.80	-0.59	1.56	2.01	-0.28	-6.80	-3.79	-0.58	0.16	2.67	0.12	7.20	6.22	0.58	7.47	6.37	0.59
3	48	Yes	-18.53	-12.85	-0.14	-11.18	-7.45	-0.52	-18.52	-12.79	-0.09	-6.34	-5.59	-0.04	-9.05	-5.59	-0.01	-11.14	-7.26	-0.20
4	24	No	-6.50	-4.63	-0.40	-2.70	-1.01	0.81	-6.65	-4.72	-0.41	-0.68	0.18	1.27	8.46	4.92	0.96	7.66	5.12	0.85
4	24	Yes	12.29	8.87	0.97	-7.85	-4.85	-0.35	11.87	8.61	0.95	13.98	8.94	1.32	-5.11	-0.51	0.42	-5.12	-0.54	0.42
4	48	No	1.26	2.17	0.17	7.41	4.15	0.10	1.08	2.01	0.15	-9.27	-4.27	-0.24	0.99	3.40	0.72	0.96	3.44	0.72
4	48	Yes	15.51	10.94	1.90	-0.38	0.25	0.87	14.17	10.13	1.85	2.66	0.44	0.19	6.84	4.32	1.08	7.96	5.21	1.16
6	24	No	-0.25	1.79	0.06	3.54	3.29	0.24	-0.20	1.73	0.06	9.56	6.75	1.29	10.62	8.74	1.28	10.62	8.74	1.28
6	24	Yes	1.10	0.65	0.10	-3.48	-2.53	0.04	1.20	0.72	0.11	5.22	3.08	0.46	7.41	3.81	0.33	7.30	3.74	0.32
6	48	No	-1.40	1.23	-0.20	0.39	1.65	0.27	-1.80	1.08	-0.30	-4.06	0.06	-0.45	5.84	4.69	0.50	5.46	4.55	0.48
6	48	Yes	-16.51	-10.05	-0.79	-4.78	-4.03	-0.54	-15.24	-9.42	-0.75	-4.46	-5.01	-0.70	-7.01	-5.23	-0.71	-6.92	-5.23	-0.71
8	24	No	-6.05	-4.40	0.14	-0.72	-0.27	-0.32	-5.97	-4.30	0.14	-1.95	-0.87	0.03	1.37	0.18	0.10	1.29	0.20	0.10
8	24	Yes	-1.15	-0.31	-0.49	3.89	3.57	0.58	-0.18	0.30	-0.42	-3.35	-1.63	-0.85	-10.59	-6.10	-1.14	-10.55	-6.28	-1.17
8	48	No	-5.76	-5.08	-0.69	-0.13	-1.38	-1.83	-4.58	-4.46	-0.66	1.70	0.57	0.28	-0.31	-2.21	0.38	-0.23	-2.21	0.38
8	48	Yes	-1.52	-3.98	0.09	-9.84	-6.49	-0.19	-1.38	-3.92	0.11	-6.59	-3.65	-0.19	-11.50	-8.35	-0.43	-11.17	-8.14	-0.44
12	24	No	-0.96	-0.86	0.56	2.65	1.44	0.03	-0.89	-0.80	0.58	7.84	5.36	0.71	2.57	2.57	0.55	2.79	2.75	0.57
12	24	Yes	-1.21	-0.10	-0.18	-12.36	-2.81	-0.48	-1.33	-0.27	-0.20	2.64	4.81	0.25	-10.26	-3.59	-0.21	-10.70	-3.88	-0.24
12	48	No	-0.11	-0.62	-1.64	1.23	1.31	-0.87	-1.69	-1.50	-1.56	-14.08	-8.21	-2.10	-3.59	-2.64	-1.15	-3.71	-2.62	-1.10
12	48	Yes	-8.30	-4.96	0.28	-15.41	-10.78	-0.51	-8.99	-5.48	0.23	-16.75	-9.99	-0.88	-13.01	-5.55	0.14	-13.49	-5.83	0.12
24	24	No	1.59	0.27	0.30	4.73	1.69	-0.07	1.47	0.20	0.27	0.39	-0.30	0.24	0.92	0.36	0.88	0.28	0.07	0.84
24	24	Yes	-5.74	-2.03	0.15	-5.47	-3.02	-0.15	-5.42	-2.09	0.14	-5.40	-2.84	0.02	-8.33	-2.65	-0.14	-9.33	-3.44	-0.23
24	48	No	1.07	0.03	-0.55	-0.94	-1.69	-1.69	1.06	-0.19	-0.59	0.46	0.84	-0.72	2.37	-0.52	-0.25	3.18	-0.22	-0.20
24	48	Yes	-8.89	-3.18	0.24	-18.10	-11.75	-0.84	-9.02	-3.23	0.24	-14.24	-6.83	-0.01	-7.69	-4.42	0.17	-8.24	-4.73	0.15
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table C.16. Summary of the effects of using 48 instead of 24 lags of PV power in LightGBM

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	0.22	-9.09	83	43%
MAE	-0.38	-5.80	96	50%
MAPE	0.19	-0.62	91	47%
Total	**0.01**	**-5.07**	**270**	**47%**

Table C.17. Summary of the effects of using the maximum installed capacity scaling in LightGBM

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	15.00	-15.16	22	11%
MAE	9.62	-11.57	20	10%
MAPE	0.47	-1.14	53	28%
Total	**8.36**	**-6.58**	**95**	**16%**

Table C.18. Summary of the effects of using 12360 instead of 8736 hours to train LightGBM

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-2.98	-8.72	116	60%
MAE	-1.42	-5.54	111	58%
MAPE	0.04	-0.68	90	47%
Total	**-1.45**	**-5.32**	**317**	**55%**

Stacking

Table C.19. Effect of using 48 instead of 24 lags in Stacking across the performance metrics

Steps ahead	Training size	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	No	-5.66	0.34	0.81	8.33	9.38	0.47	19.92	14.14	3.44	1.66	2.83	2.20	-1.94	-2.96	-0.01	10.82	6.36	0.60
1	8736	Yes	-17.91	-12.86	-1.91	-29.60	-20.35	-2.81	-41.40	-26.98	-3.11	-30.31	-20.79	-2.54	-7.62	-6.14	-1.03	-4.79	-4.64	-0.92
1	12360	No	5.38	0.01	3.01	3.41	-2.59	1.11	-10.91	-12.60	-0.88	13.39	4.20	4.18	24.14	11.66	2.34	18.81	12.59	4.07
1	12360	Yes	-80.68	-49.78	-5.67	-0.41	-1.20	0.51	12.63	8.46	1.62	-34.71	-22.82	-1.48	-13.18	-9.71	-0.88	7.59	6.52	0.33
2	8736	No	15.44	10.44	3.44	33.42	21.22	4.36	35.24	22.62	4.80	38.35	27.01	4.81	33.97	21.29	3.03	14.93	10.52	4.91
2	8736	Yes	22.84	14.83	1.00	-20.23	-15.08	-2.40	-19.88	-13.25	-2.12	-3.48	-2.19	-1.66	-8.17	-5.75	-1.58	19.49	13.03	0.33
2	12360	No	6.09	3.10	1.47	-9.74	-5.83	-0.81	-0.46	-0.42	0.04	1.71	0.00	1.14	-9.34	-7.56	0.50	-13.20	-9.49	-0.76
2	12360	Yes	-31.31	-20.46	-3.29	-16.05	-12.58	-1.71	-17.82	-13.90	-2.06	-35.93	-27.65	-3.73	-1.54	-2.78	-0.87	-24.34	-18.41	-2.45
3	8736	No	8.53	6.44	2.72	11.50	9.78	2.91	25.78	16.84	4.01	8.15	5.15	0.83	6.64	4.27	2.53	11.21	7.93	3.06
3	8736	Yes	8.80	6.57	0.93	8.07	5.80	0.27	-4.83	-2.18	-0.36	4.08	3.15	0.07	13.40	8.26	0.45	7.89	5.18	0.25
3	12360	No	-7.66	-5.21	-1.11	-6.88	-1.57	0.54	-3.44	-1.20	1.08	-5.12	-2.02	0.95	0.87	2.58	0.42	-4.13	-3.56	-1.07
3	12360	Yes	-27.62	-17.71	-1.21	-12.30	-8.97	-0.82	-14.15	-10.26	-1.19	-29.44	-19.38	-1.45	-10.21	-8.09	-0.51	-8.40	-6.32	-0.10
4	8736	No	-1.06	-0.04	1.00	-1.32	-1.91	0.94	6.02	4.30	1.50	6.26	3.76	1.33	1.03	0.75	0.43	2.04	-1.00	-0.42
4	8736	Yes	3.72	-0.70	-4.88	5.13	2.73	0.42	7.28	4.66	0.44	3.96	1.05	-0.11	8.50	6.24	1.05	19.85	12.20	1.42
4	12360	No	-1.29	-2.59	0.13	2.56	-2.12	0.50	-2.53	-2.97	-0.13	-7.62	-5.85	-0.67	-0.19	-0.44	0.61	-3.86	-2.51	0.62
4	12360	Yes	-15.50	-9.17	-0.11	-3.28	0.58	1.55	3.72	4.12	1.82	-12.77	-7.06	0.21	-13.40	-6.16	0.51	-8.43	-5.05	0.35
6	8736	No	-5.29	-4.32	-2.60	-2.51	-3.44	-4.35	-15.62	-12.71	-4.84	-5.96	-3.01	-1.32	4.46	0.58	-2.65	2.39	-0.54	-2.66
6	8736	Yes	0.70	-0.69	-3.46	-2.27	-1.60	0.29	-0.35	0.46	0.34	6.32	4.29	0.21	2.90	1.82	0.40	9.11	5.64	1.12
6	12360	No	-4.15	-2.82	1.16	-4.64	-2.86	1.14	-4.92	-2.02	1.77	2.63	1.56	1.43	0.79	-0.04	2.19	0.39	-0.20	1.02
6	12360	Yes	-6.41	-3.42	0.57	4.27	4.50	0.89	2.09	2.73	0.61	-3.20	-1.18	0.31	-2.66	-1.15	0.35	-4.10	-2.83	0.05
8	8736	No	-3.68	-5.82	-5.10	-10.20	-6.02	-1.93	-6.65	-3.27	-1.82	-1.44	0.15	0.12	-3.66	-1.95	-0.30	13.19	7.27	-0.81
8	8736	Yes	-8.61	-4.73	-3.00	-2.40	-1.99	-3.22	-7.54	-4.65	-3.37	0.57	0.52	-3.91	8.01	6.46	-2.73	5.76	4.51	-2.41
8	12360	No	-0.06	-0.03	-0.38	6.38	3.85	1.51	9.41	7.90	1.80	7.53	4.56	0.12	9.25	6.36	1.83	8.16	4.64	0.53
8	12360	Yes	-1.86	0.29	1.94	-3.21	-0.39	2.89	-4.83	-1.19	2.59	-3.60	-0.57	1.53	-0.27	1.42	2.01	-0.12	1.06	3.31
12	8736	No	7.80	4.11	1.67	14.51	8.34	-0.19	7.76	5.46	0.00	10.88	8.23	-0.38	12.00	7.34	0.14	4.96	2.23	0.65
12	8736	Yes	-1.45	-2.29	-3.48	-1.25	-2.67	-6.02	-2.26	-3.90	-5.50	-0.88	-2.34	-7.52	1.64	-0.77	-7.79	-4.25	-3.94	-6.94
12	12360	No	3.27	1.78	0.96	6.21	2.84	1.18	5.79	2.20	1.33	15.93	7.46	1.49	5.75	1.79	1.36	8.69	4.14	1.21
12	12360	Yes	6.73	3.49	0.00	7.69	3.75	-0.98	8.62	3.97	-0.76	7.29	2.73	-1.74	6.34	3.18	-0.10	6.25	2.60	-0.74
24	8736	No	8.40	5.53	0.77	12.07	7.63	1.12	17.21	11.01	1.36	13.63	8.34	-0.13	16.72	10.68	1.92	13.45	7.03	0.84
24	8736	Yes	5.65	2.67	0.53	1.90	0.30	-2.61	3.15	1.18	-1.53	5.72	2.83	-2.38	3.82	1.74	-2.65	3.76	0.91	-1.61
24	12360	No	-0.47	1.26	-0.37	6.28	4.92	2.05	8.47	5.77	1.29	2.53	2.47	-0.12	-0.70	0.98	1.48	-0.15	1.41	1.07
24	12360	Yes	-5.47	-4.41	-1.78	-1.30	-1.31	-0.49	-2.81	-2.58	-0.59	-2.26	-2.69	-0.87	-0.99	-1.11	0.08	-3.04	-1.75	-0.35

Table C.20. Effect of using installed capacity scaling in Stacking across the performance metrics

Steps ahead	Training size	Lags	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	8736	24	16.68	10.98	-3.69	-29.26	-17.99	-8.19	1.43	0.56	-5.18	-8.41	-4.70	-6.20	-37.20	-25.16	-7.65	-20.74	-13.72	-5.83
1	8736	48	4.43	-2.21	-6.41	-67.18	-47.73	-11.47	-59.89	-40.56	-11.73	-40.38	-28.31	-10.94	-42.87	-28.34	-8.66	-36.36	-24.73	-7.34
1	12360	24	110.71	67.09	2.98	-6.31	-6.09	-4.76	-8.70	-8.32	-5.12	34.62	20.55	-2.52	14.41	8.09	-2.05	-5.16	-7.57	-3.27
1	12360	48	24.65	17.31	-5.71	-10.13	-4.70	-5.37	14.84	12.74	-2.62	-13.48	-6.46	-8.18	-22.92	-13.28	-5.27	-16.38	-13.64	-7.01
2	8736	24	-21.94	-14.14	-2.40	-9.66	-7.82	-1.40	0.98	-1.80	-0.54	9.51	3.62	0.53	0.08	-1.02	-0.63	-10.43	-8.03	-1.10
2	8736	48	-14.53	-9.75	-4.84	-63.31	-44.12	-8.16	-54.14	-37.66	-7.45	-32.32	-25.58	-5.95	-42.06	-28.06	-5.24	-5.87	-5.52	-5.68
2	12360	24	67.88	45.27	5.69	8.05	6.03	-0.46	25.17	17.92	1.18	41.57	28.36	3.04	11.29	8.00	1.00	26.09	18.28	1.19
2	12360	48	30.48	21.71	0.93	1.75	-0.72	-1.35	7.82	4.44	-0.92	3.92	0.72	-1.82	19.09	12.77	-0.37	14.95	9.37	-0.50
3	8736	24	6.14	4.50	0.49	-12.91	-9.71	-0.28	-1.01	-2.50	0.34	6.35	2.72	0.65	-5.94	-5.43	0.08	6.18	2.91	0.74
3	8736	48	6.41	4.64	-1.30	-16.34	-13.69	-2.92	-31.62	-21.53	-4.03	2.28	0.72	-0.11	0.82	-1.44	-2.00	2.86	0.15	-2.06
3	12360	24	46.16	31.32	1.78	13.20	9.97	-0.22	17.14	10.81	0.54	32.40	21.96	1.74	25.38	18.34	0.16	19.42	12.70	-1.61
3	12360	48	26.21	18.82	1.67	7.79	2.57	-1.58	6.43	1.75	-1.73	8.08	4.60	-0.65	14.29	7.67	-0.77	15.15	9.94	-0.64
4	8736	24	22.10	18.15	6.06	-7.41	-5.33	-1.12	-0.80	0.83	0.14	15.47	12.20	1.54	4.81	3.10	-0.43	3.69	1.25	-0.66
4	8736	48	26.88	17.49	0.17	-0.96	-0.68	-1.64	0.46	1.19	-0.92	13.18	9.49	0.10	12.28	8.59	0.19	21.51	14.46	1.18
4	12360	24	28.56	19.50	2.52	19.44	11.52	0.76	8.20	6.12	0.02	16.91	11.10	0.58	14.99	10.20	0.32	13.46	9.77	0.60
4	12360	48	14.35	12.92	2.28	13.60	14.22	1.81	14.45	13.21	1.96	11.77	9.89	1.47	1.78	4.47	0.23	8.89	7.23	0.34
6	8736	24	2.96	3.42	1.30	5.14	3.08	-3.81	-1.77	-2.37	-4.05	4.39	4.51	-0.14	12.57	6.82	-2.34	6.88	3.04	-3.24
6	8736	48	8.94	7.05	0.45	5.38	4.92	0.83	13.50	10.80	1.12	16.66	11.80	1.39	11.01	8.06	0.71	13.61	9.23	0.54
6	12360	24	17.76	10.86	0.52	-1.77	-1.62	-0.44	-4.05	-1.74	-0.51	11.26	6.98	0.19	12.64	7.98	0.57	11.23	7.14	0.27
6	12360	48	15.49	10.26	-0.07	7.14	5.75	-0.70	2.97	3.00	-1.67	5.42	4.24	-0.93	9.19	6.87	-1.27	6.73	4.51	-0.70
8	8736	24	18.12	8.85	-0.90	5.02	4.48	2.86	11.99	8.95	2.51	16.14	12.54	4.45	10.86	9.45	3.94	19.14	12.58	3.71
8	8736	48	13.20	9.94	1.21	12.82	8.51	1.57	11.10	7.57	0.97	18.15	12.90	0.42	22.53	17.86	1.51	11.71	9.82	2.11
8	12360	24	5.38	4.55	0.64	6.73	3.65	-0.18	11.64	8.08	0.24	10.86	6.34	-0.36	6.65	3.31	0.34	3.60	2.16	-1.34
8	12360	48	3.58	4.88	2.96	-2.86	-0.59	1.20	-2.60	-1.01	1.03	-0.27	1.21	1.05	-2.87	-1.64	0.52	-4.68	-1.42	1.45
12	8736	24	20.46	12.39	4.85	26.77	16.17	6.62	19.04	13.31	5.51	31.13	23.37	8.72	30.91	20.24	9.79	31.45	20.59	9.53
12	8736	48	11.20	5.98	-0.30	11.01	5.17	0.78	9.01	3.95	0.00	19.37	12.80	1.58	20.55	12.14	1.87	22.23	14.42	1.94
12	12360	24	21.40	13.85	3.57	11.97	7.91	2.75	9.26	6.93	2.46	19.67	11.23	3.43	10.24	5.20	1.93	15.87	9.08	2.98
12	12360	48	24.87	15.56	2.61	13.44	8.82	0.59	12.09	8.70	0.38	11.03	6.49	0.20	10.83	6.59	0.47	13.43	7.53	1.02
24	8736	24	8.75	5.19	0.92	3.74	3.30	3.08	11.35	7.40	2.90	12.49	8.55	3.16	15.74	10.56	4.94	11.71	7.41	2.75
24	8736	48	6.00	2.33	0.68	-6.43	-4.03	-0.66	-2.70	-2.43	0.01	4.57	3.04	0.91	2.84	1.63	0.36	2.02	1.28	0.29
24	12360	24	0.84	3.64	2.08	-7.79	-1.28	2.50	-2.17	2.18	1.92	-2.94	1.22	2.26	-3.37	1.84	1.71	-5.45	0.58	1.97
24	12360	48	-4.16	-2.03	0.66	-15.37	-7.52	-0.04	-13.45	-6.17	0.03	-7.73	-3.95	1.51	-3.66	-0.25	0.31	-8.34	-2.58	0.55

Table C.21. Effect of using 12360 instead of 8376 hours for Stacking training across the performance metrics

Steps ahead	Lags	Scaling	All vars			Excluding Day of week			Excluding Quarter			Excluding Seasons			Excluding Seasons + Day of week			Excluding Seasons + Day of week + quarter		
			RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	24	No	-2.68	0.99	-0.57	-38.49	-22.32	-5.54	-13.74	-6.88	-2.42	-37.84	-22.55	-4.66	-40.65	-26.76	-6.09	-17.70	-9.69	-3.67
1	24	Yes	91.35	57.10	6.10	-15.54	-10.42	-2.12	-23.86	-15.76	-2.37	5.19	2.70	-0.98	10.96	6.49	-0.49	-2.11	-3.53	-1.11
1	48	No	8.37	0.66	1.64	-43.40	-34.29	-4.90	-44.57	-33.62	-6.74	-26.11	-21.18	-2.68	-14.56	-12.14	-3.73	-9.71	-3.46	-0.20
1	48	Yes	28.58	20.18	2.34	13.65	8.73	1.20	30.16	19.68	2.36	0.80	0.67	0.08	5.39	2.91	-0.34	10.27	7.63	0.14
2	24	No	-4.97	-1.84	-1.10	-12.04	-10.33	-1.73	-13.87	-11.26	-1.64	-12.50	-9.49	-1.88	-8.53	-6.49	-2.27	-5.69	-4.24	-0.76
2	24	Yes	84.85	57.57	6.99	5.67	3.52	-0.78	10.33	8.46	0.08	19.56	15.26	0.64	2.69	2.53	-0.63	30.83	22.06	1.53
2	48	No	-14.31	-9.19	-3.07	-55.20	-37.37	-6.90	-49.56	-34.29	-6.40	-49.13	-36.50	-5.55	-51.84	-35.34	-4.79	-33.83	-24.25	-6.43
2	48	Yes	30.70	22.28	2.70	9.86	6.02	-0.09	12.39	7.81	0.14	-12.89	-10.20	-1.43	9.31	5.49	0.08	-13.00	-9.37	-1.25
3	24	No	-0.39	-0.44	1.08	-17.44	-13.32	-0.90	-12.45	-8.83	-0.84	-14.81	-11.66	-1.53	-21.42	-16.55	-0.92	-8.31	-6.04	1.54
3	24	Yes	39.63	26.38	2.36	8.68	6.36	-0.84	5.71	4.48	-0.64	11.24	7.57	-0.45	9.90	7.23	-0.84	4.93	3.76	-0.82
3	48	No	-16.58	-12.09	-2.75	-35.83	-24.67	-3.26	-41.67	-26.87	-3.77	-28.08	-18.83	-1.41	-27.19	-18.24	-3.04	-23.64	-17.54	-2.59
3	48	Yes	3.22	2.10	0.22	-11.69	-8.41	-1.93	-3.62	-3.59	-1.47	-22.28	-14.96	-1.96	-13.71	-9.12	-1.80	-11.36	-7.74	-1.17
4	24	No	7.97	5.46	-0.39	-24.36	-16.23	-2.83	-11.24	-6.69	-1.40	-3.92	-2.42	-0.88	-18.93	-12.17	-2.30	-11.78	-9.54	-2.45
4	24	Yes	14.43	6.81	-3.93	2.49	0.62	-0.95	-2.24	-1.40	-1.52	-2.48	-3.52	-1.83	-8.74	-5.08	-1.54	-2.01	-1.02	-1.18
4	48	No	7.74	2.91	-1.26	-20.49	-16.43	-3.27	-19.79	-13.96	-3.02	-17.81	-12.03	-2.88	-20.15	-13.36	-2.12	-17.68	-11.05	-1.41
4	48	Yes	-4.78	-1.67	0.85	-5.93	-1.53	0.18	-5.80	-1.93	-0.14	-19.21	-11.63	-1.51	-30.65	-17.48	-2.08	-30.29	-18.28	-2.26
6	24	No	-3.84	-2.33	-3.50	-3.68	-1.34	-4.69	-6.40	-4.61	-4.79	-4.63	-0.28	-1.72	-4.71	-2.48	-4.25	-2.04	-1.43	-4.00
6	24	Yes	10.96	5.11	-4.28	-10.58	-6.04	-1.33	-8.68	-3.98	-1.25	2.24	2.18	-1.39	-4.64	-1.32	-1.34	2.31	2.66	-0.50
6	48	No	-2.70	-0.83	0.27	-5.81	-0.76	0.80	4.30	6.08	1.82	3.96	4.29	1.03	-8.37	-3.10	0.59	-4.03	-1.09	-0.32
6	48	Yes	3.85	2.38	-0.25	-4.05	0.06	-0.73	-6.24	-1.72	-0.97	-7.28	-3.28	-1.29	-10.20	-4.29	-1.40	-10.91	-5.81	-1.56
8	24	No	2.64	-0.66	-4.14	-8.93	-3.77	-3.06	-9.10	-5.16	-3.33	-7.26	-2.61	-1.16	-3.62	1.22	-1.76	-0.25	1.27	-1.63
8	24	Yes	-10.10	-4.95	-2.60	-7.22	-4.60	-6.10	-9.45	-6.02	-5.61	-12.55	-8.81	-5.96	-7.83	-4.92	-5.36	-15.80	-9.14	-6.68
8	48	No	6.27	5.13	0.59	7.65	6.11	0.38	6.96	6.02	0.29	1.70	1.80	-1.16	9.29	9.54	0.38	-5.29	-1.35	-0.30
8	48	Yes	-3.35	0.07	2.33	-8.03	-3.00	0.01	-6.74	-2.57	0.35	-16.71	-9.89	-0.53	-16.11	-9.96	-0.62	-21.67	-12.59	-0.96
12	24	No	1.83	0.23	-0.98	8.12	5.08	-2.03	3.13	2.60	-2.18	-2.39	2.80	-2.05	8.16	7.03	-0.87	-1.98	-0.45	-0.99
12	24	Yes	2.78	1.70	-2.26	-6.68	-3.18	-5.90	-6.65	-3.77	-5.23	-13.85	-9.34	-7.35	-12.52	-8.01	-8.73	-17.56	-11.97	-7.54
12	48	No	-2.71	-2.10	-1.69	-0.18	-0.41	-0.66	1.16	-0.66	-0.86	2.66	2.04	-0.18	1.91	1.48	0.36	1.75	1.46	-0.43
12	48	Yes	10.96	7.48	1.22	2.25	3.25	-0.86	4.23	4.09	-0.48	-5.68	-4.27	-1.57	-7.82	-4.06	-1.04	-7.05	-5.43	-1.35
24	24	No	12.47	6.20	1.94	10.77	5.82	-1.06	12.33	6.79	0.48	14.33	8.79	0.00	16.30	9.10	1.19	14.15	6.36	-0.44
24	24	Yes	4.56	4.65	3.09	-0.76	1.24	-1.63	-1.20	1.57	-0.51	-1.09	1.46	-0.90	-2.81	0.38	-2.04	-3.01	-0.47	-1.22
24	48	No	3.60	1.94	0.80	4.98	3.12	-0.13	3.58	1.56	0.41	3.23	2.92	0.00	-1.12	-0.60	0.74	0.55	0.73	-0.21
24	48	Yes	-6.57	-2.43	0.78	-3.96	-0.37	0.49	-7.16	-2.18	0.43	-9.06	-4.06	0.60	-7.61	-2.47	0.69	-9.81	-3.13	0.05

Table C.22. Summary of the effects of using 48 instead of 24 lags of PV power in Stacking

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	0.33	-9.07	93	48%
MAE	-0.11	-6.30	93	48%
MAPE	-0.11	-2.02	83	43%
Total	**0.04**	**-5.94**	**269**	**47%**

Table C.23. Summary of the effects of using the maximum installed capacity scaling in Stacking

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	5.76	-15.94	54	28%
MAE	3.88	-10.92	53	28%
MAPE	-0.23	-3.07	79	41%
Total	**3.14**	**-9.04**	**186**	**32%**

Table C.24. Summary of the effects of using 12360 instead of 8736 hours to train in Stacking

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-4.82	-13.40	125	65%
MAE	-2.89	-9.09	117	61%
MAPE	-1.40	-2.26	143	74%
Total	**-3.04**	**-7.95**	**385**	**67%**

Appendix D – Model results analysis (Multivariate Approach)

As a remark, three modeling scenarios with different feature sets were employed for hyperparameter tuning and model training. Scenario 1 contained the month, hour, and solar radiation variables, Scenario 2 included month, hour, solar radiation, and cloud cover, and lastly, Scenario 3 contained month, hour, solar radiation, cloud cover, and humidity.

XGBoost

Table D.1. Effect of using installed capacity scaling in XGBoost across the performance metrics

Steps ahead	Training size	Scenario 1			Scenario 2			Scenario 3		
1	8736	-13.80	-10.98	5.09	23.36	17.66	8.67	7.63	3.85	7.08
1	12360	-5.98	-3.70	3.71	9.81	10.89	4.55	8.62	7.03	4.22
2	8736	-8.11	-5.79	3.55	-2.09	-1.76	-1.94	1.30	-1.95	-0.66
2	12360	-6.88	-4.49	1.69	-4.61	-2.54	0.50	1.80	1.36	0.81
3	8736	-0.44	0.01	3.95	5.85	3.84	0.48	11.53	5.95	4.24
3	12360	13.92	7.69	4.84	18.86	9.97	1.60	13.45	5.22	1.38
4	8736	-3.33	0.79	2.98	4.87	2.83	3.55	13.37	8.00	2.82
4	12360	-3.68	-2.00	4.51	14.52	10.53	0.69	19.96	13.46	0.72
6	8736	-10.01	-3.46	3.47	-9.31	-6.82	-0.41	0.04	-5.09	0.62
6	12360	-1.76	0.74	7.28	-2.42	-3.08	2.25	20.37	12.61	2.10
8	8736	4.73	4.34	6.69	3.62	1.65	1.25	8.27	6.80	4.54
8	12360	-0.02	-0.31	3.92	4.98	3.38	3.30	14.61	11.63	2.80
12	8736	-8.14	-1.57	6.42	-7.54	-3.96	3.23	8.18	3.05	3.65
12	12360	-5.56	2.62	10.18	17.94	10.61	3.57	14.67	8.87	7.29
24	8736	-13.26	-7.33	3.74	-12.59	-11.26	-0.41	-12.89	-9.53	2.49
24	12360	-10.57	-3.40	8.39	-1.00	-4.57	1.48	-3.73	-4.44	3.26
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.2. Effect of using 12360 instead of 8376 hours in XGBoost across the performance metrics

Steps ahead	Scaling	Scenario 1			Scenario 2			Scenario 3		
1	No	-32.55	-23.38	-3.04	-8.56	-7.57	-1.69	-21.30	-15.66	-3.23
1	Yes	-24.74	-16.10	-4.41	-22.10	-14.33	-5.80	-20.31	-12.48	-6.09
2	No	-3.59	-2.92	-2.17	-7.24	-5.90	-5.28	-14.58	-12.57	-5.52
2	Yes	-2.36	-1.62	-4.02	-9.76	-6.68	-2.83	-14.08	-9.26	-4.05
3	No	-19.81	-14.10	-3.21	-21.40	-15.14	-4.73	-11.04	-6.80	-1.48
3	Yes	-5.45	-6.43	-2.33	-8.39	-9.01	-3.61	-9.11	-7.53	-4.33
4	No	-1.53	0.48	-2.93	-15.09	-14.16	-0.99	-10.42	-9.06	-2.35
4	Yes	-1.88	-2.30	-1.41	-5.43	-6.46	-3.85	-3.83	-3.60	-4.44
6	No	-10.50	-7.20	-2.36	-17.88	-11.55	-2.85	-33.49	-25.74	-3.52
6	Yes	-2.24	-3.00	1.45	-11.00	-7.81	-0.19	-13.15	-8.03	-2.04
8	No	-1.61	-0.37	-2.54	-9.98	-7.22	-8.03	-14.50	-10.00	-3.32
8	Yes	-6.37	-5.02	-5.31	-8.61	-5.49	-5.98	-8.16	-5.16	-5.07
12	No	-4.96	-7.24	-3.08	-30.74	-18.25	-3.70	-11.50	-7.30	-2.37
12	Yes	-2.38	-3.05	0.68	-5.27	-3.67	-3.35	-5.01	-1.48	1.26
24	No	-8.28	-6.19	-0.91	-14.45	-7.37	-3.54	-11.27	-4.86	-1.56
24	Yes	-5.59	-2.26	3.74	-2.86	-0.68	-1.65	-2.11	0.23	-0.80
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.3. Summary of the effects of using the maximum installed capacity scaling in XGBoost

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	2.47	-6.42	23	48%
MAE	1.61	-4.67	21	44%
MAPE	3.34	-0.85	4	8%
Total	**2.47**	**-5.19**	**48**	**33%**

Table D.4. Summary of the effects of using 12360 instead of 8736 hours to train XGBoost

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-11.18	-11.18	48	100%
MAE	-7.94	-8.30	46	96%
MAPE	-2.89	-3.32	44	92%
Total	**-7.34**	**-7.71**	**138**	**96%**

Random Forest

Table D.5. Effect of using installed capacity scaling in RF across the performance metrics

Steps ahead	Training size	Scenario 1			Scenario 2			Scenario 3		
1	8736	28.60	18.68	1.59	8.55	7.66	2.58	12.48	10.39	2.14
1	12360	6.99	9.76	0.28	4.57	5.98	0.99	-1.98	0.42	-0.38
2	8736	16.87	10.14	0.83	-14.51	-11.72	0.16	10.40	7.17	2.49
2	12360	6.55	3.25	-1.30	-7.07	-7.08	-1.12	-15.83	-9.20	-3.31
3	8736	14.79	9.57	1.36	-12.74	-11.55	0.53	3.09	2.01	1.08
3	12360	12.45	7.10	0.40	0.61	-0.24	0.36	4.44	2.75	-0.97
4	8736	5.47	3.79	-0.24	-13.82	-12.58	0.72	-0.23	1.02	0.50
4	12360	3.46	2.36	0.61	-9.29	-4.81	-0.30	-5.50	-2.79	-1.02
6	8736	8.98	4.25	0.87	0.18	-0.74	1.65	4.07	0.00	1.04
6	12360	-4.88	-2.87	-0.80	-9.49	-7.34	0.11	-5.15	-2.54	0.06
8	8736	5.41	4.70	0.55	0.40	0.20	2.15	-4.72	-2.00	0.04
8	12360	-3.49	-1.36	0.49	-17.60	-8.77	-0.49	-15.60	-6.14	-0.41
12	8736	6.49	4.74	1.09	-2.07	-4.38	0.89	-3.28	-3.37	1.24
12	12360	-5.23	-3.71	-0.68	-10.43	-13.46	-0.14	-1.10	-3.75	-0.03
24	8736	1.72	2.20	0.74	-5.19	-5.21	0.00	-4.45	-4.41	1.20
24	12360	-12.34	-4.81	0.60	-21.01	-14.86	0.47	-26.11	-18.29	-2.20
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.6. Effect of using 12360 instead of 8376 hours in RF across the performance metrics

Steps ahead	Scaling	Scenario 1			Scenario 2			Scenario 3		
1	No	21.16	9.34	2.30	-0.65	0.70	-0.64	6.24	4.31	0.40
1	Yes	-0.45	0.42	0.99	-4.63	-0.99	-2.23	-8.22	-5.67	-2.12
2	No	9.64	5.56	1.38	-12.37	-6.37	-0.85	20.74	14.87	3.07
2	Yes	-0.68	-1.33	-0.75	-4.93	-1.74	-2.13	-5.48	-1.50	-2.73
3	No	0.05	-0.59	-0.36	-14.24	-10.26	-1.29	1.85	0.21	0.55
3	Yes	-2.29	-3.05	-1.32	-0.89	1.05	-1.46	3.19	0.95	-1.50
4	No	-0.59	0.02	-1.55	-8.34	-7.84	-0.77	3.01	2.87	-0.13
4	Yes	-2.60	-1.41	-0.70	-3.82	-0.07	-1.80	-2.26	-0.94	-1.65
6	No	8.67	5.75	0.54	6.75	3.63	0.58	3.93	0.99	-0.22
6	Yes	-5.19	-1.37	-1.13	-2.92	-2.97	-0.97	-5.29	-1.56	-1.19
8	No	5.04	3.15	-0.80	12.84	6.72	0.99	8.81	4.93	-1.22
8	Yes	-3.86	-2.91	-0.86	-5.16	-2.24	-1.65	-2.07	0.79	-1.67
12	No	6.89	5.81	0.35	9.90	6.95	-0.24	0.22	-0.71	-0.09
12	Yes	-4.83	-2.65	-1.42	1.54	-2.13	-1.26	2.40	-1.09	-1.36
24	No	6.59	1.85	-0.73	6.26	3.76	-1.21	9.25	6.78	1.24
24	Yes	-7.47	-5.15	-0.87	-9.57	-5.88	-0.75	-12.41	-7.11	-2.17
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.7. Summary of the effects of using the maximum installed capacity scaling in RF

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-1.39	-8.97	26	54%
MAE	-1.04	-6.46	26	54%
MAPE	0.34	-0.89	15	31%
Total	**-0.69**	**-6.19**	**67**	**47%**

Table D.8. Summary of the effects of using 12360 instead of 8736 hours to train RF

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	0.49	-5.05	26	54%
MAE	0.29	-3.10	25	52%
MAPE	-0.65	-1.18	37	77%
Total	**0.04**	**-2.87**	**88**	**61%**

LigthGBM

Table D.9. Effect of using installed capacity scaling in LightGBM across the performance metrics

Steps ahead	Training size	Scenario 1			Scenario 2			Scenario 3		
1	8736	3.89	3.96	1.36	-17.94	-12.28	-0.54	-42.73	-29.11	-2.61
1	12360	-11.10	-4.72	-0.77	-12.15	-4.24	0.45	-42.27	-24.61	-2.69
2	8736	7.91	7.81	2.19	-25.84	-18.28	1.12	-7.36	-8.25	-0.32
2	12360	-11.53	-6.23	0.78	-17.98	-12.35	-0.26	-15.42	-9.20	-0.38
3	8736	15.65	12.01	2.03	5.47	1.39	1.87	-3.89	-3.91	1.33
3	12360	-5.89	-1.80	0.41	-8.70	-5.53	0.11	-3.75	-1.11	0.88
4	8736	-2.92	2.19	1.33	2.92	1.48	0.25	2.15	0.11	0.83
4	12360	15.88	11.50	1.94	-0.11	-0.83	0.90	6.00	1.89	1.17
6	8736	27.02	19.27	3.20	-6.31	-6.88	1.76	17.86	9.60	2.85
6	12360	4.70	2.44	0.61	-2.43	-1.12	-0.92	-9.41	-1.20	-2.29
8	8736	35.00	25.73	3.34	-13.52	-9.47	-0.13	2.48	1.60	1.61
8	12360	7.78	7.35	1.00	-1.29	-0.18	0.25	0.43	0.19	0.11
12	8736	19.11	13.07	2.60	5.94	0.25	1.87	7.62	1.22	1.81
12	12360	-8.95	-4.59	0.55	2.16	-2.47	-0.73	17.08	8.84	0.66
24	8736	22.63	15.25	2.00	3.21	-4.15	-0.22	-5.38	-5.39	0.16
24	12360	-12.72	-5.17	0.20	0.11	-1.50	-0.03	-16.87	-11.13	-2.00
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.10. Effect of using 12360 instead of 8376 hours for LightGBM across the performance metrics

Steps ahead	Scaling	Scenario 1			Scenario 2			Scenario 3		
1	No	-1.44	-1.40	0.63	-14.51	-11.71	-0.85	-0.62	-2.77	0.09
1	Yes	-16.43	-10.09	-1.50	-8.72	-3.67	0.15	-0.17	1.73	0.00
2	No	-0.71	0.75	-0.41	0.41	0.13	1.21	-2.95	-4.69	-0.89
2	Yes	-20.15	-13.30	-1.82	8.27	6.07	-0.17	-11.01	-5.64	-0.95
3	No	-1.87	-2.44	-0.64	5.06	3.81	1.81	0.14	0.56	0.84
3	Yes	-23.41	-16.25	-2.27	-9.11	-3.11	0.05	0.28	3.36	0.38
4	No	-21.16	-9.54	-0.81	-3.98	-3.00	-0.83	-2.93	-0.92	-0.30
4	Yes	-2.35	-0.24	-0.19	-7.02	-5.32	-0.17	0.92	0.86	0.04
6	No	3.49	5.57	0.28	-4.45	-3.77	0.40	-0.54	-3.61	1.61
6	Yes	-18.83	-11.26	-2.31	-0.57	1.99	-2.28	-27.81	-14.41	-3.52
8	No	-3.72	-0.97	0.39	-12.98	-7.95	-1.65	-0.60	-0.23	-0.36
8	Yes	-30.94	-19.35	-1.95	-0.75	1.34	-1.27	-2.65	-1.64	-1.85
12	No	-2.16	0.31	-0.90	-5.35	1.13	-0.76	-4.05	-5.61	0.28
12	Yes	-30.22	-17.35	-2.95	-9.13	-1.59	-3.36	5.41	2.01	-0.87
24	No	4.13	0.87	-0.29	-5.94	-3.80	-1.59	-5.42	-3.95	0.05
24	Yes	-31.22	-19.56	-2.09	-9.04	-1.16	-1.39	-16.91	-9.69	-2.11
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.11. Summary of the effects of using the maximum installed capacity scaling in LightGBM

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-1.53	-12.26	25	52%
MAE	-1.01	-7.25	27	56%
MAPE	0.62	-0.99	14	29%
Total	**-0.64**	**-7.82**	**66**	**46%**

Table D.12. Summary of the effects of using 12360 instead of 8736 hours to train LightGBM

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	-7.16	-9.53	39	81%
MAE	-3.95	-6.67	33	69%
MAPE	-0.73	-1.35	32	67%
Total	**-3.95**	**-6.11**	**104**	**72%**

Stacking

Table D.13. Effect of using installed capacity scaling in Stacking across the performance metrics

Steps ahead	Training size	Scenario 1			Scenario 2			Scenario 3		
1	8736	10.51	5.55	-2.34	-27.30	-21.91	-9.44	-17.06	-11.25	-6.86
1	12384	-9.71	-5.56	-4.56	33.86	23.56	1.07	-56.39	-41.54	-17.80
2	8736	49.48	32.36	0.27	13.47	8.68	-2.16	-14.22	-8.82	-4.37
2	12384	-29.48	-18.03	-4.70	-45.40	-30.02	-9.78	-2.68	-1.47	-3.04
3	8736	26.26	19.41	2.43	5.82	1.80	-3.45	7.47	3.26	-2.99
3	12384	24.57	14.31	2.12	14.03	6.33	-7.59	8.69	5.83	-2.20
4	8736	19.42	14.83	1.42	2.34	0.71	-2.34	1.59	2.98	-2.70
4	12384	4.53	5.00	0.75	2.42	1.35	-1.26	-0.60	2.80	-2.84
6	8736	6.25	4.16	-1.88	-4.16	-5.40	-4.77	4.62	3.65	-2.48
6	12384	25.66	15.07	0.87	-2.97	-1.42	-1.49	5.61	5.57	-5.00
8	8736	16.49	13.13	2.01	-7.12	-4.03	-2.41	8.25	3.45	-0.75
8	12384	-9.84	-5.99	0.12	-8.93	-3.00	-2.77	-10.81	-4.05	-2.94
12	8736	16.34	10.54	1.82	2.30	-0.39	-0.57	8.61	2.83	0.57
12	12384	10.41	5.29	3.44	-0.38	-1.32	0.48	7.02	5.65	0.06
24	8736	-1.74	0.19	0.63	-8.74	-9.27	-1.91	-24.56	-17.55	-2.97
24	12384	-15.95	-9.20	-1.47	-16.94	-10.99	-1.84	-20.20	-10.40	-0.88
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.14. Effect of using 12360 instead of 8376 hours for Stacking across the performance metrics

Steps ahead	Scaling	Scenario 1			Scenario 2			Scenario 3		
1	No	34.21	22.51	3.71	-25.93	-18.79	-7.06	54.51	40.33	10.70
1	Yes	13.99	11.39	1.49	35.23	26.68	3.45	15.17	10.04	-0.25
2	No	53.96	34.47	5.62	52.42	36.97	7.15	-6.03	-2.59	-3.06
2	Yes	-25.00	-15.92	0.65	-6.45	-1.72	-0.47	5.51	4.77	-1.73
3	No	4.60	5.68	0.68	-7.28	-1.27	3.90	4.35	2.11	-1.27
3	Yes	2.91	0.58	0.37	0.94	3.27	-0.23	5.57	4.69	-0.49
4	No	-9.15	-5.96	-1.18	1.37	-0.05	-1.56	8.00	6.19	0.71
4	Yes	-24.04	-15.79	-1.85	1.45	0.59	-0.48	5.81	6.01	0.56
6	No	-8.91	-3.56	-2.12	-1.13	-2.57	-4.10	-1.53	-0.61	-0.51
6	Yes	10.50	7.35	0.63	0.06	1.41	-0.83	-0.54	1.31	-3.02
8	No	13.73	9.10	0.30	-3.90	-3.45	-0.59	14.17	6.74	1.73
8	Yes	-12.60	-10.02	-1.59	-5.71	-2.43	-0.96	-4.90	-0.76	-0.46
12	No	2.61	3.12	-0.93	-1.30	0.07	-1.96	-4.56	-4.72	-0.22
12	Yes	-3.33	-2.13	0.69	-3.99	-0.86	-0.91	-6.15	-1.90	-0.73
24	No	9.27	5.29	1.16	5.26	1.85	-0.79	-13.25	-12.69	-3.33
24	Yes	-4.94	-4.09	-0.93	-2.94	0.12	-0.72	-8.89	-5.54	-1.24
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE

Table D.15. Summary of the effects of using the maximum installed capacity scaling in Stacking

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	2.92	-13.01	20	42%
MAE	1.93	-8.91	19	40%
MAPE	-1.97	-3.40	33	69%
Total	**0.96**	**-7.52**	**72**	**50%**

Table D.16. Summary of the effects of using 12360 instead of 8736 hours to train in Stacking

Metric	Average performance change (All configurations)	Average improvement (Only improved configurations)	Nr. of configurations improved	Percentage of configurations improved
RMSE	0.70	-7.86	31	65%
MAE	0.51	-5.14	31	65%
MAPE	-0.55	-1.80	32	67%
Total	**0.22**	**-4.90**	**94**	**65%**