Energy 336 (2025) 138396

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Can machine learning enhance day-ahead renewable power forecasts? A study on data-driven methods for solar photovoltaic and wind turbines

Francesco Superchi [a], Antonis Moustakis [b], George Pechlivanoglou [b], Alessandro Bianchini [a,*]

[a] Università Degli Studi di Firenze, Department of Industrial Engineering, Via di Santa Marta 3, 50139, Firenze, Italy
[b] Eunice Energy Group, 29, Vas. Sofias Ave, 10674, Athens, Greece

ABSTRACT

This study evaluates and compares various data-driven approaches to improve renewable energy production forecasts. In particular, focus is given here to photovoltaic (PV) fields and wind turbines (WTs), which commonly represent the backbone of modern Hybrid Power Stations (HPSs). With the aim of increasing the accuracy of commercial forecasting, the study investigates the potential improvement that machine learning (ML) techniques in correcting case-specific prediction errors. This improvement is particularly valuable for optimizing HPS scheduling and battery management, ultimately boosting system efficiency and reliability. With reference to a small Mediterranean island as the case study, for the PV Linear Regression (LR) achieved notable results, reducing the Normalized Mean Absolute Error (NMAE) from 8.6 % down to 3.59 %. Support Vector Regression (SVR) and Neural Networks (NNs) further enhanced accuracy, with NNs providing the best results (NMAE reduced to 3.34 %). In contrast, the improvement in wind power forecasts was more modest, with SVR performing best (NMAE only reduced from 10.25 % to 9.84 %). While ML successfully eliminated bias errors in the commercial forecast, they were less effective in addressing the variability and oscillations in wind production, highlighting inherent data limitations. Sensitivity analyses also revealed that shorter training windows improved algorithm performance, capturing recent patterns while minimizing noise from older, less relevant data. Results suggest that the introduction of ML correction methods into RES planning can support policy strategies aimed at enhancing renewable integration and improving the resilience of isolated power systems.

1. Introduction

Wind and solar energy, abundant and widely distributed across the globe, are essential for achieving the decarbonization goals critical to mitigating climate change [1–3], but their inherently intermittent nature results in power generation from renewable energy sources (RES) being variable and non-continuous.

Renewable generators, especially photovoltaic (PV) fields and wind turbines (WT), are being adopted in regions with outdated electrical grids that face challenges related to non-constant power injections [4,5]. To maintain the balance and stability of electricity distribution, grid operators impose constraints and operational limits on power producers to ensure compliance with technical and infrastructural boundaries [6–9]. In response to this request, renewable generators are paired with energy storage systems to enhance the flexibility and reliability of renewable energy output [10,11], forming Hybrid Power Stations (HPSs). By introducing advanced storage technologies and implementing optimized control and scheduling strategies, HPSs can mimic the operational characteristics of traditional dispatchable power generation units. However, achieving this level of functionality requires precise knowledge of the time evolution and magnitude of power generation from RES.

To implement robust and efficient strategies to manage the fluctuating output from intermittent resources [12,13], the ability to accurately predict the RES production potential plays a key role. Advanced forecasting methods are indeed essential for anticipating power output and optimizing dispatch scheduling for all types of generators. These forecasts enable grid operators and producers to mitigate the variability of renewable generators, improve reliability, and maximize the penetration of RES in the energy mix [14].

Current forecasting techniques for RES production primarily rely on meteorological data provided by the nearest weather stations or external companies specialized in meteorological data. Given that both solar and wind power generation are highly dependent on weather conditions, such forecasts are inherently subject to uncertainties. Additionally, the physical distance between weather stations and power generation sites can introduce further inaccuracies. Discrepancies in local environmental conditions that are not captured by distant weather stations may lead to bias errors in the forecasts, reducing the reliability of power output predictions. To address these challenges, more localized and precise forecasting methods are essential for improving the performance and

* Corresponding author.
E-mail address: alessandro.bianchini@unifi.it (A. Bianchini).

https://doi.org/10.1016/j.energy.2025.138396
Received 14 February 2025; Received in revised form 19 August 2025; Accepted 6 September 2025
Available online 11 September 2025

Nomenclature

AI	Artificial Intelligence
HPS	Hybrid Power Station
LR	Linear Regression
MAE	Mean Absolute Error
ML	Machine Learning
NE	Normalized Error
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Squared Error
NNs	Neural Networks
P	Power
PV	Photovoltaic
RMSE	Root Mean Squared Error
RES	Renewable Energy Sources
SVR	Support Vector Regression
WT	Wind Turbine

scheduling accuracy of renewable power plants.

1.1. Literature review of renewable power production forecasting

Data-driven approaches proved to be an interesting tool to forecast energy commodities such as renewable production [15] and hourly load profiles [16,17], and more.

Ruhanau et al. [18] demonstrated that the correlation between the forecast errors and the market price spread determines the economic implications, a phenomenon known as the 'correlation effect'. Chen et al. [19] examined how stochastic forecast errors (SFE) in renewable energy and load demands impact the stability of energy storage systems (ESS) by analyzing state of charge (SOC) deviations.

More specifically, several methods have been developed and tested, directly targeting the generation of reliable solar production forecasts. Accurate PV power forecasting is indeed essential for transmission and distribution system operators to optimize energy flow management [20].

Ma et al. [22] analyzed the sources of PV power forecasting errors, attributing them primarily to numerical weather predictions and forecasting processes. They proposed a short-term PV power forecasting method based on irradiance correction and error forecasting. Bright et al. [23] introduced a methodology to enhance satellite-derived PV power predictions by combining two data sources.

Data-driven approaches, particularly those incorporating artificial intelligence (AI), show significant prospects in improving renewable power forecasting. Sobri et al. [24] noted in their review that AI techniques outperform traditional methods due to their ability to handle the nonlinear and complex nature of the data. Sharifzadeh et al. [25] compared three predictive methods to forecast the behavior of renewable energy from wind and solar resources and electricity demand. Their findings demonstrated that all the models were capable of predicting wind and solar power.

Thaker and Höller [26] compared various methods for predicting photovoltaic (PV) power generation in fixed and single-axis tracking systems at two locations, focusing on intra-day forecasts (1–6 h ahead). Their results highlighted a highly skilled machine learning-based hybrid model as the most effective approach.

While simple Linear Regression (LR) can yield good results, more advanced models tend to consistently improve predictions. Markovics and Mayer [27] compared 24 machine learning models and found that the best model outperformed LR by 13.9 % in terms of Root Mean Squared Error (RMSE). They also emphasized that hyperparameter tuning is crucial to fully leverage the potential of these models. Das et al. [28] applied a Support Vector Regression (SVR) based model for PV power generation forecasting and found it to be more accurate than Neural Networks (NNs) in their application. Similarly, De Leone et al. [29] achieved accurate PV production forecasts using SVR. In contrast, AlShafeey and Csáki [30] compared regression techniques with NNs for 24-h-ahead PV forecasts, with NNs outperforming regressions regardless of the input method used. Çakır [31] proposed a hybrid methodology combining intuitionistic fuzzy time series (FTS) with intuitionistic fuzzy c-means clustering for renewable energy generation (REG) forecasting in Turkey. The model showed high accuracy and is applicable to other time series forecasting problems.

Lindberg et al. [32] evaluated the value of probabilistic forecasts for hybrid power parks (HPPs) using two years of data from Sweden. The authors found that probabilistic models enhance energy storage efficiency and profit under certain conditions, while deterministic models may still be preferable for specific battery-to-power ratios. Lee et al. [33] developed a hybrid deep-learning model combining a variational auto-encoder (VAE) and long short-term memory (LSTM) networks to forecast renewable electricity supply in South Korea. The model outperformed other machine learning approaches and led to significant error reductions.

Attempts have also been made to forecast the power production from wind turbines. In their review of wind, solar, and electrical load forecasting methods, Wang et al. [34] show that the uncertain forecasting results of wind power are higher than those of solar power. In their critical review of wind power forecasting methods, Hanifi et al. [35] provided guidelines for the selection of methods. The authors suggest advancing wind power forecasting through hybrid and computationally efficient methods, improving data processing for larger datasets, addressing offshore wind prediction challenges, and developing a standardized baseline model for benchmarking forecasting methods.

The literature focuses on short-term wind power production forecasting, and not many studies have been conducted on the day-ahead prediction of the generation. Yand et al. [36] proposed a method based on the clustering of equivalent power curves to perform a day-ahead wind power forecast. The method uses weather predictions as input, and tests show a promising MAE of 5.49 %. Zhang et al. [37] proposed a combined forecasting model using BP, Wavelet, and RVM with an information fusion strategy to improve the accuracy and uncertainty of short-term wind power forecasting. The results show that the model helps optimize EV charging schedules, maximizing wind power utilization, improving grid reliability, and reducing pollution.

Bashir et al. [38] proposed two hybrid models, CNN-ABiLSTM and CNN-Transformer-MLP, for improving solar and wind power forecasting. The authors demonstrated that CNN-Transformer-MLP excels in day and week-ahead predictions, while CNN-ABiLSTM is superior for month-ahead wind power forecasting, achieving lower MAE, RMSE, and MSE values.

More attempts have been made for the day-ahead forecasting of the wind speed. Nguyen et Phan [39] proposed a method to perform hourly day-ahead wind speed forecasting using a hybrid model of EEMD, CNN-Bi-LSTM embedded with GA optimization. Their approach confirmed that hybrid models usually perform better. Hong and Satriani [40] proposed a day-ahead spatiotemporal wind speed forecasting using convolutional neural networks. The authors determined the parameters of the network by robust design and proved that CNN outperforms existing methods. NNs were proven to be promising for wind power forecasting also by Lin et al. [41].

Pappa et al. [42] applied the Analog Ensemble (AnEn) technique to improve wind speed and solar radiation forecasts for renewable energy predictions across Europe. AnEn enhances forecasts, outperforming multi-model ensemble forecasting in most cases.

1.2. Research gaps and contributions

The literature review revealed that there are many studies proposing different ML methods for the direct forecasting of power production from intermittent renewable sources or other time series related to the operation of energy systems. On the other hand, many managers of microgrids, HESs, or power plants are provided with commercial forecasts of the generation capabilities of the devices that power their systems. These forecasts can be characterized by poor accuracy and strong biases. However, the topic of enhancing forecasts of energy commodities has been scarcely explored in the literature.

Maciejowska et al. [43] proved that the utilization of day-ahead wind production enhanced forecasts is helpful in the day-ahead choice of a market. The authors analyze the German electricity market, focusing on forecasting fundamental variables (load, wind, and solar generation) and their impact on electricity prices and utility decision-making. They employ ARX models to improve load forecasts, while only slightly enhancing wind and solar predictions. They integrate these improved forecasts into utility decision-making, demonstrating a 13.5 % increase in annual revenue, confirming the findings of [44].

One of the most common problems of commercial forecasts is that providers often offer generic predictions that may not account for site-specific conditions such as variable shading and conversion losses. Böök and Lindfors [45] proposed a method for adjusting a generic PV output forecast to specific sites, successfully reducing the MAE from 8 % to 6 %. Unlike the previously cited studies, the approach proposed in this work is most similar to Böök and Lindfors' methodology but offers a more direct solution.

Yin et al. [21] examined the error distribution in short-term PV power predictions and proposed a correction method based on NNs that successfully reduced the Mean Absolute Error (MAE) by approximately 2.6 %. The present study advances the status of research in the field by comparing the effectiveness of NNs with regression methods and by applying the same correction mechanisms also to wind forecasts.

This study aims to contribute to the field of ML correction of commercial forecasts by applying machine learning (ML) techniques specifically to enhance solar and wind power production forecasts. ML techniques can use historical data of both prediction and actual production and then be applied to identify patterns and correct biases in the initial forecasts. By directly addressing the inherent uncertainties and errors associated with commercial predictions, this study aims to enhance the reliability of PV and wind power output estimations, contributing to a more effective and efficient management of renewable energy resources.

The primary objective is to develop and evaluate a robust, data-driven methodology for directly improving the accuracy of power generation forecasts of a PV field and a WT. To achieve this, three ML modelling approaches were employed and compared: Linear Regression (LR), Support Vector Regression (SVR), and Neural Networks (NNs).

The renewable energy generators analyzed in this study are part of the HPS located on the Greek island of Tilos. This HPS integrates solar and wind power generation with an advanced energy storage system. Each day, the station is required to develop an output schedule based on renewable energy forecasts to optimize the operation of the storage device (a high-temperature battery). This battery plays a crucial role by converting the inherently intermittent renewable energy production into dispatchable power, which can be reliably supplied to the grid operator [46].

The methods proposed in this study could also be applied to the scheduling and operation of more complex hybrid energy systems, as suggested in prior research by the same authors [47]. In such systems, enhanced forecasts could inform the activation and coordination of various system components, such as energy storage units, auxiliary generators, or load management systems. Improved forecasting directly enhances the efficiency and reliability of these systems by enabling better storage utilization and energy dispatch planning. Accurate predictions of renewable energy output are key for maximizing the performance of HPSs, as forecast accuracy has a direct impact on the effectiveness of energy storage scheduling. By improving forecast precision, these methods enable hybrid power stations to increase the utilization of renewable resources, reduce waste, and optimize economic returns through more effective energy trading with grid operators.

A data-driven methodology is particularly well-suited to enhance the accuracy in predicting renewable power production because it enables the correlation of generic forecasts with the actual operational performance of a specific plant. Often, the real production of RES diverges from predictions due to a variety of factors, including the degradation of system components, the unique characteristics of the geographic location of the plant, site-specific environmental conditions, and numerous other variables affecting device performance. These discrepancies, which are challenging to account for using traditional forecast models, can be systematically addressed through ML algorithms. By learning from historical production data, data-driven models have the potential to identify and capture the nuanced effects of these factors. A well-trained model can post-process generic predictions to incorporate the influence of these phenomena, ultimately producing more accurate and reliable forecasts tailored to the specific characteristics of the HPS.

A preliminary version of the proposed methodology was first presented in Ref. [48] and tested on the PV production forecast only. Unlike traditional approaches, the methods presented in this work focus on refining commercial forecasting outputs by leveraging the production history of an actual HPS. Having access to the historical behavior of a real HPS, the present study uses the historical time series from an actual PV field and WT. By testing the effectiveness of ML-based models in a real-world scenario, this research underscores the practical applicability of data-driven approaches for enhancing RES predictions.

The paper is structured as follows. Section 2 introduces the machine learning (ML) techniques used in this study, detailing the specific algorithms applied to enhance renewable energy production forecasts. This section also explains the rationale behind selecting these methods and outlines the workflow used to simulate their application, aiming at mimicking the actual operation for a day-by-day correction. The error metrics used to evaluate the effectiveness of the proposed methodologies in reducing forecasting inaccuracies are then presented. These metrics are critical for quantifying the improvement achieved by each approach and for enabling a fair comparison between the methods.

Section 3 presents the results of comprehensive sensitivity analyses, which examine how variations in input parameters and model configurations affect forecasting accuracy. This section also identifies the most effective ML technique for minimizing errors in solar and wind power production forecasts. The findings highlight the relative strengths and weaknesses of each method and point to the one that shows the most interesting prospects for practical applications.

Finally, Section 4 provides a detailed summary of the main findings of the study, highlighting the key outcomes and insights gained from the analysis.

2. Methods

Prediction errors, and bias errors in particular, can be reduced following a procedure based on data-driven methods. An ML algorithm can be trained using the difference between the forecast and actual production over a certain time frame.

The proposed method utilizes a moving window approach to train the ML algorithm on several days of operation, which is then applied to enhance the forecast for the following day. This approach mimics the actual operation of the algorithm, allowing it to improve forecasts for the next day before executing the algorithm that optimizes the energy scheduling of the HPS.

Three machine learning methods were tested and compared to determine the best fit for this application. The first method, LR, was selected for its simplicity and low computational effort, as it can yield good results given the repetitive nature of PV production. The second method, SVR, is a more complex model with tunable parameters that may improve performance for this specific task. The third method, NNs, represents the most advanced algorithm tested in this study, capable of capturing non-linear relationships within the training time series and potentially delivering better results.

All methods were implemented through functions from the *scikit-learn* library [49]. Those algorithms were optimized by varying the number of days that the algorithm can use for the training phase, and, for SVR and NNs, some hyper-parameters were tuned to tailor the algorithms to the specific application.

2.1. Training and application of ML algorithms

During the training phase (Fig. 1), the algorithm receives as input the historical production data from a generator (either PV or WT) along with the historical production forecasts provided by the commercial provider. The result of this phase is a tuned model that establishes a relationship between the initial forecast from the provider and the actual production.

In the case of PV forecast, the algorithm also exploits the autocorrelation of solar production as supplementary information to enhance the forecast correction. The time series is shifted by 24 h, and during the forecast adjustment phase (Fig. 2), the algorithm also takes into account the PV production from the previous day. In the case of the WT forecast, the time series is not lagged, since no correlation subsists between the production at a given moment and the production at the same time the day before.

During the application phase (Fig. 2), the algorithm takes as an input the RES production forecast from the commercial provider for the upcoming day of operation. Using this information, the algorithm processes and adjusts the forecast to predict the actual production for the following day. The ideal output is an accurate estimation of the RES production for the next day, reflecting a correction based on both historical data and the forecast provided. When correcting the PV forecast, the algorithm also uses the actual production data from the 24 h preceding the forecast application.

2.2. Moving window approach and walk-forward validation

This analysis simulates an approach to enhance next-day renewable forecasting over a one-year period. The algorithm is initially trained using data from the first N days of operation and then applied to improve the forecast for the first "application day." Subsequently, the training window moves forward, incorporating data from the next N days to refine the forecast for the second "application day," and this process continues iteratively throughout the year. Fig. 3 presents a scheme of the proposed moving window approach. The results are presented as the average error that would characterize a forecast improved by a specific method when applied throughout the entire year of operation. In this way, the algorithms are validated following a “not anchored walk-forward validation” technique, showing how well algorithms can generalize beyond the data they are trained on for the first time. Unlike a conventional cross-validation, this strategy inherently prevents over-fitting in forecasting of time series by ensuring that future data are never used in the training phase [50].

To accurately capture the factors contributing to the mismatch between the commercial forecast and the actual production from RES generators, it is key to train the algorithm using data from a time window close to the application day. This approach ensures that all relevant phenomena influencing the real operation are taken into account. Conversely, training the algorithm with data from distant time periods may introduce bias errors into the predictions, as it may not reflect current operational conditions.

The results section (Section 3) presents several sensitivity analyses aimed at optimizing the method for this particular application. The training window size is adjusted to determine the optimal number of previous days required to train the algorithm effectively. Additionally, ML models have been fine-tuned to meet the specific needs of this problem.

2.3. RES historical production data and power production forecasts

The HPS of Tilos integrates both wind and solar energy production. The Energy Management System (EMS) of the system receives forecasts from a commercial provider, along with real-time performance data from each generator. Both the predicted and actual production data are stored in datasets, which are available for analysis.

The PV field installed in Tilos is characterized by a peak power of 160 kWp and is composed of 592 solar panels of 270 Wp each, mounted with a tilt angle of 30° and oriented towards the south.

The wind turbine has a nominal power production of 800 kW and is installed in the northwest region of Tilos. The WT is a commercial Enercon E−53 model and features an independent pitch system for its rotor blades, enabling active control of the pitch angle. The turbine begins generating power at a cut-in wind speed of 2 m/s, reaches its nominal output at a rated speed of 13 m/s, and has a safety cut-out speed of 28 m/s. Additionally, the E−53 is equipped with a storm control mode, which allows the turbine to gradually slow down when wind speeds range between 28 and 34 m/s.

The time series of historical production of the PV field and WT have

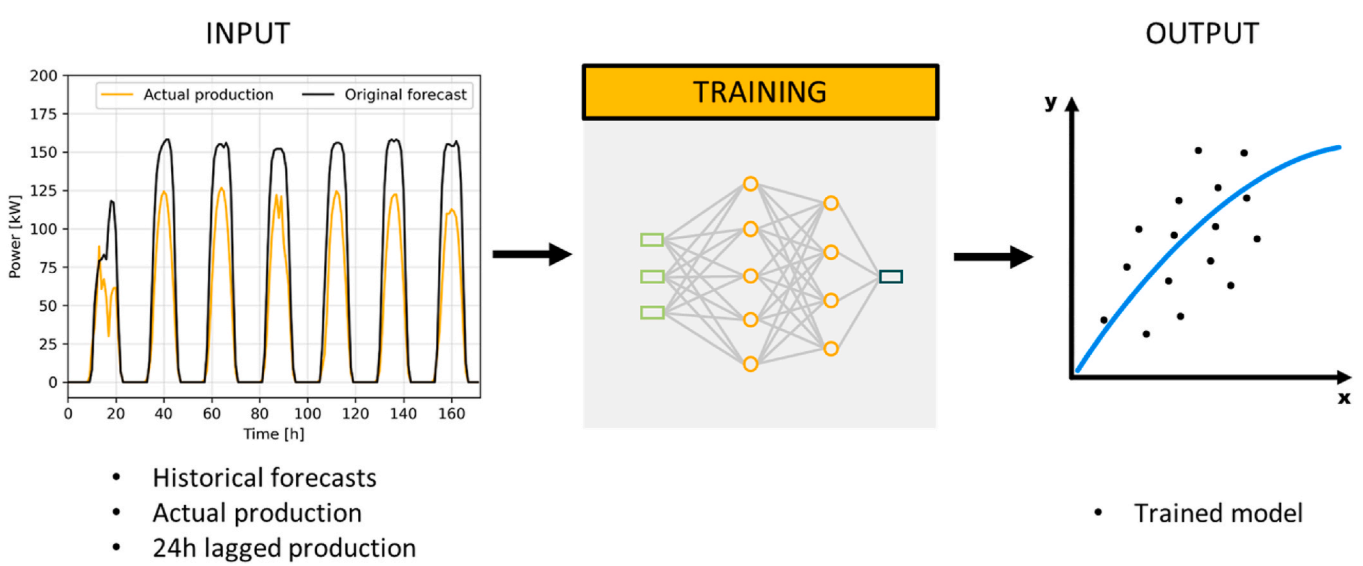

Fig. 1. Training Phase (e.g., considering PV forecast enhancement).

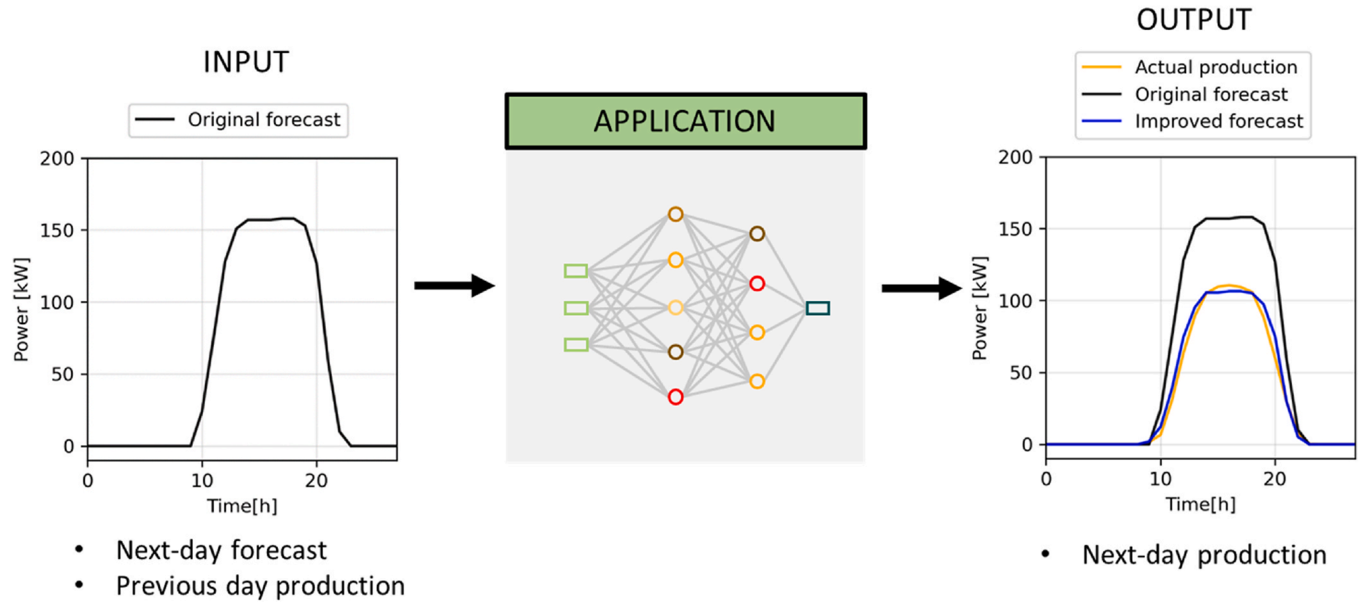

Fig. 2. Application Phase (e.g. considering PV forecast enhancement).

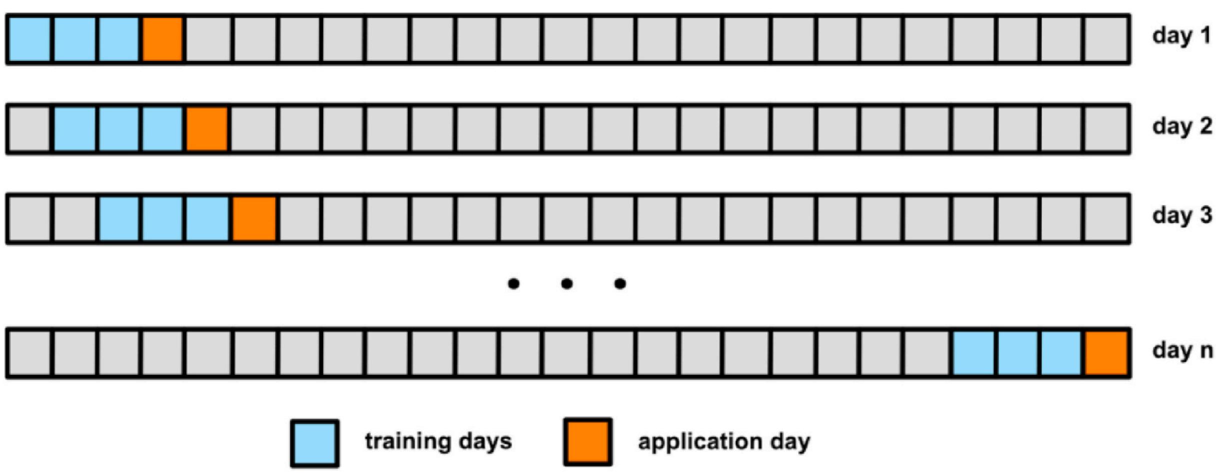

Fig. 3. Moving window approach to test the actual performance of the algorithm.

been kindly provided by Eunice Energy Group (EEG), owner and manager of the HPS of Tilos. The production dataset has a 1-h time resolution, recording the average power output from the generators for each time step. The historical forecasts, provided by a commercial provider, also have an hourly resolution and span the same period as the historical production dataset.

The training window runs through thirteen months of operation of an actual HPS, from September 1st, 2022, to October 1st, 2023. The application day is instead selected on a year of operation, from October 1st, 2022, to October 1st, 2023.

First, the historical datasets containing the day-ahead forecasts and actual production data were time-aligned. The historical forecasts were directly employed for the training and application of ML algorithms to demonstrate the correction margin that these methods could have when applied to commercial forecasts as they are provided to users.

A data pre-processing and quality control phase was instead applied to the historical dataset of actual production from PV and wind generators to ensure data reliability. The time series were first analyzed to replace missing values using techniques as linear interpolation to fill short gaps. Then, power production consistency was ensured by comparing historical measures with nominal capabilities of generators according to their specifics reported in datasheets. Historical records of wind speeds were used with the E−53 power curve to ensure that wind power measures do not exceed the maximum output capabilities of the machine. Similarly, historical records of the solar radiation were used with the STC and NOCT efficiencies of solar panels to verify the plausibility of solar generation records.

Finally, the cleaned and time-aligned datasets were used to study the application of the ML correction algorithms. To ensure numerical stability during the training phase and to avoid issues due to the different magnitudes of features, all input variables were normalized using the min–max scaling function.

2.4. Machine learning methods and hyperparameters optimization

The methods presented here can be fine-tuned by adjusting the parameters of the algorithms and the time series data used during the training phase. The parameters governing the learning process of ML algorithms are known as "hyperparameters." These hyperparameters are specific to each algorithm and must be optimized for the particular application at hand. Additionally, to determine the optimal training window, the number of preceding days used for training was varied ranging from three to thirty days prior to the "application day". The algorithms proposed in this work (LR, SVR, and NNs) are described in more detail in Ref. [51].

2.4.1. Linear regression (LR)

LR is a linear model that correlates a predictor (commercial forecast) and an outcome (enhanced forecast). The training phase tunes the coefficients of the model to minimize the distance between the model's outcome and the target (historical production). Fig. 4 represents a simple linear regression applied between an input x and an output y. The LR method was implemented through the *LinearRegression* function from the *scikit-learn* library [49], and it was optimized by varying the number of days that the algorithm can use for the training phase.

2.4.2. Support Vector Regression (SVR)

The SVR, in principle, works as the LR, with the possibility of defining a degree of acceptance for the error to tailor the algorithm to a specific task [43]. Errors that fall inside a tolerance band (within a distance ε from the regression) are not penalized during the training phase, unlike predicted points at a distance ζ from the tolerance bands (Fig. 5).

In addition to the optimization of the number of training days, two hyperparameters of the SVR were tuned.

- ε: the maximum error that the model allows during the training phase. The initial trial value of ε was set to 0.0 and then varied up to 1.0.
- C: the hyperparameter that controls the tolerance for errors exceeding ε. A higher value of C increases the tolerance for data points that fall outside the maximum error ε. When C is set to zero, no tolerance is allowed for values outside this range. The initial trial value of C was 1, and then varied within a range from 1 to 50.

SVR was implemented through the *LinearSVR* function from the *scikit-learn* library.

2.4.3. Neural networks (NNs)

The present study also considers the application of NNs to determine whether a more complex approach could yield improved results, potentially capturing non-linear relationships within the data time series. Fig. 6 represents a fully-connected NNs topology and zooms in on the structure of a single neuron. During training, the weights and biases of neurons are adjusted at each iteration to minimize the error between the predicted output and the target [52].

NNs were implemented using the MLPRegressor function from the scikit-learn library, leveraging the "adam" solver, a stochastic gradient-based optimization algorithm. To control the step size during weight updates, a learning rate of 0.0001 was chosen, ensuring stable convergence during training.

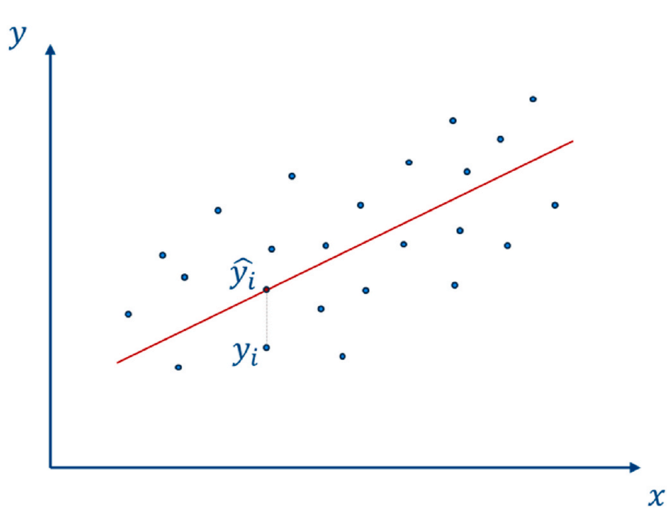

Fig. 4. Linear Regression representation.

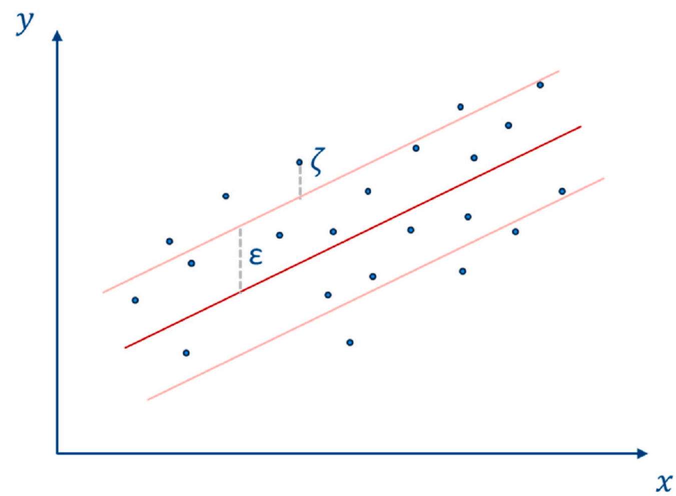

Fig. 5. Support Vector Regression representation.

A sensitivity analysis was carried out to assess the impact that the number of training days may have on the performance of the predictive performance of the model. Moreover, focus was given on fine-tuning the network topology, a crucial hyperparameter that defines the structure of the neural network. This involved testing different numbers of hidden layers between the input and the output of the network, and different numbers of neurons within each layer. Adjusting the network topology may help reach a better balance between model complexity and generalization, enhancing the ability of the network to capture intricate patterns within the time series data while avoiding overfitting.

2.5. Error metrics

2.5.1. Normalized Mean Absolute Error (NMAE)

The error that characterizes the forecast when trying to predict the power production of a generator in a specific moment can be normalized and expressed as a Normalized Error (NE). NE is indeed computed by dividing the difference between the forecasted power ($P_{forecast}$) and the actual power production (P_{actual}) by the rated power (P_{rated}):

$$NE = \frac{P_{forecast} - P_{actual}}{P_{rated}} \cdot 100 \tag{1}$$

To evaluate the overall accuracy of the forecast, the present work employs the Normalized Mean Absolute Error (NMAE) metric. The NMAE is calculated as the arithmetic mean of the absolute values of the NE for each timestep (n) considered in the analysis. Since the study analyses the performance throughout one year of operation, n corresponds to the total number of hours in a year (i.e., 8760).

$$NMAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|P_{forecast} - P_{actual}|}{P_{rated}} \cdot 100 \tag{2}$$

The NMAE aims at providing a comprehensive measure of forecast performance over the entire evaluation period.

For the PV field, the forecasting error is normalized by the nominal production capacity of 160 kWp. The original solar power forecast provided by the commercial provider has a NMAE of 8.6 % for the considered year. This forecast exhibits a significant bias error, with predictions often exceeding the actual production.

For the WT, the error is normalized by the nominal production capacity of 800 kW. The original wind power forecast from the commercial provider shows an NMAE of 10.25 % for the same year.

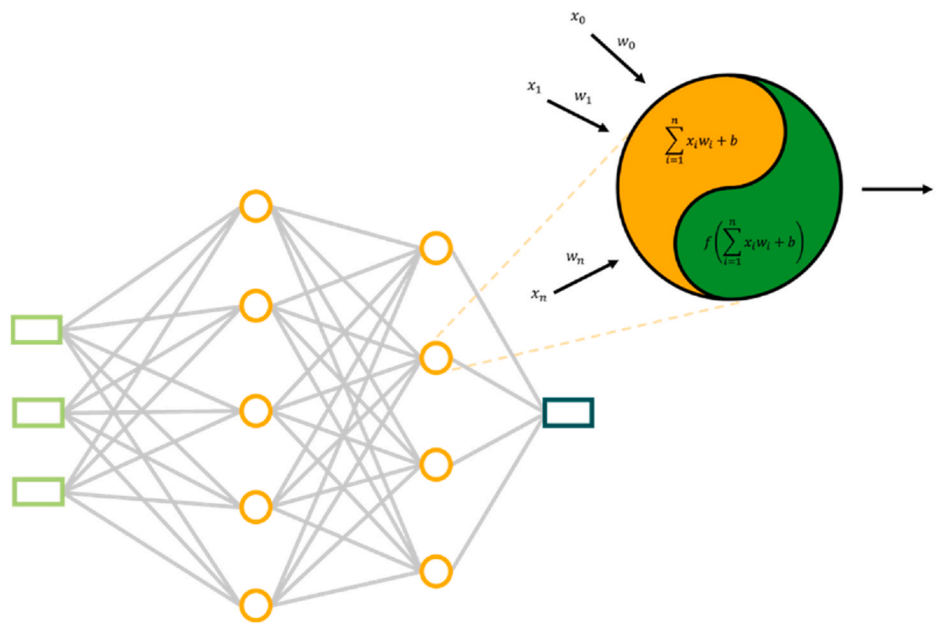

Fig. 6. Representation of Neural Network Topology and zoom on the structure of a neuron.

The proposed models are tested, and their performance is evaluated by recalculating this NMAE metric based on the adjusted predictions they generate, allowing a direct comparison of their effectiveness against the original forecasts.

2.5.2. Normalized Root Mean Squared Error (NRMSE)

The Root Mean Square Error (RMSE) represents the square root of the mean of the squared differences between predicted and actual values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{forecast} - P_{actual}\right)^2} \tag{3}$$

Unlike the NMAE, the RMSE places greater emphasis on larger errors due to the squaring process, making it particularly sensitive to outliers. This sensitivity highlights significant deviations in predictions, complementing the focus of the NMAE on overall error magnitude. Evaluating both metrics provides a more comprehensive understanding of forecasting accuracy, as they capture different aspects of prediction performance.

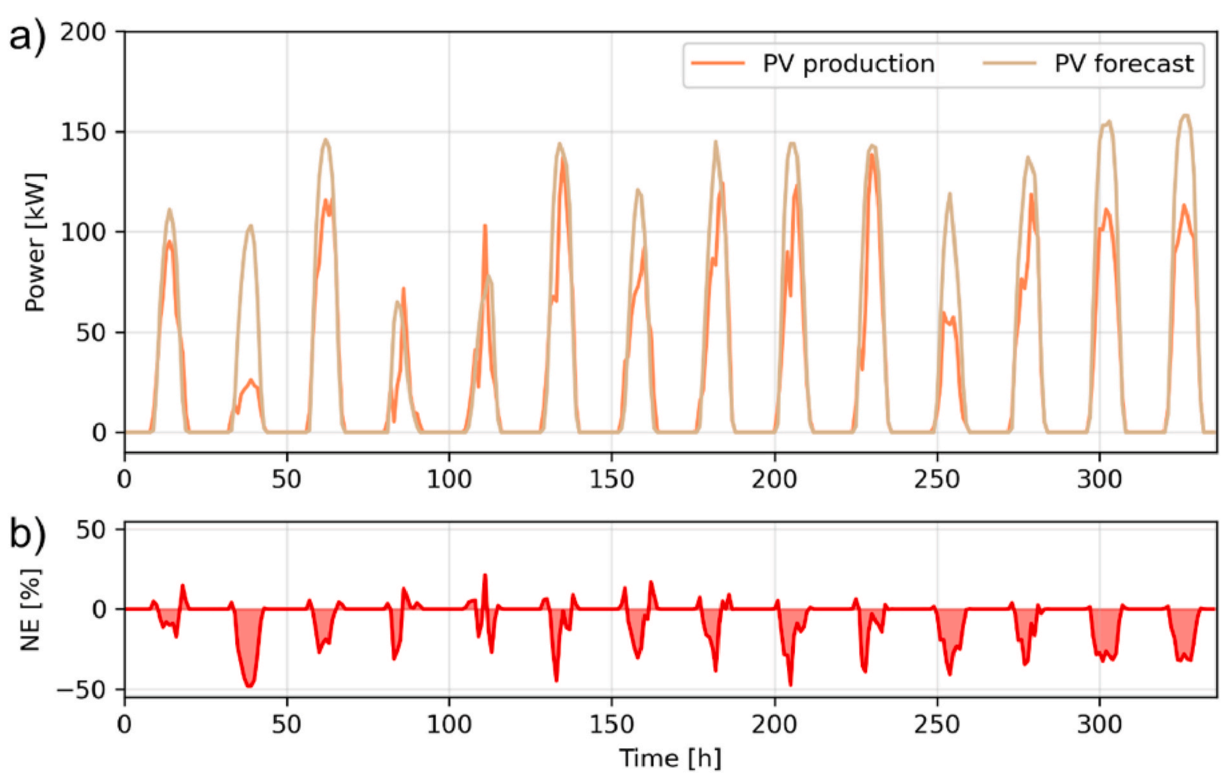

Fig. 7. Comparison on a sample dataset of PV forecast with actual production (a) and NE between actual production and predicted production (b).

Similar to the NMAE, the RMSE can be normalized to the nominal power of the system and expressed as a percentage:

$$NRMSE = \frac{RMSE}{P_{rated}} \cdot 100 \quad (4)$$

The NRMSE of the original forecast from the commercial provider was 15.95 % for the PV field and 14.6 % for the WT.

Although the forecasting methods were primarily optimized to minimize the NMAE, the results section (Section 3) also analyzes the NRMSE of the final results. This secondary evaluation provides insight into whether the forecasts improved from the perspective of this metric and to what extent.

3. Results and discussion

3.1. PV forecast enhancement

The original forecast of solar power production from the commercial provider is characterized by an average NMAE for the selected year of 8.6 %. Fig. 7 (a) considers a sample dataset to show a comparison of the PV forecasts with actual production. The forecast is characterized by a pronounced bias error and is, most of the time, higher than the actual production. Fig. 7 (b) shows the NE between actual production and forecast from the commercial provider. The plot shows how the commercial provider often foresees a power production higher than the actual capabilities. Indeed, the forecast is characterized by a mean NE of −12.2 % and a median NE of −9.7 %, highlighting the presence of an important bias error. In the scatter plot correlating the actual power production with the respective forecast in Fig. 8, most of the points are higher than the bisector line representing the ideal prediction. This behavior means that the predictions, especially of upper power levels, are most of the time higher than the actual realizations.

The histogram of the NE in Fig. 9 confirms this behavior: the distribution of the error shows a peak at low errors, meaning that most of the time the error committed is lower than ±1 %, but another peak around −23 %, confirming the existence of a bias error.

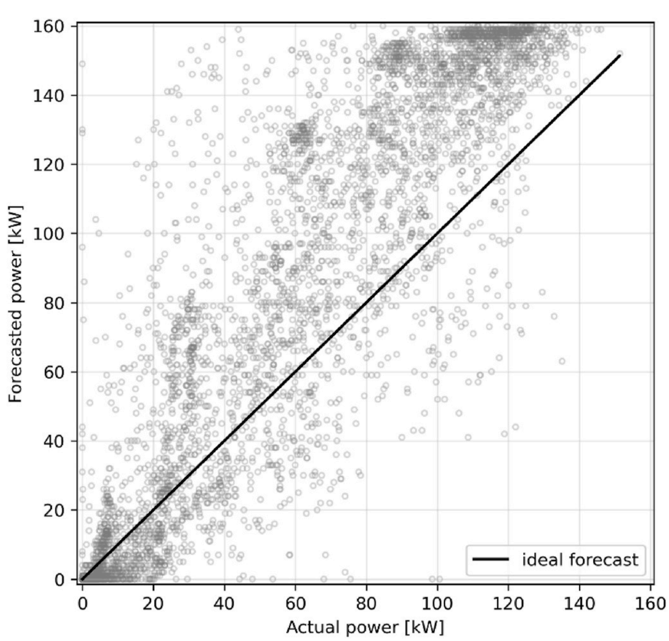

Fig. 8. Scatter of the commercial forecast versus the actual power production from the PV field.

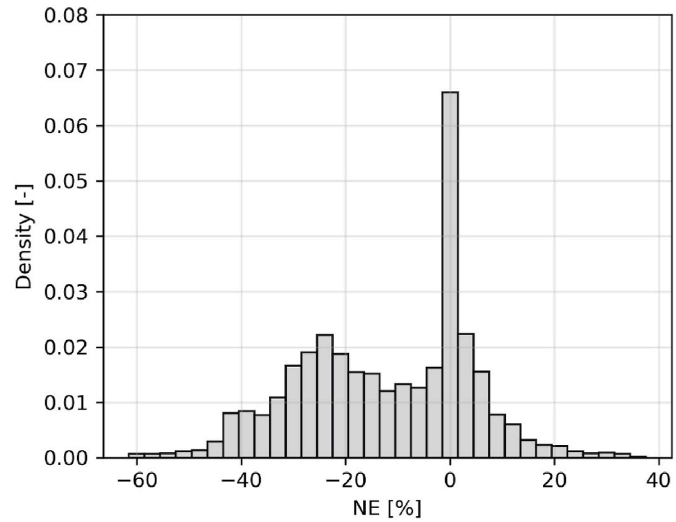

Fig. 9. Histogram of the Normalized Error distribution characterizing the commercial forecast of the PV production.

3.1.1. PV forecast enhancement - linear regression (LR) sensitivity analysis on the number of training days

The LR model was tested by training the algorithm on an expanding range of prior operational data, starting with 3 days and increasing incrementally up to 30 days.

As shown in Fig. 10, the NMAE of the predictions adjusted by the LR method decreases as the number of training days increases. This trend suggests that the algorithm benefits from analyzing a larger historical dataset, resulting in progressively more accurate adjusted forecasts. When trained on 30 days of prior operation, the LR model achieves a forecast with an NMAE of 3.59 % and an NRMSE of 7.68 %, demonstrating significant accuracy improvements with extended training periods.

3.1.2. PV forecast enhancement - Support Vector Regression (SVR) sensitivity analysis on C hyperparameter and number of training days

The SVR model was also tested by training the algorithm on an expanding range of prior operational data. Given the poor results when using narrow training windows, the sensitivity analysis of the SVR considers the range between 18 and 30 days.

The contour plot in Fig. 11 shows the results of the sensitivity

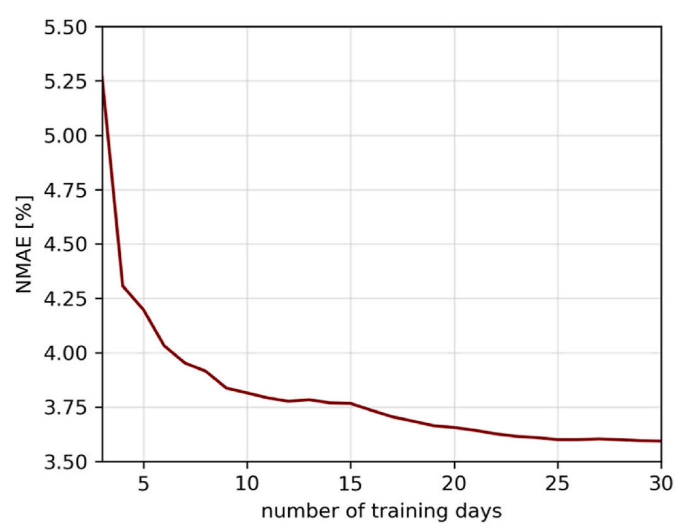

Fig. 10. Sensitivity analysis NMAE of PV forecast corrected by LR, varying the number of days used for the training phase of the algorithm.

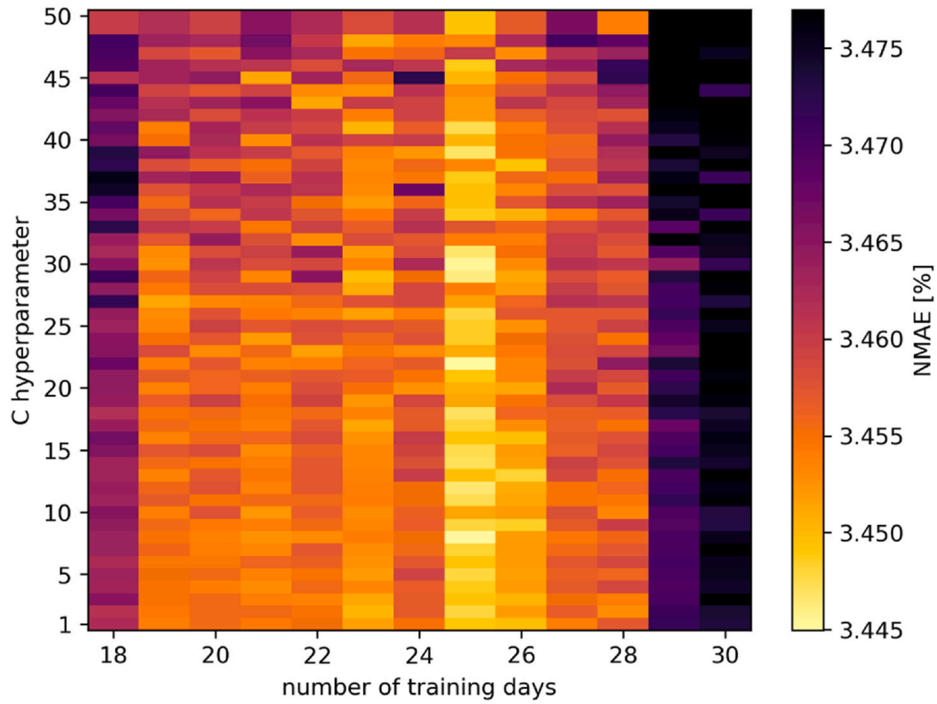

Fig. 11. Sensitivity analysis on the NMAE of the forecast corrected by SVR, varying the C parameter and the number of training days used for the training phase of the algorithm, considering an ε equal to zero.

analysis made by varying the value of the hyperparameter C and the number of days used for the training phase. The value of C varies from 1 to 50, with a resolution of 1, while the value of ε was fixed to zero. Results show that the best combination to minimize the forecasting error is given by 25 training days. In this case, the optimal number of training days for the algorithm differs from the maximum tested. Unlike LR, the correlations identified by SVR can be influenced or biased by data points that are too distant in time. This suggests that SVR performs best when trained on a carefully selected, more recent subset of historical data, allowing the method to capture relevant patterns without being skewed by older, potentially less relevant information.

The dependency on the C parameter was instead proved to be much weaker.

Fig. 12 shows how the NMAE varies with different combinations of C and ε, considering a number of days for the training phase equal to 25. This time, the hyperparameter ε varies from zero to one, with a

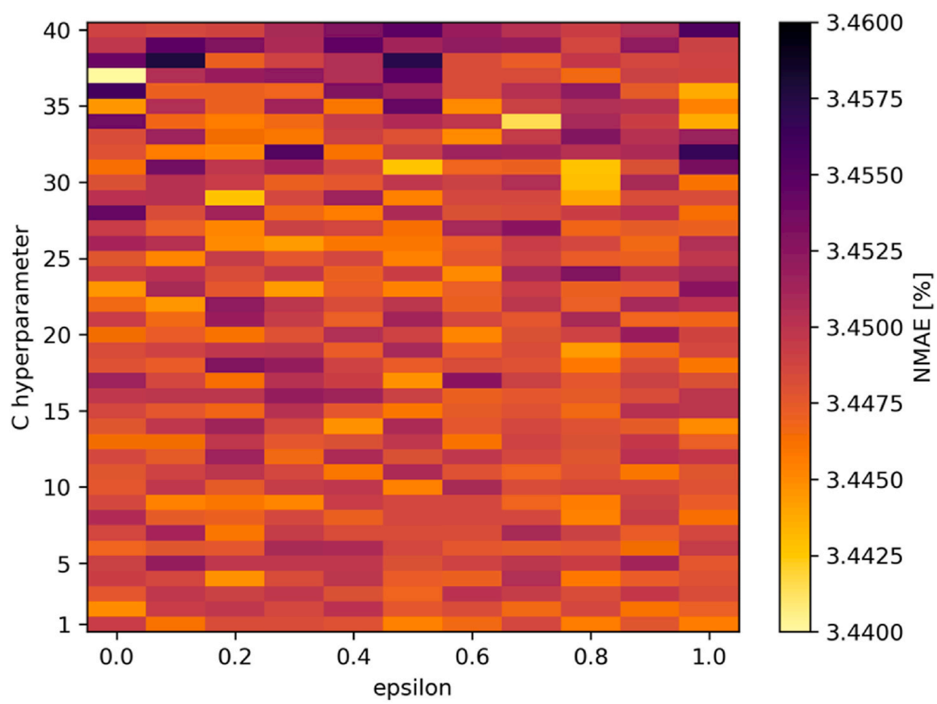

Fig. 12. Sensitivity analysis on the NMAE of the PV forecast corrected by SVR, varying the C parameter and ε for the training phase of the algorithm, considering a number of training days equal to 25.

resolution of 0.1. Extensive testing has revealed that the most effective results in minimizing the NMAE of predictions are achieved when the parameter ε is set to zero. In this specific application, it appears advantageous to always apply penalties to predicted points that deviate significantly from the target, ensuring greater precision by discouraging large prediction errors.

While the optimal null value of ε indicates that penalizing errors during the training phase is always beneficial, the best-performing method made use of a C value of 37. This relatively high C suggests that, in addition to strict error penalization within the tolerance band, the model benefits from allowing increased tolerance for errors outside the band. Indeed, the impact of outliers is higher in this case, and the optimal SVR performs slightly worse than LR in NRMSE terms, lowering the metric down to 7.8 %.

3.1.3. PV forecast enhancement - neural networks (NNs) sensitivity analysis on topology and number of training days

The NNs were tested by adjusting the network topology, the maximum number of iterations allowed during training, and the number of days used for training.

The first tests revealed that the default maximum of 200 iterations was consistently insufficient to achieve convergence for this type of application. A more reasonable limit, which typically ensures convergence without significantly increasing computational time, is $1{\bullet}10^4$ iterations. All results presented in this section were obtained using this iteration limit.

The contour plots in Figs. 13 and 14 illustrate the results of the sensitivity analysis, exploring the combined effects of varying the number of training days and the neural network (NN) topology. The NN topology is represented by a sequence of numbers in round brackets, where each number indicates the number of neurons in each layer, progressing from the input layer to the output layer.

Various network topologies were tested to identify the most effective one. A first sensitivity analysis considered small networks (Fig. 13), involving structures made by three to four hidden layers, each containing a number of neurons between five to ten. A second sensitivity analysis considered larger networks (Fig. 14), made by up to eight layers and comprising a higher number of neurons, up to forty per layer.

Structures with fewer than 28 neurons performed the worst. Given the complexity of the forecasting problem, a larger number of neurons is necessary to detect patterns that simpler models might miss. However, results indicate that adding a fourth layer or increasing the number of neurons beyond a certain point did not improve performance. This suggests a potential issue of overfitting, where the network memorizes the training data rather than generalizing well. Since the training phase only involves two time series, an overly complex model may fail to capture any additional trends. The optimal balance appears to be a three-layer structure, each with ten neurons.

The results indicate that the optimal configuration for the neural network method involves training over 25 days. The choice of 25 training days aligns with the optimal duration for the SVR model, suggesting that using too many training days could introduce bias into the algorithm.

After the application of the best-performing NNs, the NMAE of the improved prediction is reduced to 3.34 %. Additionally, the NNs model performs better also in terms of NRMSE, with the optimal topology trained over 25 days achieving a reduction of the metric to 7.62 %.

3.2. WT forecast enhancement

The original forecast of WT production from the commercial provider is characterized by an average NMAE for the selected year of 10.25 %. Fig. 15 (a) shows a sample two-weeks dataset to show a comparison between the WT forecasts with actual production. The WT forecast succeeds in capturing the average trend of wind production but fails in correctly predicting the oscillations around it.

Fig.15 (b) shows the NE between actual production and predicted production. Unlike the error that affects solar production, the error behind the WT forecast is much more randomly distributed. The scatter plot correlating the actual power production with the respective forecast in Fig. 16 shows that this time the points are distributed less predictably with respect to the case of solar prediction (Fig. 8). The high concentration of points above the bisector line in the lower left part of the graph means that the commercial forecast is prone to overestimating wind

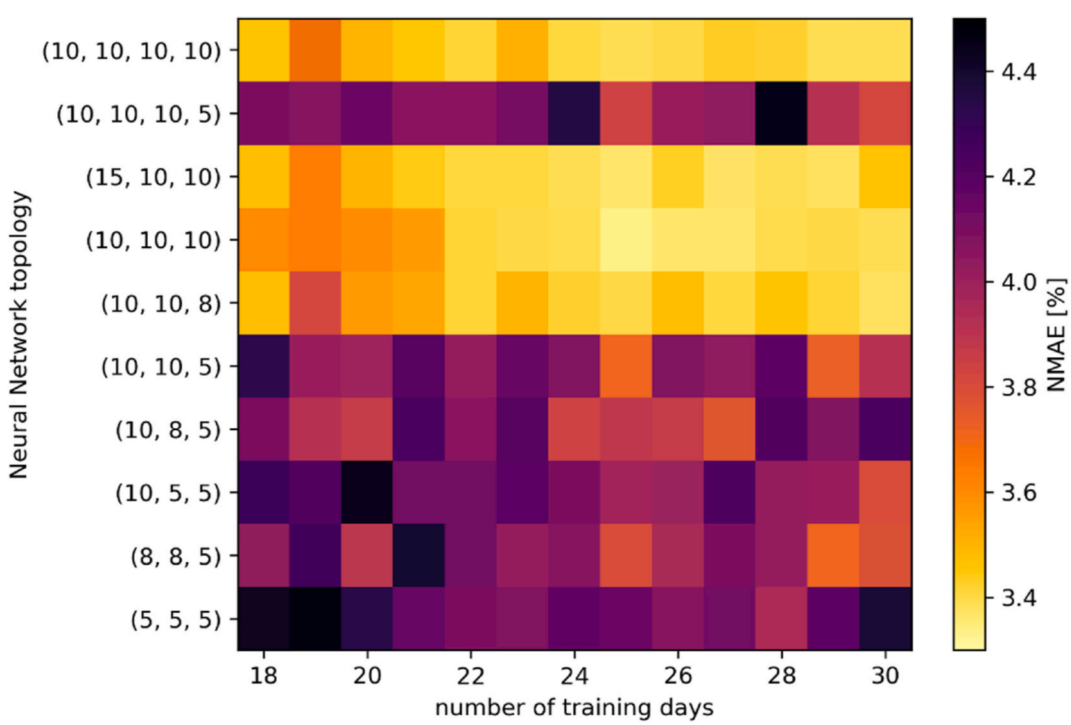

Fig. 13. Sensitivity analysis on NMAE of PV forecast varying the topology and the number of training days for NNs – small networks.

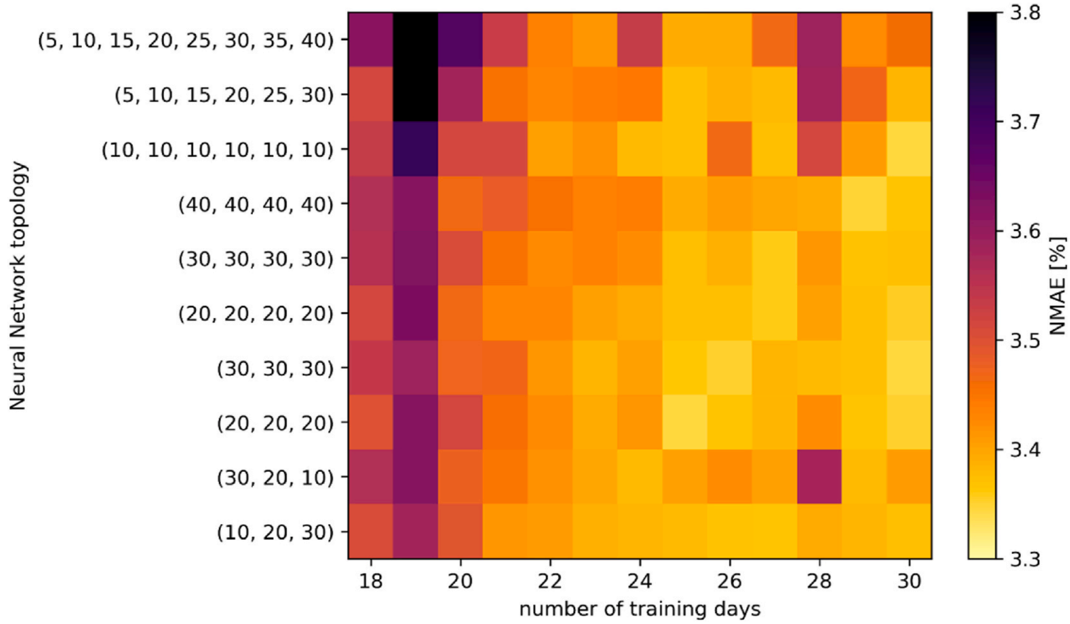

Fig. 14. Sensitivity analysis on NMAE of the PV forecast varying the topology and the number of training days for NNs – larger networks.

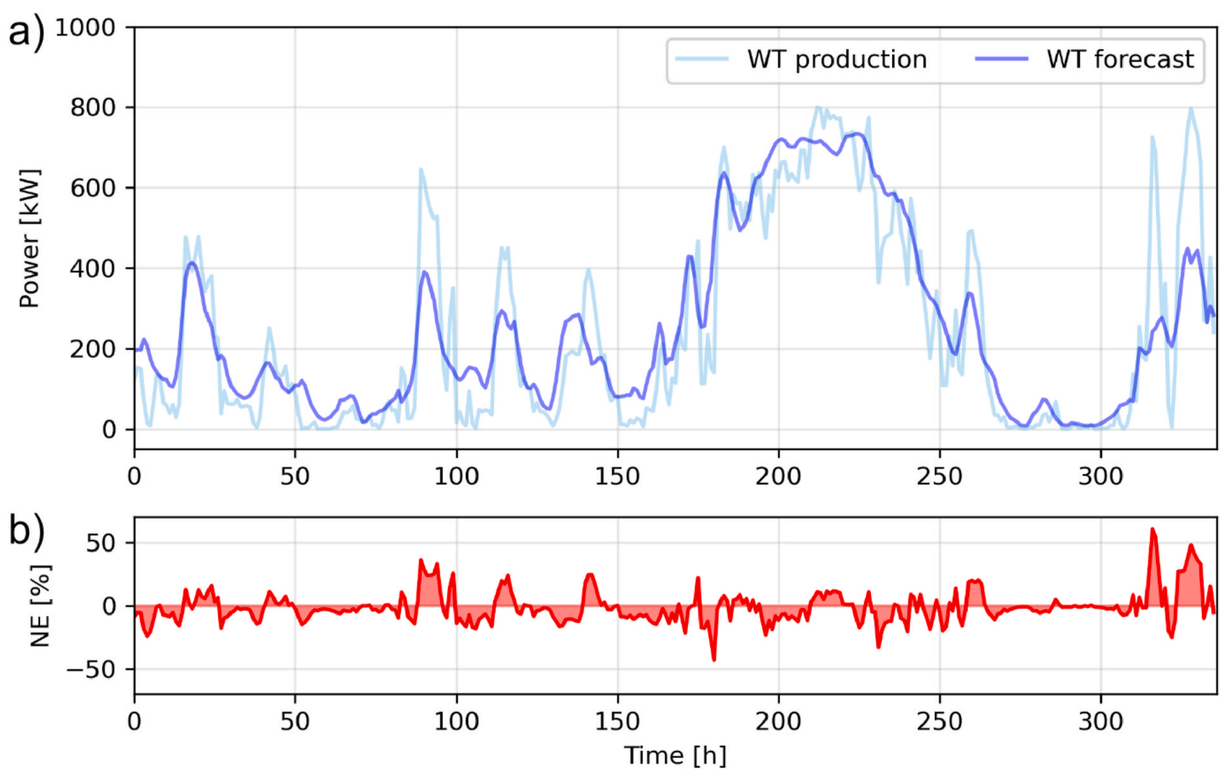

Fig. 15. Comparison of two weeks of operation of WT forecast with actual production (a) and NE between actual production and predicted production (b).

power production at low power ratings. The mean NE is indeed equal to −1.9 % and a median NE of −2.5 %. The histogram of the NE in Fig. 17 shows that the distribution of the error is almost symmetric but slightly shifted towards negative errors.

To test the capabilities of data-driven approaches, the same methodology applied to enhance the PV power production was applied to the WT time series.

3.2.1. Linear regression (LR) sensitivity analysis on the number of training days

Fig. 18 shows that, at first, the average NMAE of the WT prediction decreases with the number of previous days used for the training phase of the LR method. Unlike results of the same method applied for the PV forecast correction (Fig. 10), the error increases also in LR when an excessive number of days is used for the training. The curve presents a minimum after 21 days, then, when the algorithm is trained on too many days, the accuracy drops and the error increases again. Trained on 21 days of previous operation, LR produces a WT production forecast

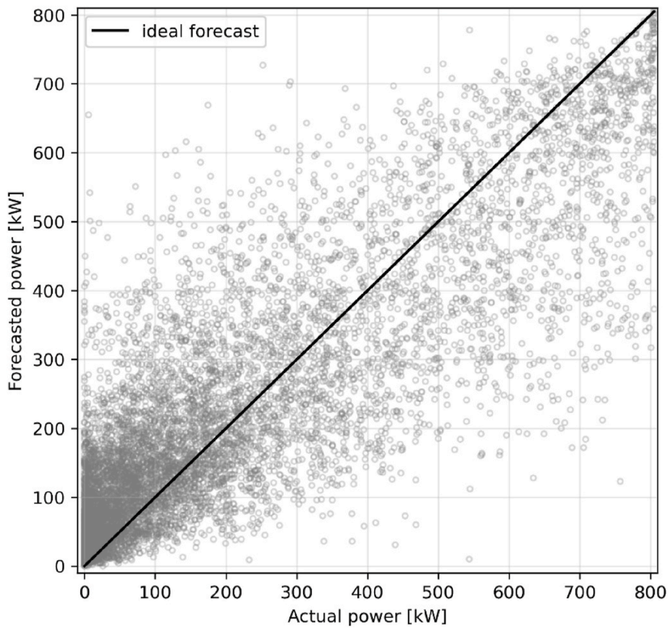

Fig. 16. Scatter of the commercial forecast versus the actual power production from the WT.

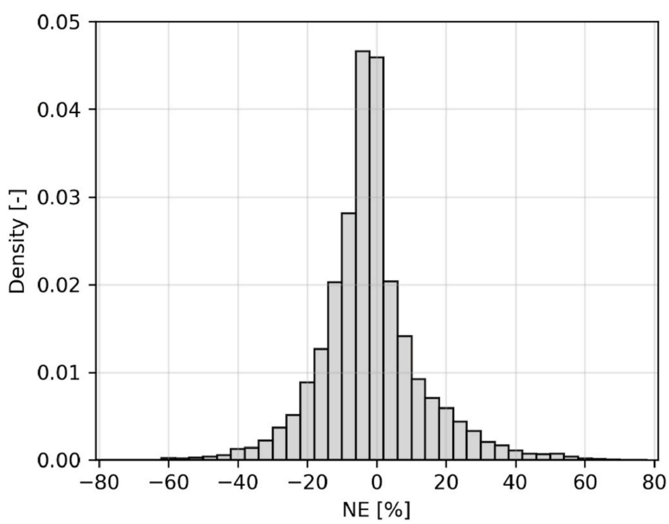

Fig. 17. Histogram of the Normalized Error distribution characterizing the commercial forecast of the WT production.

characterized by an NMAE of 9.86 %, but the NRMSE remains at 14.6 %.

3.2.2. WT forecast enhancement - Support Vector Regression (SVR) sensitivity analysis on C hyperparameter and number of training days

As for the PV, the SVR method was tested first varying the number of training days and the value of the hyperparameter C. The contour plot in Fig. 19 shows the results of the sensitivity analysis varying the number of training days from 16 to 30 and the C parameter from 1 to 50. Results show that the best combination to minimize the error in forecasting the WT is given by 22 training days. Again, the effect of the C variation on the effectiveness of the algorithm is much weaker than the effect of changing the size of the training window.

As for the LR, the optimal number of days to train the algorithm is different than the maximum, since the correlation that SVR can find may be biased by values too far away in time.

Then, the SVR method was tested with the hyperparameter ε varying from zero to one, with a resolution of 0.1, in various combinations with the C parameter (Fig. 20). Differently from the PV, this time the best strategy involves a value for ε different from zero. Due to the different nature of wind production, it appears advantageous to allow for the existence of a tolerance band and not penalize points whose deviation is still close from the target.

Even with tuned parameters, the SVR performs only slightly better than the LR. Trained on a time window of 22 days, the combination that produced the lowest error combined a C of 37 with an ε of 0.4 to reach a NMAE of 9.84 %.

3.2.3. WT forecast enhancement - neural networks (NNs) sensitivity analysis on topology and number of training days

The sensitivity analysis regarding the application of NNs to correct the WT forecast focused on the optimization of the topology in combination with different time windows (number of previous days) used for the training. As for the PV field, at least $1 \bullet 10^4$ iterations were required to reach convergence during the training phase.

The contour plots in Figs. 21 and 22 show the results of the sensitivity analysis varying the number of days to train the algorithm and the topology of the neural network. Fig. 21 involved smaller networks, while in Fig. 22 larger networks with a higher number of neurons are considered.

Networks involving a higher number of neurons performed better than simpler networks. The presence of a layer involving too few neurons always proved to be a limitation: all networks comprising a layer with only 5 neuros performed worse than networks involving only layers with a higher number of elements.

The lowest NMAE reached by the sensitivity analysis is 10.04 with NNs involving three hidden layers each containing 30 neurons trained on 27 days of operation. NNs performed considerably worse than regression methods, meaning that too complex algorithms may be detrimental to the improvement of the WT forecast.

3.3. Comparative analyses between the approaches

This section presents a comparison of the performance of the three tested methods (SVR, NNs, and LR) across different numbers of training days. The discussion also examines how the improved forecast, achieved using the method with the best performance, compares to the original one. Finally, the results of the sensitivity analysis are summarized, focusing on the variation and impact of hyperparameters on the outcomes. All comparisons are conducted both for solar and wind energy predictions.

3.3.1. PT forecast enhancement - comparison between the approaches

Fig. 23 shows a comparison among the NMAE of the PV improved forecast resulting from LR (brown line), SVR (yellow line), and NNs (orange line), varying the number of days that the algorithm can use for the training phase. The results from SVR and NNs are derived from the optimal hyperparameter tuning for each specific number of training days. LR performs best when trained on the maximum possible number of days (i.e., 30), whereas both NNs and SVR achieve their highest performance when trained on 25 days before the "application day." Additionally, LR consistently improves the original forecast, but the other two algorithms (NNs and SVR) always outperform LR, regardless of the number of training days used.

SVR is the most effective choice when the training period is fewer than 22 days, as this method accurately captures the linear relationship between the production of the following day and its trend from previous days. However, when more than 22 days of training data are available, NNs outperform SVR. NNs excel in capturing trends that linear models, such as LR and SVR, cannot detect, leading to more precise predictions when sufficient training data is provided.

Fig. 24 shows the application of NNs trained on 8, 25, and 30 days to correct the forecast of a conventional day, after the realization of an

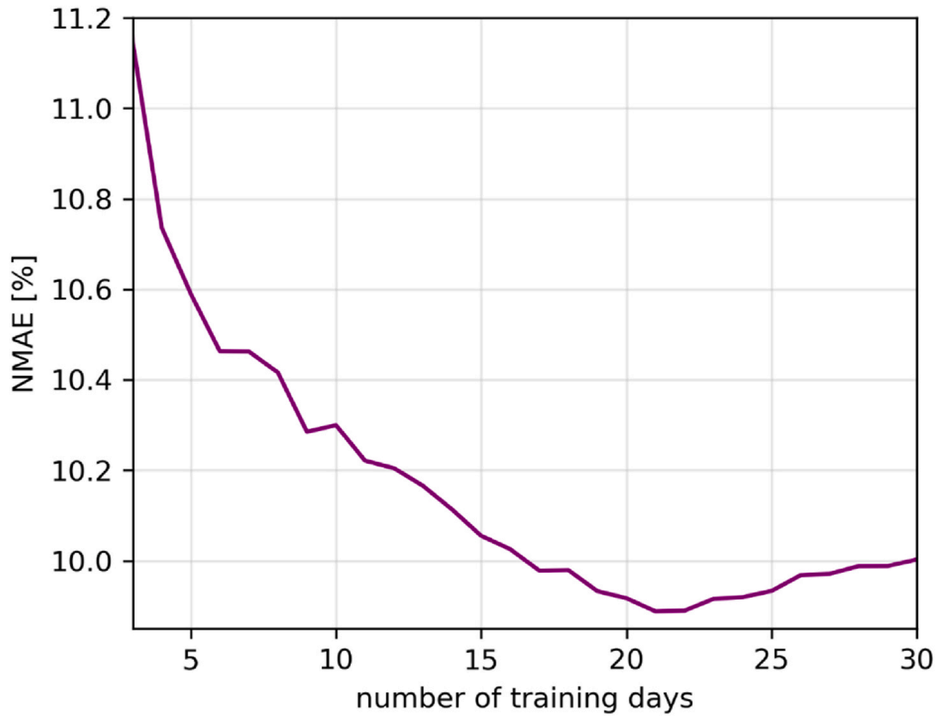

Fig. 18. Sensitivity analysis NMAE of WT production forecast corrected by LR, varying the number of days used for the training phase of the algorithm.

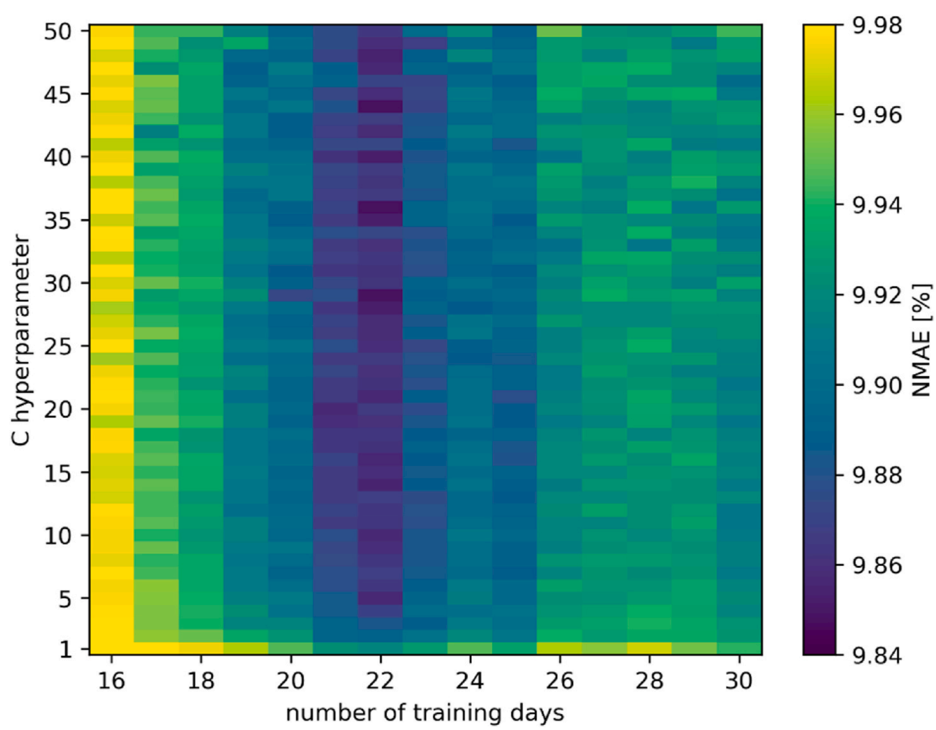

Fig. 19. Sensitivity analysis on the NMAE of the WT forecast corrected by SVR, varying the C parameter and the number of training days used for the training phase of the algorithm, considering $\varepsilon = 0$.

atypical day. When trained on an excessively long dataset, the algorithm may find a strong seasonal correlation that it then applies all the time. In this case, the algorithm trained on 30 days (orange line), predicts a high peak of production based on a previous month characterized by a high solar irradiance. Shorter training periods prevent the model from capturing longer-term dependencies but may introduce a bias error if the recent period is atypical. In the example, the algorithm trained on 8 days (green line) underestimates the power production in the application because the day before was characterized by a low producibility. The algorithm trained on 25 days was instead able to provide a forecast much closer to the actual production, finding a trade-off between using enough days to prevent recent biased data without capturing uncorrelated long-term trends.

The example of forecast correction over a two-week period in Fig. 25

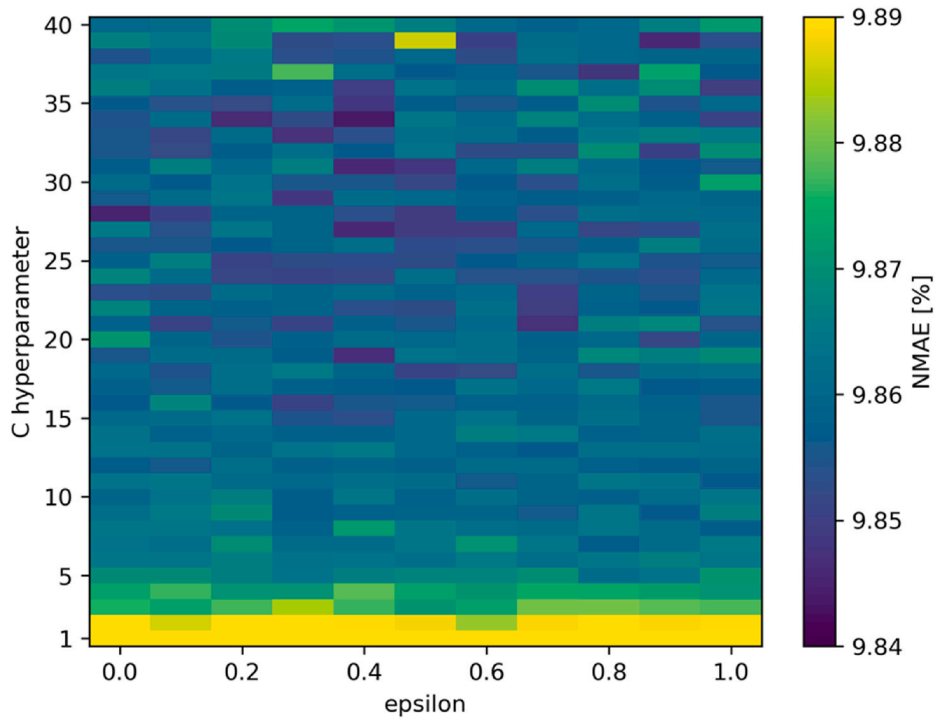

Fig. 20. Sensitivity analysis on the NMAE of the WT forecast corrected by SVR, varying the C parameter and ε used for the training phase of the algorithm, considering a number of training days equal to 22.

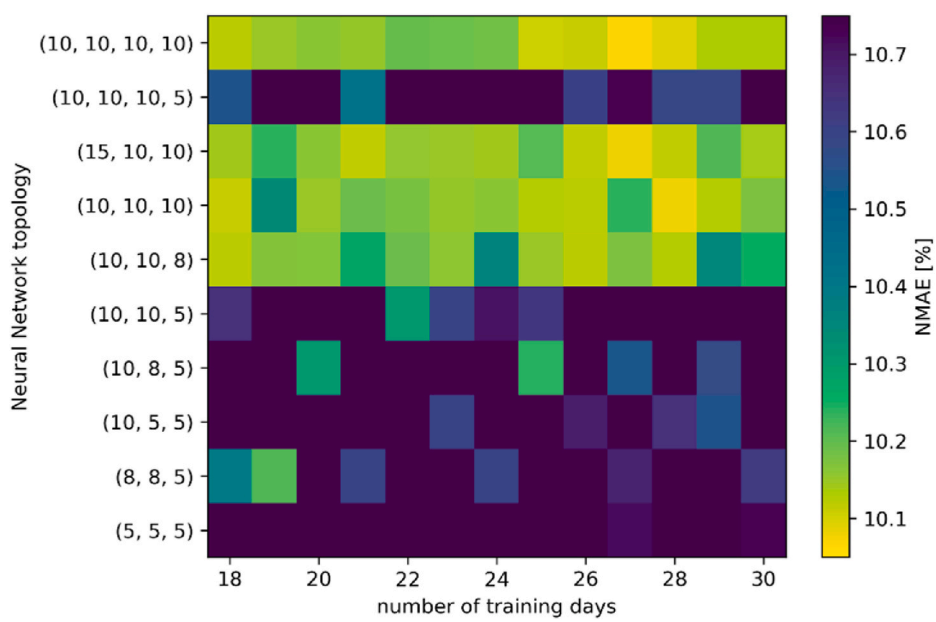

Fig. 21. Sensitivity analysis on NMAE of the WT forecast varying the topology and the number of training days for NNs – small networks.

shows how the best-performing method tested in this study (NNs trained on 25 days of operation) modifies the time series. While the results represent the average accuracy of forecasts over the entire year of operation, this graph provides a focused view of the comparison for a specific period.

Fig. 25 (a) shows the time comparison between the original forecast, the improved forecast, and the actual production. The improved forecast consistently aligns more closely with the actual production, as the model is able to reduce the initial bias error. The subsequent figures highlight the trend of the NE between the actual production and the original forecast (b), as well as between the actual production and the improved forecast (c). Fig. 25 (b) shows that the original NE is consistently negative, indicating that the forecast consistently overestimated production (bias error). In contrast, Fig. 25 (c) demonstrates that the NE of the improved forecast is reduced and more evenly distributed. This shows that the NNs method successfully eliminates the bias error and enhances the accuracy of the forecast.

Table 1 summarizes the results of the sensitivity analysis conducted on the three tested methods, highlighting their best possible outcomes for improving PV forecast accuracy.

Starting with an original NMAE of 8.6 %, LR reduces the error to 3.59 % when trained on 30 days of past operation. LR performs

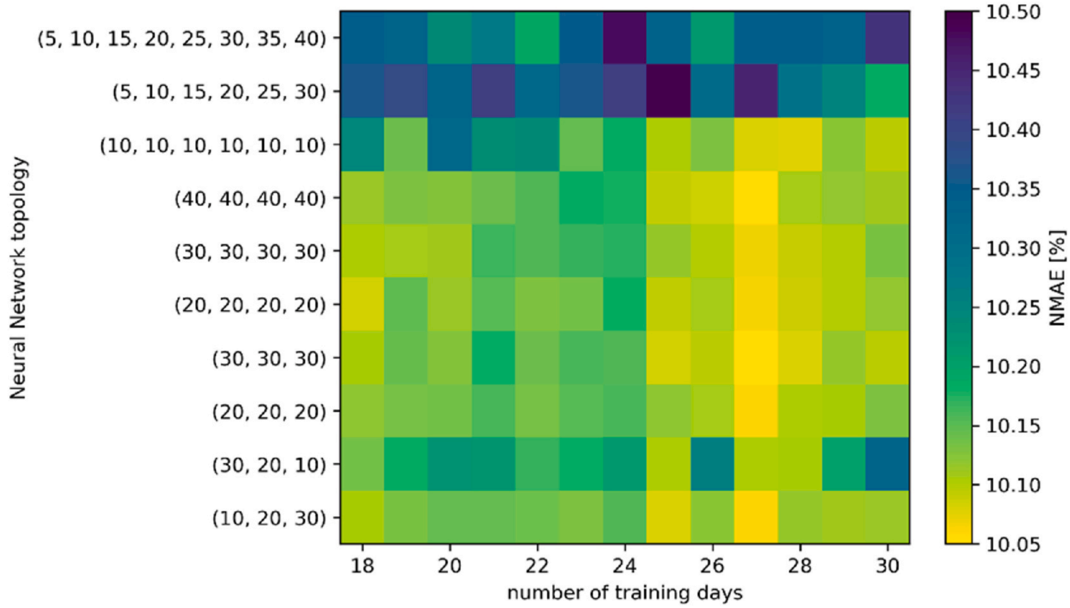

Fig. 22. Sensitivity analysis on NMAE of the WT forecast varying the topology and the number of training days for NNs – larger networks.

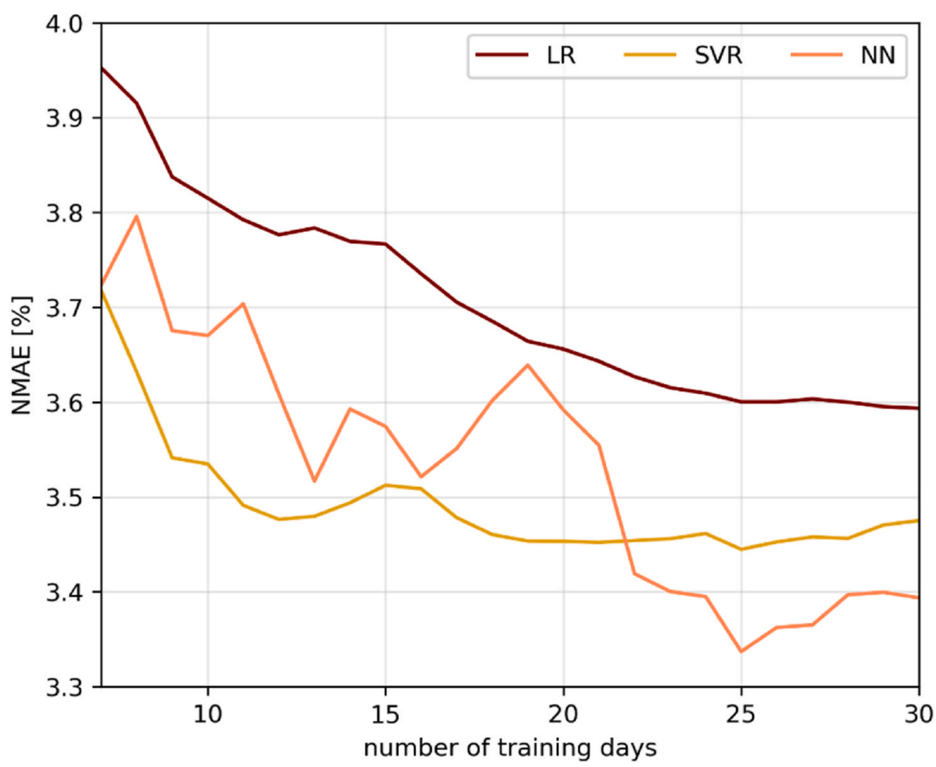

Fig. 23. Comparison among the NMAE of the PV improved forecast between LR (brown line), SVR (yellow line), and NNs (orange line) varying the number of days that the algorithm can use for the training phase.

optimally when trained on the maximum number of previous days. In contrast, SVR and NNs show improved performance when trained on a higher number of previous days, but not an excessively large dataset.

SVR, with ε set to zero and C set to 37, achieves a further reduction in NMAE when trained on 25 days of operation, reaching 3.44 %. The sensitivity analysis on SVR reveals that always applying a penalty to errors during training enhances its performance, but better forecasts come from methods that also mitigate this penalty.

NNs provide the best forecast, lowering the NMAE to 3.34 % when trained on 25 days of operation. However, NNs require careful configuration: they must have a sufficient number of neurons to capture the complexity of the data patterns, while avoiding excessive complexity to prevent overfitting. Additionally, a sufficiently high number of iterations (10k) is necessary for the method to converge during training.

From an NRMSE perspective, all tested methods show improvement. LR reduces the NRMSE from the original 15.59 %–7.68 %. SVR slightly underperforms in this regard, lowering the NRMSE to 7.8 %. NNs again offer the greatest benefit, reducing the metric to 7.62 %.

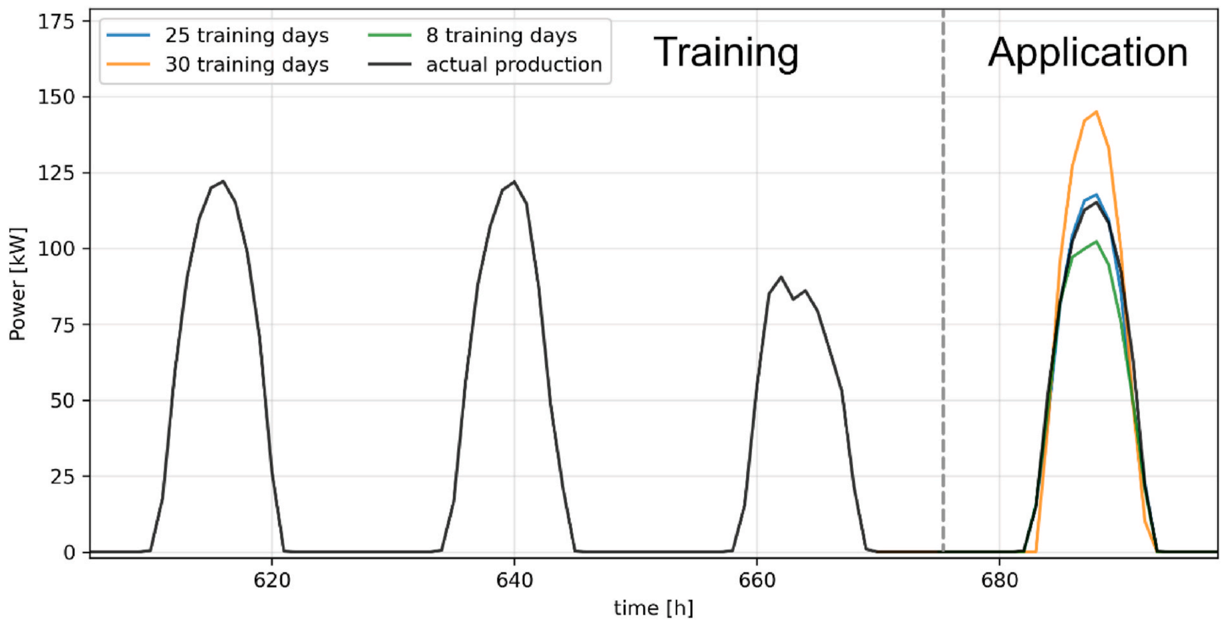

Fig. 24. Application of NNs trained on 8, 25 and 30 days to correct the forecast of a conventional day, after the realization of an atypical day.

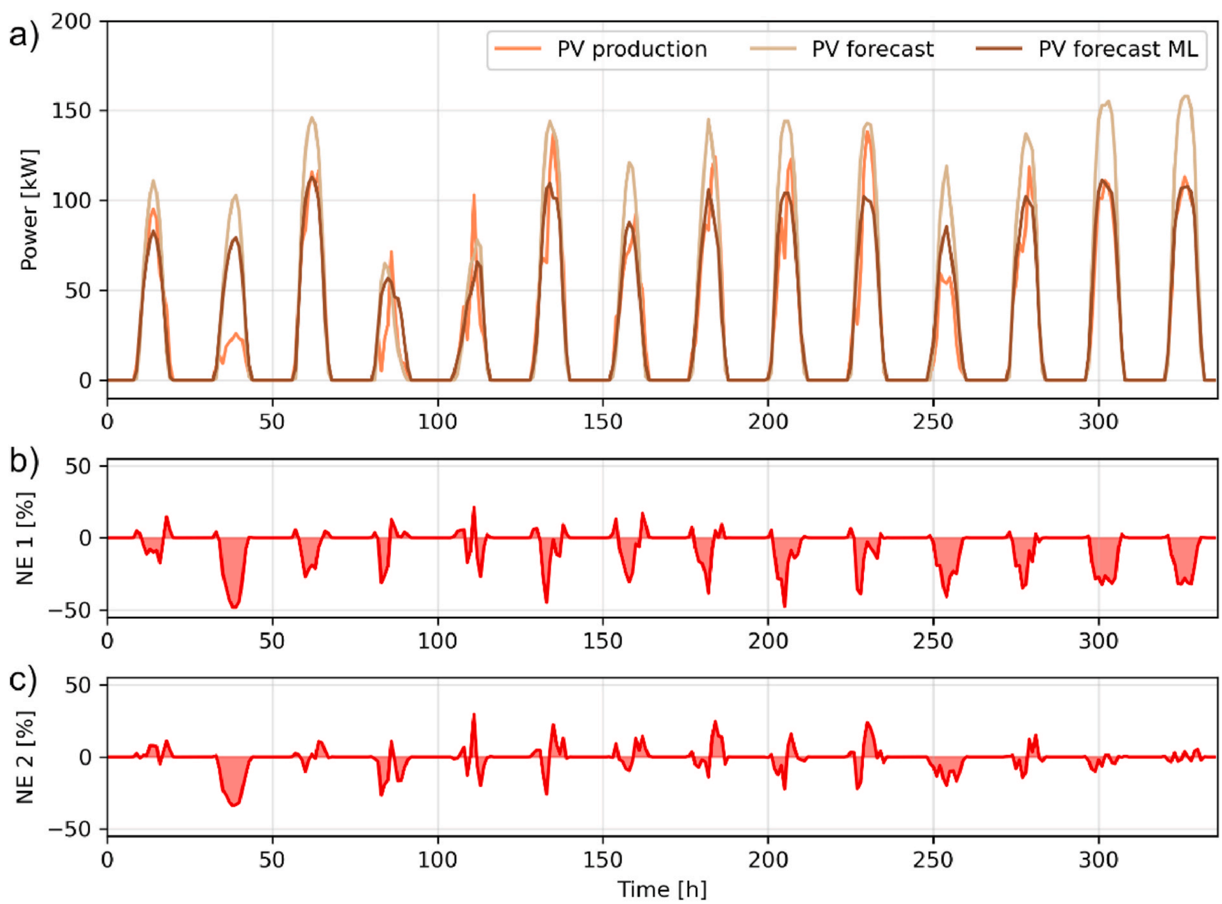

Fig. 25. Example of forecast correction of solar production during two weeks of operation. a) comparison between the original forecast, the improved forecast, and the actual production b) normalized error trend between the original forecast and the actual production c) normalized error trend between the improved forecast and the actual production.

Fig. 26 compares the NE distribution of the PV production forecast corrected using LR, SVR, and NNs and shows the scatter of the relationship between the actual power production and the predicted one after the application of the three methods. All methods successfully removed the second peak of the error distribution characterizing the original commercial forecast (Fig. 9). LR remains slightly more biased

Table 1
Comparison of best results from different algorithms for the PV forecast enhancement.

Metric	Original	LR	SVR	NNs
Optimal number of training days	–	30	25	25
Optimal Hyperparameters	–	–	ε = 0 C = 37	Max iter. = 10k Topology = 10,10,10
Final NRMSE	15.95 %	7.68 %	7.8 %	7.62 %
Final NMAE	8.6 %	3.59 %	3.44 %	3.34 %
Max NE	93.1 %	53.4 %	58.7 %	75.6 %
Mean NE	−12.2 %	−0.59 %	−0.57 %	−0.61 %
Median NE	−9.7 %	−0.95 %	0.21 %	0.12 %

towards negative errors with respect to SVR and NNs, with the latter methods showing the most symmetrical error distribution among the three. All methods also reduced the maximum NE from 93.1 % to 53.4 % (LR), 58.7 % (SVR), and 75.6 % (NNs).

To facilitate the comparison between the distributions, the box-and-whiskers plot in Fig. 27 directly compares the distribution of the original NE and the NE after the application of the three methods. From the original error distribution with a mean NE of −12.2 % and a median NE of −9.7 %, the application of machine learning succeeded in centering the error distribution around zero and removing biases in the forecast.

3.3.2. *WT forecast enhancement - comparison between the approaches*

Fig. 28 shows the comparison among the NMAE of the improved WT forecast resulting from LR (purple line), SVR (blue line), and NNs (turquoise line), varying the number of days that the algorithm can use for the training phase. Results from SVR and NNs come from the best hyperparameters tuning when working with that specific number of training days.

SVR and LR perform best when using a limited number of previous days for the training phase. Both present the best accuracy when working with a time window of around 21–22 days. NNs perform better when working with more previous days, but they always perform worse than regression methods. In general, for the WT forecast enhancement, the SVR always represents the best solution.

Fig. 29 shows an example of WT forecast correction during two weeks of operation, using the best method tested in this work: SVR trained on 22 days of operation. In particular, Fig. 29 (a) shows the time comparison between the original forecast, the improved forecast, and the actual production. Due to the much higher complexity and the origin of the forecast error, the improved forecast is scaled in some points and not always closer to the actual production. The figures below show instead the NE trend between the actual production and the original forecast (Fig. 29 (b)) and the improved forecast (Fig. 29 (c)). The difference between the time series is hardly noticeable this time.

Table 2 summarizes the results from the sensitivity analysis made on the three tested methods and their best possible outcome for the enhancement of the WT forecast. From the original NMAE of 10.25 %, LR trained on 21 days of previous operation can reduce the error down to 9.89 %. SVR tuned with ε equal to 0.4 and C equal to 37 can reduce the NMAE even more when trained on 22 days of operation, bringing the metric to 9.86 %. Results of the sensitivity analysis on SVR show that it is better to always assign a penalty to errors during the training phase of the method, but better forecasts come from methods that mitigate this

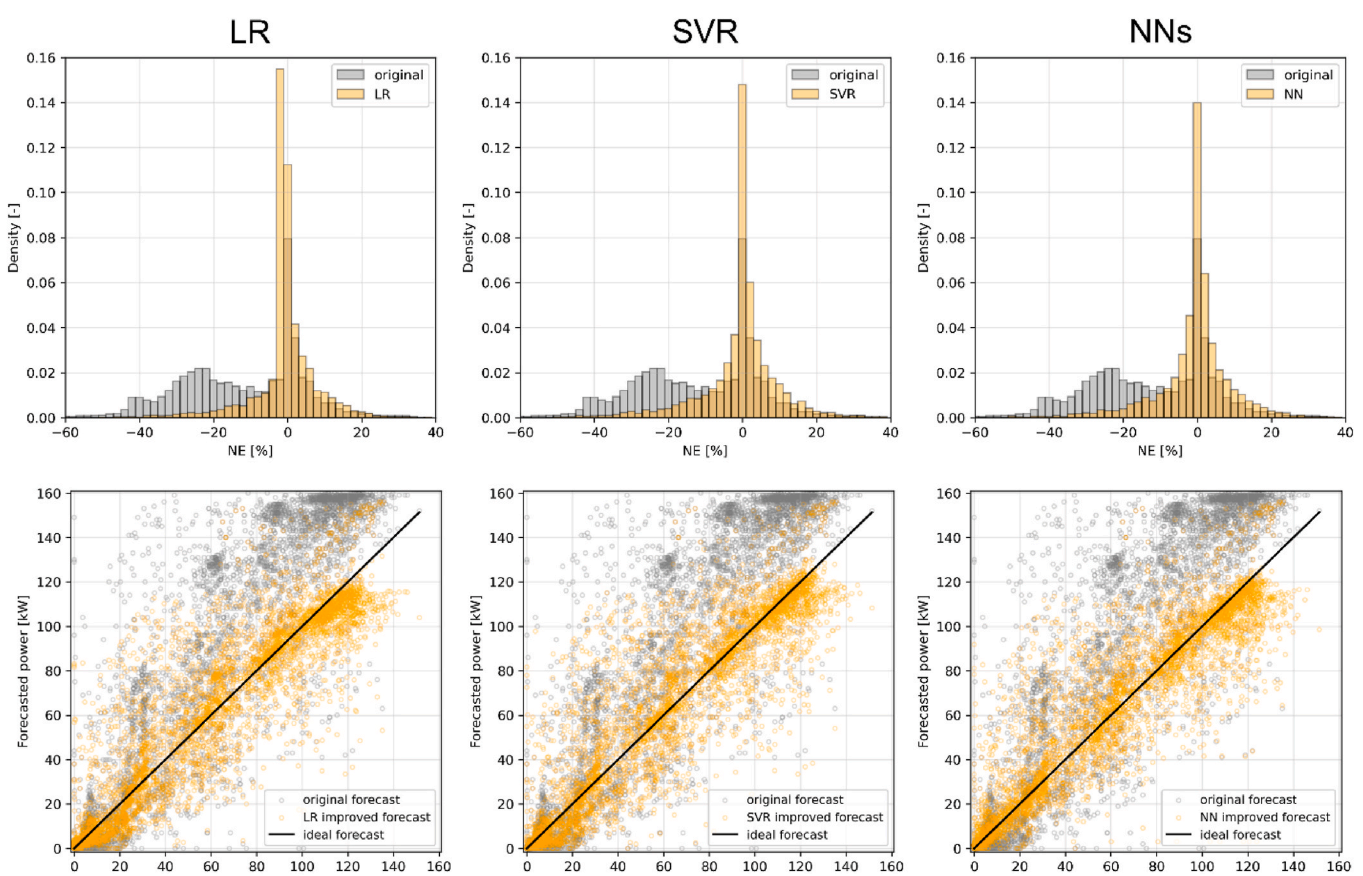

Fig. 26. Histograms of normalized error distribution and scatter of the forecasted power vs actual power production from the PV field after the application of LR, SVR, and NNs.

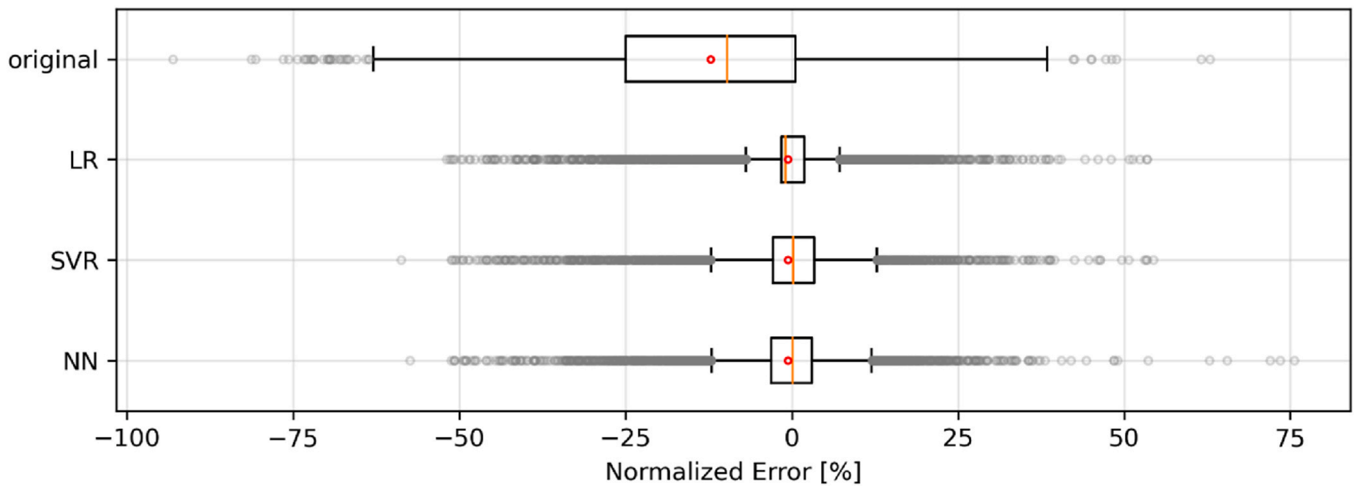

Fig. 27. Box-plot comparison of the normalized error characterizing the original forecast of the PV field production and forecast improved by LR, SVR and NNs.

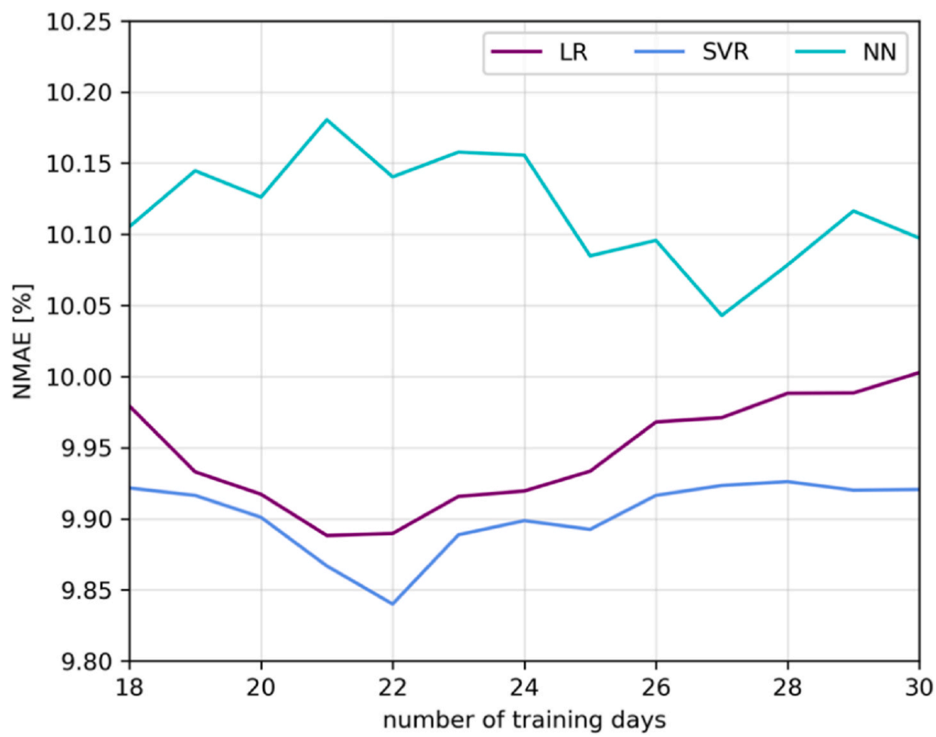

Fig. 28. Comparison among the NMAE of the WT improved forecast between LR (purple line), SVR (blue line), and NNs (turquoise line) varying the number of days that the algorithm can use for the training phase.

penalty.

NNs lead to a worse forecast and lower the error only down to 10.04 % when they are trained on 27 days prior to their application. When working with uncorrelated time series, complex methods try to find complex correlations that end up increasing the final error and compromising the accuracy of the forecast.

In this case, regression methods are able to decrease a specific error metric as the NMAE. However, from other perspectives, like the one of the NRMSE, the accuracy of the forecast remains the same. Since the NRMSE is more sensitive to outliers, ML methods are not able to significantly improve this metric. A data-driven approach is effective in reducing bias errors, and in the case of wind energy forecasting, this results in a slight improvement in the NMAE. However, as the ML methods are unable to address the underlying issue, i.e., that the forecasts tend to follow the average production rather than capturing its fluctuations, there is little improvement from the perspective of NRMSE. This limitation prevents any substantial change in the NRMSE, as the methods do not fully account for the variability in the production data.

Fig. 30 compares the NE distribution of the WT production forecast corrected using LR, SVR, and NNs, and shows the scatter of the relationship between the actual power production and the predicted one after the application of the three methods. As for the application to the solar forecast, the methods improved the symmetry of the error, reducing the negative bias of the commercial forecast at low power levels. The median NE was reduced from −2.5 % to −0.8 % (LR), 0 % (SVR), and −1.5 % (NNs), centering the distribution much closer to zero. The methods also improved the mean NE, which was reduced from −1.9 % to −0.24 % (LR), 0.77 % (SVR), and −0.35 % (NNs). Even if the gain in terms of NMAE was limited, the improvement in the error distribution makes the prediction less prone to introducing problems related to

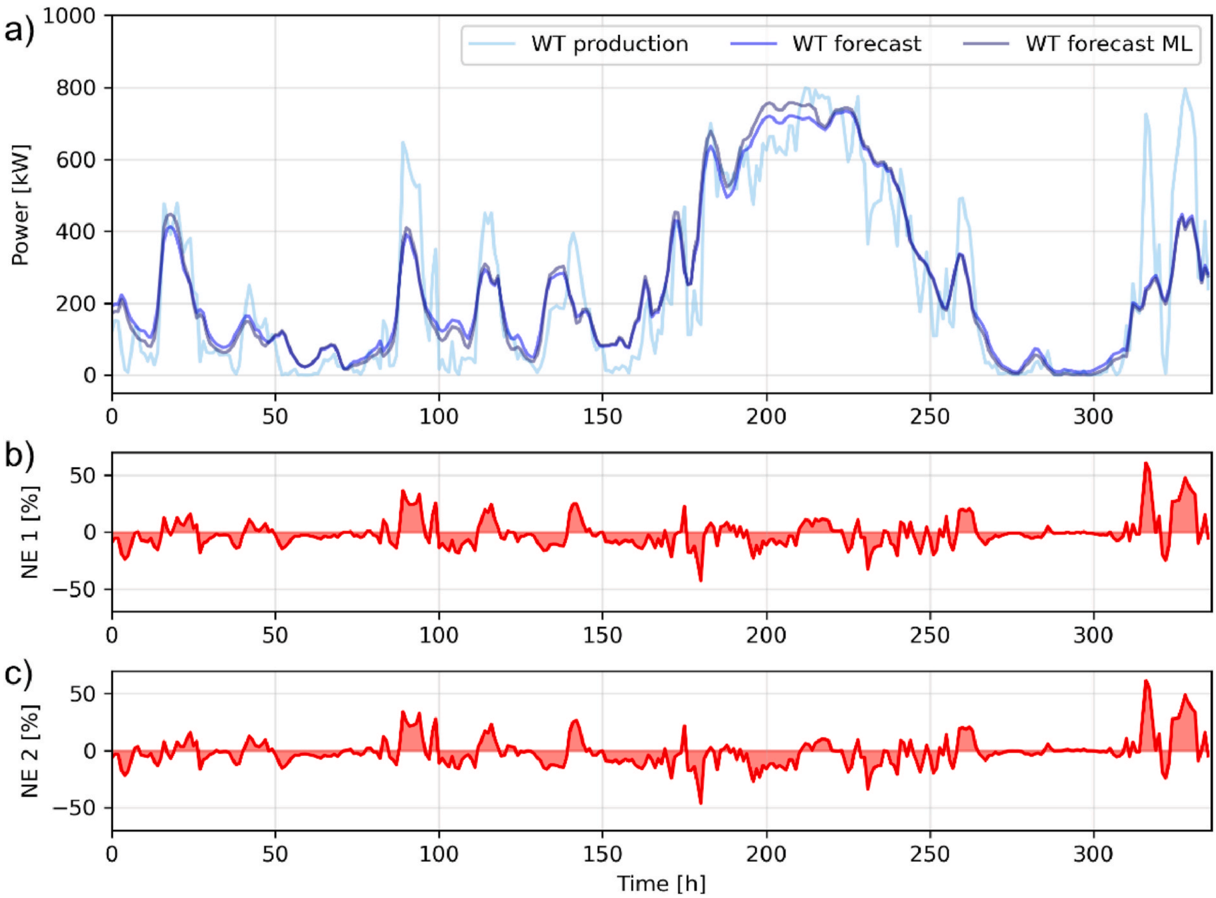

Fig. 29. Example of forecast correction of wind production during two weeks of operation. a) comparison between the original forecast, the improved forecast, and the actual production b) normalized error trend between the original forecast and the actual production c) normalized error trend between the improved forecast and the actual production.

Table 2
Comparison of the best results from different algorithms for the WT forecast enhancement.

Metric	Original	LR	SVR	NNs
Optimal number of training days	–	21	22	27
Optimal Hyperparameters	–	–	ε = 0.4 C = 34	Max iter. = 10k Topology = 30,30,30
Final NRMSE	14.6 %	14.6 %	14.6 %	14.6 %
Final NMAE	10.25 %	9.89 %	9.84 %	10.04 %
Max NE	81.14 %	81 %	85.7 %	82.7 %
Mean NE	−1.9 %	−0.24 %	0.77 %	−0.35 %
Median NE	−2.5 %	−0.8 %	0 %	−1.5 %

biased behaviors.

The box-and-whiskers plot in Fig. 31 directly compares the NE distribution of the original forecast with the predictions improved by the three algorithms. Even if the gain is much more limited than in the case of the application to solar forecast, the approach of the mean and the median NE to zero represents a benefit for several applications.

4. Discussion

This work proposed the application of three machine learning algorithms to reduce the error of commercial forecasts of solar and wind power production.

The study also analyzed the problem of overfitting, preventing the algorithms from failing to fit to additional data or predict future observations reliably. In this work, the models were not trained once on a fixed dataset but re-trained daily using a sliding window of the previous 30 days and applied to predict the next day. This approach can also be referred to as a "walk-forward strategy" and inherently limits overfitting by ensuring that future data are never used in training. Each prediction is made on completely unseen data, and the model cannot "memorize" the entire historical dataset. As an additional check, some standard overfitting prevention strategies were still tested. No improvement was achieved by implementing the model regularization via the alpha parameter (L2 penalty). Similarly, when using early stopping, the solver led to premature convergence and increased the error in this specific application, probably because of the limited training windows used.

The commercial solar forecasts analyzed in this study were characterized by considerable bias error, with a mean NE of −12.2 % and an error distribution with two peaks (Fig. 9). The application of ML proved to be able to drastically reduce this bias error, lowering the mean NE to −0.6 % on average. The NMAE was also improved from 8.6 % to 3.34 % by using NNs. The order of magnitude of these results is in line with findings from Böök and Lindfors [45], who managed to reduce the MAE from 0.08 to 0.06 W/Wp and bias from 0.05 to 0.01 W/Wp after the application of their site-specific adjustment of NWP-based PV forecasts. Results from this work show that it is possible to achieve a comparable improvement by directly correcting commercial forecasts with a standard ML algorithm available via open-source libraries, if properly tuned.

The error distribution of the ML corrected forecast is similar to what

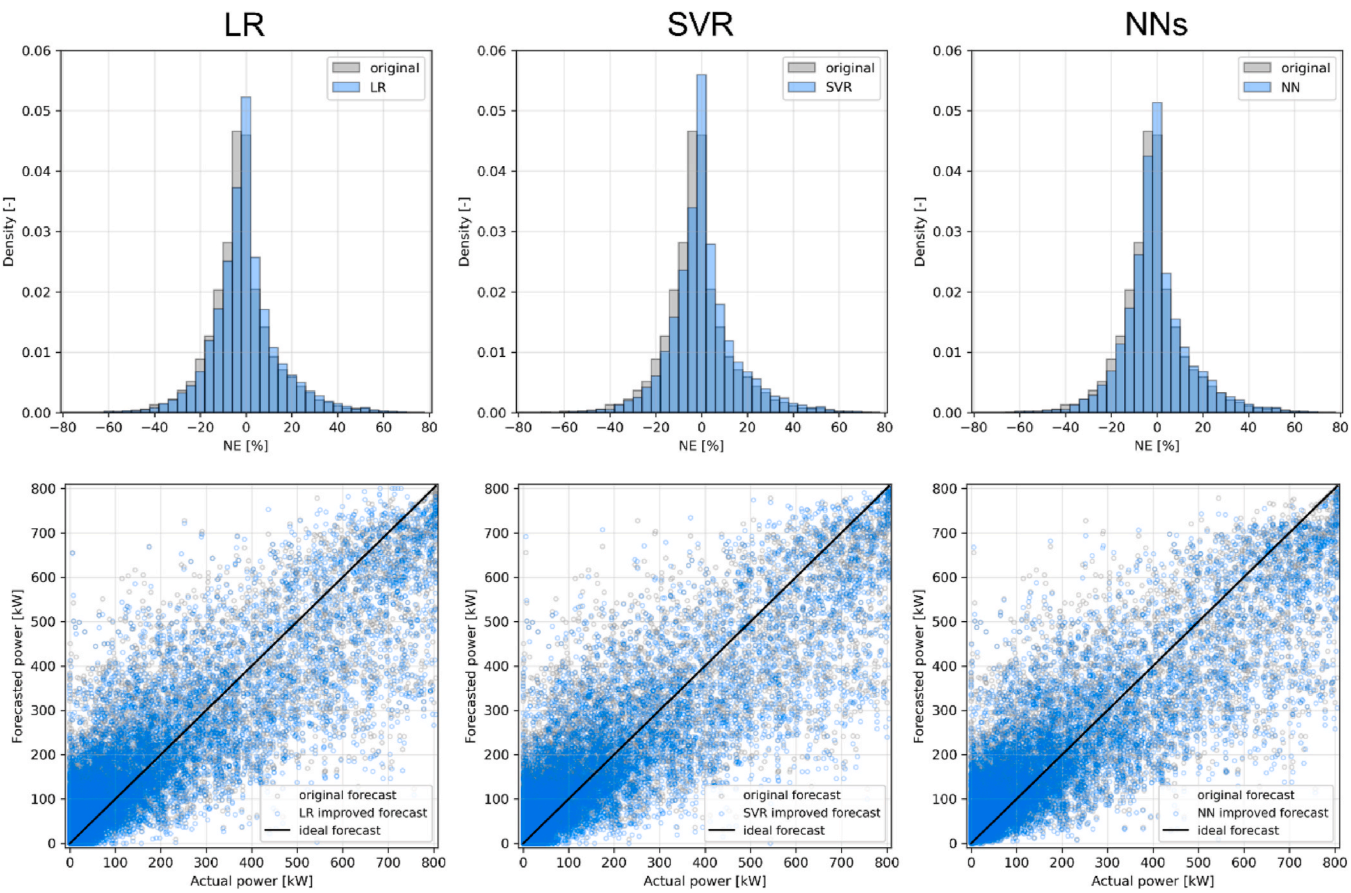

Fig. 30. Histograms of normalized error distribution and scatter of the forecasted power vs actual power production from the WT after the application of LR, SVR, and NNs.

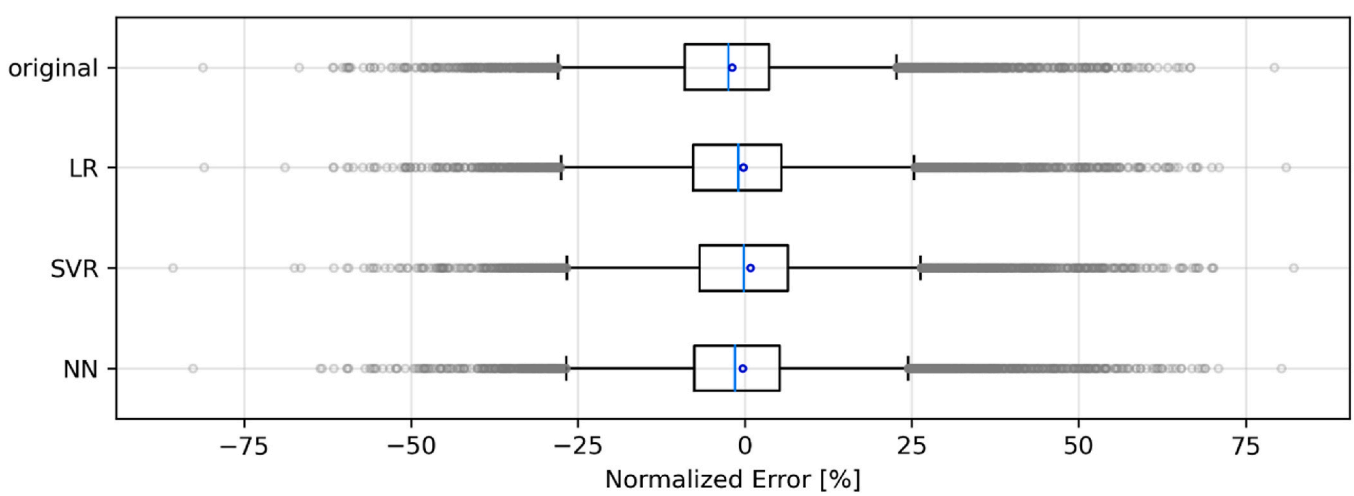

Fig. 31. Box-plot comparison of the normalized error characterizing the original forecast of the WT production and forecast improved by LR, SVR, and NNs.

was found by Yin et al. [21]. With respect to the non-iterative correction method proposed by the authors of the latter study, the walk-forward technique developed here achieved similar improvements in error metrics by correlating the commercial forecast with historical power production without requiring other inputs such as the irradiance, temperature, humidity, and wind speed.

The analysis of the RES production forecasts revealed that these quantities can often be biased, as found by Maciejowska et al. [43]. As found by the authors of the latter study, results from this study proved that regression models can improve the accuracy of predictions. However, the comparison between the three algorithms tested in this study proved that NNs outperformed regression models for PV production forecasting, in line with findings from studies performing solar predictions from weather data [24,27,30].

In line with findings from Wang et al. [34], the uncertainty of wind power forecasts is higher than solar power. In terms of NRMSE, the proposed methods were not able to improve the accuracy of the WT production forecast. Considering the other error metrics, the room for improvement of the WT forecast is much more contained than in the case of solar predictions.

The sensitivity varying the size of the time window used for the training of the methods has shown that, in most cases, correction algorithms performed better when trained on a smaller number of days, rather than the maximum available training period. The correlations identified by data-driven methods can indeed be influenced or biased by data points that are too distant in time. Methods performed better when working with more recent subsets of historical data, allowing the algorithm to capture relevant patterns without the noise deriving from less relevant information. On the other hand, using too short training periods resulted in biased forecasts after atypical days, showing the need to find the right trade-off between the two extremes. In general, training periods from 25 to 27 days performed best.

The statistical analysis of the residual NE of corrected forecasts has shown that the proposed methods were successful in eliminating bias errors: from −12.2 % to −0.57 % for solar using SVR and from −1.9 % to −0.24 % for wind using LR. In cases where a commercial forecast presents an offset from the actual performance of a generator (e.g., overlooked aspects of real production or poor calibration), ML algorithms have shown significant potential in greatly improving accuracy.

However, these methods were not able to reduce other types of errors. For example, while wind production forecasts can capture the average generation, they struggle to consider the actual oscillations of production. Even when trained on the discrepancy between historical production and previous predictions, ML algorithms did not improve this issue. This limitation likely arises because there is no inherent link between the predicted values and actual production, making it impossible to extract meaningful patterns from the data, even using complex methods. In fact, more advanced techniques, such as NNs, often performed worse than simpler regression models, possibly due to their attempt to model correlations that do not exist in the data.

The improvement that ML correction can bring to commercial forecasting suggests that policies promoting the integration of these tools can enhance the planning and scheduling phase of HESs or isolated networks. A more accurate forecast can indeed reduce the reliance of these systems on fossil-based power plants and support a more efficient use of battery storage, thus lowering operational costs and greenhouse gas emissions. The adoption of ML-based forecasting frameworks can also increase the resilience of power systems in remote areas, providing valuable guidance for decision-makers aiming to accelerate renewable energy penetration while maintaining system stability.

The ML improved forecast can drive the dispatch optimization of the HPS of the island of Tilos. This study proved that ML could reduce bias error and improve the accuracy of commercial forecasts. Despite this, the new predictions must be tested on a real application to assess if they could actually result in better decision-making for the storage device management and power allocation, ultimately leading to an actual improvement in the outcome of the system. The bias-corrected forecasts proved to be the ideal starting point for the generation of a set of possible realization scenarios of PV and WT production, the input for a robust optimization algorithm to improve the reliability of the HPS energy export, as suggested recently by the authors in Ref. [53].

This study applied ML methods to directly correct a commercial forecast, trying to identify patterns linking the biased prediction with the actual production from generators. This may represent a strength point for some applications, limiting the number of time series required for the application of the proposed correction method. On the other hand, the inclusion of other physical quantities in the training phase of the algorithms, such as the temperature, irradiance, wind speed, etc., may allow the algorithms to find other correlations that may enhance the accuracy of final predictions. In further studies, if historical predictions of other quantities related to RES production become available, new correction methods involving new inputs could be tested.

The discussion and analysis of results have provided some general insights on the application of ML algorithms for the correction of commercial forecasts, as the higher improvement margin for bias errors and the trade-off represented by a medium number of training days. However, the use of data related to a small, geographically specific location limits the direct generalizability of the proposed algorithms. With the availability of additional datasets related to different locations, further developments of the study could involve the application of the proposed techniques to new sites. The comparison of results from this study and from the application to other locations would provide valuable insights into the necessity of tuning the algorithms for each specific case study, considering each specific set of phenomena contributing to the forecasting error.

The aim of the study was specifically to correct day-ahead commercial forecasts from a commercial provider that the manager of an HPS acquires to optimize the operation of a RES-based system in the day-ahead electricity market. These forecasts are usually classified as "short-term forecasts", different from the medium- and long-term forecasts, which are instead useful for the system planning and design phases. Further studies could focus on the correction of longer-term forecasts to discuss whether the application of ML algorithms to a longer application dataset would change the size of the optimal training window and the set of optimal hyperparameters.

Due to data availability, the maximum length of the training dataset used for the walk-forward training validation of algorithms was thirty days. While results have shown that in most cases the optimal number of days for the training phase was lower than the maximum one, longer training windows may allow the algorithms to exploit seasonal trends. In further studies, if longer historical predictions become available, the algorithms could be tested with longer training windows. However, in such case, the algorithms would be more vulnerable to suffering from overfitting problems. To avoid that, advanced techniques as dropout or early stopping should be tested to improve the effectiveness of algorithms.

5. Conclusions

The present study evaluated and compared various data-driven approaches to enhance renewable energy production forecasts. The study specifically focused on forecasting solar and wind energy production, which was tested using the HPS of a small Mediterranean island as a case study. The HPS model is also gaining attention in regions with outdated electrical grids, where the integration of intermittent RES must comply with stringent technical constraints to ensure grid stability and reliability.

Every method tested in this work succeeded in improving the original forecast of the production capabilities of both the PV field and the WT that feed an actual HPS. The results indicate that data-driven models hold significant potential for improving RES production forecasts provided by an actual commercial provider. This application is particularly promising, as enhanced prediction accuracy can greatly benefit the scheduling process of HPSs.

Regarding the enhancement of solar production forecast, LR leads to good results despite its simplicity. An LR method trained on 30 days of previous operation can reduce the original NMAE from 8.6 % to 3.59 %. SVR was the best-performing approach when the available number of days for the algorithm training was limited. When trained on 25 days, the SVR can reduce the NMAE to 3.44 %. However, tailoring the algorithm for the minimization of the NMAE led to high values of C, the parameter that relaxes the tolerance of errors outside the allowed band. Consequently, SVR performs slightly worse than LR from the perspective of NRMSE, since this error metric weights the impact of outliers more.

Overall, the method that performed best in the solar forecast comprises the NNs algorithm. An NN composed of three layers, each containing 10 neurons, can reduce the NMAE to 3.34 % when trained on 25 days prior to the application. Every tested method was able to improve the PV commercial forecast both in terms of NMAE and NRMSE.

When methods were applied to enhance the accuracy of the prediction of the wind power production, the improvement was significantly lower. NNs were the worst-performing method when applied to the WT time series. The layout that performed best comprises three layers, each containing 30 neurons. This method could reduce the error only down to 10.04 % when trained on 27 days. LR performed slightly better, reducing the error to 9.89 % when trained on 21 days. SVR was the best-performing method and could reduce the error to 9.86 %.

CRediT authorship contribution statement

Francesco Superchi: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Antonis Moustakis:** Writing – review & editing, Supervision, Data curation. **George Pechlivanoglou:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Alessandro Bianchini:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Renewables 2024 – Analysis. IEA; Oct. 2024. https://www.iea.org/reports/renewables-2024. [Accessed 5 December 2024].

[2] Huynh CM, Phan TN. Climate change and income inequality: does renewable energy matter? Renew Energy Oct. 2024;233:121147. https://doi.org/10.1016/j.renene.2024.121147.

[3] Akan T. Can renewable energy mitigate the impacts of inflation and policy interest on climate change? Renew Energy Sep. 2023;214:255–89. https://doi.org/10.1016/j.renene.2023.05.023.

[4] Blum NU, Sryantoro Wakeling R, Schmidt TS. Rural electrification through village grids—Assessing the cost competitiveness of isolated renewable energy technologies in Indonesia. Renew Sustain Energy Rev Jun. 2013;22:482–96. https://doi.org/10.1016/j.rser.2013.01.049.

[5] Ahmad E, Khan D, Anser MK, Nassani AA, Hassan SA, Zaman K. The influence of grid connectivity, electricity pricing, policy-driven power incentives, and carbon emissions on renewable energy adoption: exploring key factors. Renew Energy Oct. 2024;232:121108. https://doi.org/10.1016/j.renene.2024.121108.

[6] Dai W, et al. Increase the integration of renewable energy using flexibility of source-network-load-storage in district cooling system. J Clean Prod Feb. 2024; 441:140682. https://doi.org/10.1016/j.jclepro.2024.140682.

[7] Alyami S. Fairness and usability analysis in renewable power curtailment: a microgrid network study using bankruptcy rules. Renew Energy Jun. 2024;227: 120424. https://doi.org/10.1016/j.renene.2024.120424.

[8] Denholm P, Mai T. Timescales of energy storage needed for reducing renewable energy curtailment. Renew Energy Jan. 2019;130:388–99. https://doi.org/10.1016/j.renene.2018.06.079.

[9] Rupiper AM, Good R, Miller GJ, Loge F. Mitigating renewables curtailment and carbon emissions in California through water sector demand flexibility. J Clean Prod Jan. 2024;437:140373. https://doi.org/10.1016/j.jclepro.2023.140373.

[10] Luhaniwal J, Puppala H, Agarwal S, Mathur T. Framework for strategic deployment of hybrid offshore solar and wind power plants: a case study of India. J Clean Prod Nov. 2024;479:144009. https://doi.org/10.1016/j.jclepro.2024.144009.

[11] Psarros GN, Papathanassiou SA. Internal dispatch for RES-Storage hybrid power stations in isolated grids. Renew Energy Mar. 2020;147:2141–50. https://doi.org/10.1016/j.renene.2019.10.001.

[12] Su Y, Wu Z, Liu J. Towards financial resilience: smart energy management in virtual power plants using stochastic–robust optimization. J Clean Prod Dec. 2024; 483:144238. https://doi.org/10.1016/j.jclepro.2024.144238.

[13] Jannati M, Foroutan E. Analysis of power allocation strategies in the smoothing of wind farm power fluctuations considering lifetime extension of BESS units. J Clean Prod Sep. 2020;266:122045. https://doi.org/10.1016/j.jclepro.2020.122045.

[14] Couto A, Estanqueiro A. Wind power plants hybridised with solar power: a generation forecast perspective. J Clean Prod Oct. 2023;423:138793. https://doi.org/10.1016/j.jclepro.2023.138793.

[15] Lorenti G, Mariuzzo I, Moraglio F, Repetto M. Heuristic optimization applied to ANN training for predicting renewable energy sources production. Compel - Int J Comput Math Electr Electron Eng Mar. 2022;41(6):2010–21. https://doi.org/10.1108/COMPEL-11-2021-0420.

[16] A Data-Driven Approach to Predict Hourly Load Profiles From Time-of-Use Electricity Bills | IEEE Journals & Magazine | IEEE Xplore." Accessed: December. 6, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10151868.

[17] Galli C, Superchi F, Bianchini A. Enhancing the prediction of electric load demand: a comparative analysis of ARIMA and ANN models for the case of a small touristic island. J Phys Conf Ser Nov. 2024;2893(1):012120. https://doi.org/10.1088/1742-6596/2893/1/012120.

[18] Ruhnau O, Hennig P, Madlener R. Economic implications of forecasting electricity generation from variable renewable energy sources. Renew Energy Dec. 2020;161: 1318–27. https://doi.org/10.1016/j.renene.2020.06.110.

[19] Chen Y, et al. Impacts of stochastic forecast errors of renewable energy generation and load demands on microgrid operation. Renew Energy Apr. 2019;133:442–61. https://doi.org/10.1016/j.renene.2018.09.110.

[20] Pierro M, et al. Progress in regional PV power forecasting: a sensitivity analysis on the Italian case study. Renew Energy Apr. 2022;189:983–96. https://doi.org/10.1016/j.renene.2022.03.041.

[21] Yin W, Han Y, Zhou H, Ma M, Li L, Zhu H. A novel non-iterative correction method for short-term photovoltaic power forecasting. Renew Energy Oct. 2020;159: 23–32. https://doi.org/10.1016/j.renene.2020.05.134.

[22] Ma Y, Lv Q, Zhang R, Zhang Y, Zhu H, Yin W. Short-term photovoltaic power forecasting method based on irradiance correction and error forecasting. Energy Rep Nov. 2021;7:5495–509. https://doi.org/10.1016/j.egyr.2021.08.167.

[23] Bright JM, Killinger S, Lingfors D, Engerer NA. Improved satellite-derived PV power nowcasting using real-time power data from reference PV systems. Sol Energy Jul. 2018;168:118–39. https://doi.org/10.1016/j.solener.2017.10.091.

[24] Sobri S, Koohi-Kamali S, Rahim N Abd. Solar photovoltaic generation forecasting methods: a review. Energy Convers Manag Jan. 2018;156:459–97. https://doi.org/10.1016/j.enconman.2017.11.019.

[25] Sharifzadeh M, Sikinioti-Lock A, Shah N. Machine-learning methods for integrated renewable power generation: a comparative study of artificial neural networks, support vector regression, and gaussian process regression. Renew Sustain Energy Rev Jul. 2019;108:513–38. https://doi.org/10.1016/j.rser.2019.03.040.

[26] Thaker J, Höller R. Hybrid model for intra-day probabilistic PV power forecast. Renew Energy Oct. 2024;232:121057. https://doi.org/10.1016/j.renene.2024.121057.

[27] Markovics D, Mayer MJ. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew Sustain Energy Rev Jun. 2022;161:112364. https://doi.org/10.1016/j.rser.2022.112364.

[28] Das UK, et al. SVR-based model to forecast PV power generation under different weather conditions. Energies Jul. 2017;10(7):7. https://doi.org/10.3390/en10070876.

[29] De Leone R, Pietrini M, Giovannelli A. Photovoltaic energy production forecast using support vector regression. Neural Comput Appl Nov. 2015;26(8):1955–62. https://doi.org/10.1007/s00521-015-1842-y.

[30] AlShafeey M, Csáki C. Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. Energy Rep Nov. 2021;7:7601–14. https://doi.org/10.1016/j.egyr.2021.10.125.

[31] Çakır S. Renewable energy generation forecasting in Turkey via intuitionistic fuzzy time series approach. Renew Energy Sep. 2023;214:194–200. https://doi.org/10.1016/j.renene.2023.05.132.

[32] Lindberg O, Zhu R, Widén J. Quantifying the value of probabilistic forecasts when trading renewable hybrid power parks in day-ahead markets: a nordic case study. Renew Energy Dec. 2024;237:121617. https://doi.org/10.1016/j.renene.2024.121617.

[33] Lee Y, Ha B, Hwangbo S. Generative model-based hybrid forecasting model for renewable electricity supply using long short-term memory networks: a case study of south Korea's energy transition policy. Renew Energy Nov. 2022;200:69–87. https://doi.org/10.1016/j.renene.2022.09.058.

[34] Wang H, Zhang N, Du E, Yan J, Han S, Liu Y. A comprehensive review for wind, solar, and electrical load forecasting methods. Glob Energy Interconnect Feb. 2022; 5(1):9–30. https://doi.org/10.1016/j.gloei.2022.04.002.

[35] Hanifi S, Liu X, Lin Z, Lotfian S. A critical review of wind power forecasting methods—past, present and future. Energies Jan. 2020;13(15):15. https://doi.org/10.3390/en13153764.

[36] Yang M, Shi C, Liu H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. Energy Mar. 2021;218:119515. https://doi.org/10.1016/j.energy.2020.119515.

[37] Zhang J, Meng H, Gu B, Li P. Research on short-term wind power combined forecasting and its Gaussian cloud uncertainty to support the integration of renewables and EVs. Renew Energy Jun. 2020;153:884–99. https://doi.org/10.1016/j.renene.2020.01.062.

[38] Bashir T, Wang H, Tahir M, Zhang Y. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models. Renew Energy Feb. 2025; 239:122055. https://doi.org/10.1016/j.renene.2024.122055.

[39] Nguyen THT, Phan QB. Hourly day ahead wind speed forecasting based on a hybrid model of EEMD, CNN-Bi-LSTM embedded with GA optimization. Energy Rep Nov. 2022;8:53–60. https://doi.org/10.1016/j.egyr.2022.05.110.

[40] Hong Y-Y, Satriani TRA. Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network. Energy Oct. 2020;209:118441. https://doi.org/10.1016/j.energy.2020.118441.

[41] Lin W-H, Wang P, Chao K-M, Lin H-C, Yang Z-Y, Lai Y-H. Wind power forecasting with deep learning networks: Time-series forecasting. Appl Sci Jan. 2021;11(21): 21. https://doi.org/10.3390/app112110335.

[42] Pappa A, Theodoropoulos I, Galmarini S, Kioutsioukis I. Analog versus multi-model ensemble forecasting: a comparison for renewable energy resources. Renew Energy Mar. 2023;205:563–73. https://doi.org/10.1016/j.renene.2023.01.030.

[43] Maciejowska K, Nitka W, Weron T. Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. Energy Econ Jul. 2021;99:105273. https://doi.org/10.1016/j.eneco.2021.105273.

[44] Kiesel R, Paraschiv F. Econometric analysis of 15-minute intraday electricity prices. Energy Econ May 2017;64:77–90. https://doi.org/10.1016/j.eneco.2017.03.002.

[45] Böök H, Lindfors AV. Site-specific adjustment of a NWP-based photovoltaic production forecast. Sol Energy Nov. 2020;211:779–88. https://doi.org/10.1016/j.solener.2020.10.024.

[46] Superchi F, Giovannini N, Moustakis A, Pechlivanoglou G, Bianchini A. Optimization of the power output scheduling of a renewables-based hybrid power station using MILP approach: the case of tilos island. Renew Energy Jan. 2024;220: 119685. https://doi.org/10.1016/j.renene.2023.119685.

[47] Superchi F, Schepers S, Moustakis A, Pechivanoglou G, Bianchini A. Towards the introduction of green hydrogen in the energy mix of mediterranean islands through the integration of wind and solar power: a techno-economic case study. In: Presented at the ECOS2023; Jun. 2023. p. 3361. https://doi.org/10.52202/069564-0301.

[48] Superchi F. In: 37th international conference on efficiency, cost, optimization, simulation and environmental impact of energy systems (ECOS 2024). ECOS; 2024. https://doi.org/10.52202/077185. 2024.

[49] Pedregosa F, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12(85):2825–30.

[50] Börjesson L, Singull M. Forecasting financial time series through causal and dilated convolutional neural networks. Entropy Sep. 2020;22. https://doi.org/10.3390/e22101094.

[51] Superchi F, Mariuzzo I, Repetto M. Machine learning strategies assessment for energy commodities forecasting. Politecnico di Torino; 2021 [Online]. Available: https://webthesis.biblio.polito.it/19992/. [Accessed 23 January 2024].

[52] Catalão JPS, Mariano SJPS, Mendes VMF, Ferreira LAFM. Short-term electricity prices forecasting in a competitive market: a neural network approach. Electr Power Syst Res Aug. 2007;77(10):1297–304. https://doi.org/10.1016/j.epsr.2006.09.022.

[53] Superchi F, Moustakis A, Pechlivanoglou G, Bianchini A. On the relevance of considering the uncertainty in renewables production forecasts to optimize hybrid power stations: a robust MILP approach. Energy Convers Manag Oct. 2025;341: 120078. https://doi.org/10.1016/j.enconman.2025.120078.