Journal of Cleaner Production 423 (2023) 138793

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Wind power plants hybridised with solar power: A generation forecast perspective

António Couto *, Ana Estanqueiro

LNEG - Laboratório Nacional de Energia e Geologia, Lisboa, Portugal

A R T I C L E I N F O

Handling Editor: Panos Seferlis

Keywords:
Power forecast
Hybrid power plant
Wind power plant hybridisation
Solar PV energy
Meteorological parameters

A B S T R A C T

The association of different variable renewable technologies in hybrid power plants and the benefits of their aggregation for the operation of power systems is an area of recent research. Accurate forecasts are crucial for extracting those benefits and promote an optimal integration of such plants into power systems and electricity markets. This study focuses on the hybridisation of existing wind power plants with different shares of solar photovoltaic capacity and investigates how these power plants can reduce their combined forecast errors and thus, increasing profitability in electricity markets.

The work uses a forecast methodology based on a sequential forward feature selection algorithm which employs two different objective functions and an artificial neural network approach previously presented but, in this case, it is applied to the specific case of hybrid power plants. The methodology uses as input data from a numerical weather prediction model and iteratively selects meteorological features to achieve the different objective functions implemented, namely i) minimisation of the root mean square error; or ii) maximisation of the market remuneration.

The methodology developed was applied to three case studies in Portugal with different levels of wind and solar generation complementarity. The results show that the hybrid power plants can increase market value by up to 5% and total remuneration can increase by up to 30% when compared with the existing wind power plant, while it is possible to reduce the forecast errors by nearly 4%. The obtained results highlight the need to select the most relevant meteorological features to maximise the accuracy of the power forecast and the renewable power producers revenues in a market environment.

1. Introduction

Sustainably integrating variable renewable energy sources (vRES) as wind and solar photovoltaic power into power systems is a significant challenge due to their intrinsic generation variability (Yang et al., 2021). Accurate forecasting of vRES production is necessary to minimise the use of carbon-intensive technologies and costly reserves and to achieve optimal dispatch schedules for the power system (Wang et al., 2022). As reliance on renewable energy grows, the accuracy of forecasting systems will become increasingly crucial (Ruhnau et al., 2020), (Strbac et al., 2021).

Forecasts are also instrumental in electricity markets to enable market players to participate in different market products. Notwithstanding that vRES players will always face challenges in these markets due to the unpredictable and complex behaviour of the atmosphere, especially since forecasting accuracy depends on the forecast horizon (Yan et al., 2022). The day-ahead market (DAM) is used in many electricity markets worldwide, typically requiring forecasts for a horizon between 12 and 36 h (Algarvio et al., 2019). For this horizon, forecast methods based on numerical weather prediction (NWP) models outperform other forecast methods such as satellite and all-sky imaging (applied for solar purposes) or autoregressive that are useful for short time horizons (up to 6 h head). Comprehensive reviews are available for wind (Yang et al., 2021) and solar photovoltaic (PV) power forecasting (Yang and van der Meer, 2021), (Ahmed et al., 2020).

In recent years NWP models have improved their capabilities to simulate the atmospheric circulation but still have errors due to, for instance, inadequate physics parameterisations, sub-grid-scale phenomena, and uncertainty in initial and boundary conditions (Yan et al., 2022). Downscaling techniques can mitigate some of these issues by using statistical or machine learning methods to establish relationships between local (as wind power) and large-scale (such as wind speed,

* Corresponding author.
E-mail addresses: antonio.couto@lneg.pt (A. Couto), ana.estanqueiro@lneg.pt (A. Estanqueiro).

https://doi.org/10.1016/j.jclepro.2023.138793
Received 6 June 2023; Received in revised form 19 August 2023; Accepted 9 September 2023
Available online 12 September 2023
0959-6526/

pressure, and atmospheric stability fields from a NWP) variables and provide location-specific forecasts (Ahmed et al., 2020).

Recently, researchers have focused on advanced power forecast approaches that involve machine learning and deep learning (Wang et al., 2019), ensemble (Alessandrini et al., 2015), and hybrid (Eseye et al., 2018) models. These models are complex and require extensive hyperparameter tuning (Markovics and Mayer, 2022), which makes them difficult to interpret and adopt. Additionally, some of these models typically exhibit only a marginal improvement in performance and are often tested on a single case study, making it unclear whether the observed improvements are universally applicable or represent local effects (Verbois et al., 2022). Thus, the accuracy of wind speed forecasts is significantly influenced by the local climate conditions and, as a result, the specific location under consideration (Visser et al., 2022).

Recent studies have shown that the use of adequate meteorological input parameters from NWP data can improve the accuracy of wind and solar power forecasts. By comparing several machine learning methods for solar power, Markovics and Mayer (2022) identified that the selection of predictors is more crucial than the choice of models in achieving accurate results. Salcedo-Sanz et al. (2018) found a 20% improvement in wind power forecasting by using a feature selection (FS) approach, while Andrade and Bessa (2017) recommend effective feature extraction from raw NWP data to enhance wind and solar PV forecasting accuracy. Couto and Estanqueiro (2022) have demonstrated that selecting suitable meteorological data can reduce errors in wind power forecasts by 13–37%, and Castangia et al. (2021) showed that cloud cover and relative humidity are important meteorological features for improving solar PV forecasts.

The concept of (utility) hybrid power plants (HPPs), which allows the hybridisation of existing wind power plants, was recently introduced in several countries, bringing attention to the exploitation of wind and solar PV complementarity (Jurasz et al., 2020a). HPP refers to a power plant that utilizes two or more renewable technologies with or without an energy storage system, to *generate* electric power (Couto and Estanqueiro, 2021), (WindEurope, 2019). This power is then injected into the grid at a single point of common coupling which is the same interconnection grid busbar, regardless of the number of renewable technologies or groups being used, and up to the maximum capacity allowed by its distributed generation permits. Thus, HPPs or the hybridisation of existing wind and solar PV power plants can have benefits such as i) shared and synergetic use of electric infrastructure, ii) a combined vRES generation with lower variability when compared with a single technology; iii) a reduction of events with low vRES contribution; and iv) bringing more value to investors and to the electricity system, such as through the possibility of providing ancillary services (Dykes et al., 2020). A review of HP projects currently in operation is available in (WindEurope, 2019), (Klonari et al., 2019; Lindberg et al., 2021; Babaremu et al., 2022). In Portugal, the first wind-solar HPP began operating at the end of 2022. This HPP combines an existing wind power plant, which has eight wind turbines and an installed capacity of 11 MW, with the integration of an additional 8.4 MW of solar PV capacity (EDP, 2023). Despite the focus on wind and solar PV power in this work, the combination of different renewable technologies shows that can bring value for producers and the power system, for instance, by combining wind and hydro power (Karadö et al., 2021), solar and hydro power (Jurasz et al., 2020b), solar and biomass (É et al., 2021), or combine multiple sources as wind, solar, hydro and geothermal (Qiu et al., 2022).

Notwithstanding the benefits identified in the literature on HPPs, the specific literature regarding the power forecasts of utility HPPs is still reduced (Lindberg et al., 2023). Alessandrini and McCandless (2020) used individual forecasts of wind and solar PV and proposed a data assimilation technique to improve the forecasting accuracy within an HPP. The forecast approach consists of an analogue ensemble technique using limited meteorological predictors such as wind speed and direction, global horizontal irradiance, air temperature, and cloud cover. Pombo et al. (2022) focused on the most adequate statistical approach to obtain individual near-real-time (lead horizons of 5 and 60 min) forecasts for wind and solar PV in a hybrid power plant in Denmark. In addition to the common basic inputs (power generation, wind speed), physics-informed parameters were also computed based on thermodynamics and astronomy. The results obtained suggest that different statistical approaches need to be used depending on the forecast horizon. Authors also concluded that a random decision forest approach is the most appropriate for 5-min forecasts, while an artificial neuronal network is the most adequate for lower-resolution problems. Lindberg et al. (2023) studied the effect of aggregation at a HPP with wind and solar technology in Sweden and assessed the accuracy and value of probabilistic power forecasts through participation of this type of power plant in an electricity market environment (day-ahead and balancing markets). Different aggregation ratios of wind and solar PV within a HPP were analysed through their combined variability and forecast performance. The authors concluded that a scenario with a ratio between 50 and 60% of wind power is the one that provides the best forecast results. Although not focused on power forecast to the specific case of HPP, other authors have addressed the benefit of combining wind and solar PV at different spatial scales, *e.g.,* for reducing the forecast errors (Zhang et al., 2015) and in the context of smart grids (Qadir et al., 2021).

1.1. Research gap

Being utility-scale HPP a recent concept and area of research, the power forecasts for these power plants were only addressed in few articles with different scopes as previously discussed. In specific, none of the previous work focused on the identification of the relevant meteorological parameters needed to feed the forecast approaches. In addition, the benefits of wind and solar PV complementarity for improving the power forecasts were only analysed for one specific wind and solar PV hybrid power plant without discussing the impact of different levels of complementarity, as observed in different regions of Portugal (Couto and Estanqueiro, 2021). Most of the works used individual forecasts for each technology, neglecting the potential benefit of the combined complementary primary resources to improve the accuracy of the HPP aggregated forecast.

1.2. Novel contributions

Using the concept of HPP without storage solutions, this work focuses on the forecast for hybrid power plants (existing wind power plants combined with newly installed PV capacity) in Portugal and intends to address some previously identified research gaps. In specific, it uses the forecast methodology presented in Couto and Estanqueiro (2022) for the specific case of HPP. The methodology is applied to three specific locations with different levels of wind and solar PV generation complementarity. The forecasting methodology used a sequential forward FS algorithm to select, at each iteration, the meteorological parameters that minimise a predefined objective function. While in (Couto and Estanqueiro, 2022) the root mean square error (RMSE) was the only function analysed, here an additional objective function is added to the methodology aiming to investigate the potential increase in the profitability of the producer in an electricity market environment. First, the forecast methodology is applied to forecast wind power. Afterwards, the forecast is applied to the combined generation of wind and solar PV assuming an HPP design. Similarly to the approach in Lindberg et al. (2023), two different solar PV capacity scenarios within the HPP are analysed by overplanting the existing wind park with additional 5 and 10 MW of solar PV capacity. Despite the increase in the installed capacity, the HPP maximum injected capacity is limited to 20 MW.

Compared to the traditional approach of forecasting and bidding in the electricity markets for individual renewable technologies, this work also contributes to the understanding of whether: i) the combined forecast errors can be reduced in a HPP when compared with the individual ones; and ii) the profitability in the electricity market can increase in an HPP, when compared to an existing wind power plant. This is particularly relevant for bringing added value for the (wind) power producers and potentially for the power system. As a side effect, this research also provides insights regarding most of the relevant meteorological parameters that should be used in an HPP power forecast. To increase the interpretation of the results obtained, a feature importance technique is applied to the meteorological parameters identified with the FS algorithm.

It should be noted this work focuses on hybridisation of already existing wind power plants with newly installed PV solar capacity, since the potential for this solution is not only significant in Portugal (as shown by Couto and Estanqueiro (2021)). Due to the complementarity levels (see (Kapica et al., 2021), (Jerez et al., 2023)) the hybridisation can also have a high potential in several other counties where the vRES deployment started with wind plants and moved to solar PV in recent years. Hybridization enables to increase the share of vRES in the near future discarding the need to wait for complex and expensive expansion plans for new transmission lines to be designed, approved and implemented and, in overall, reducing the unitary investment costs of the plants.

2. Power forecasts and electricity markets

The day-ahead markets (DAM) with hourly resolution are used in numerous electricity markets worldwide as the primary platform for trading electric energy. These markets need forecasts/bids for 24-h periods for day *D,* and those bids need to be submitted on day *D-1*, usually until 11:00 a.m. or 12:00 a.m. depending on the market (Algarvio et al., 2019). Hence, power producers need to obtain the forecast for a time horizon between 12 and 36 h ahead of the actual conversion of the kinetic energy of the wind. Due to the high anticipation, wind power forecast errors are not neglectable and those have a high impact on the revenues of wind power producers in markets. It is common to characterize the power forecasts performance using technical metrics such as bias error, mean absolute error, and RMSE. Those metrics are also used to optimize the power forecast systems and evaluate their accuracy (Shirkhorshidi et al., 2015). RMSE is often recommended since it considers amplitude and phase errors (Yang et al., 2020), (Lange and Focken, 2006). However, electricity market players such as vRES power producers may be more concerned about the economic performance of those models as highlighted by several authors, *e.g.*, (Visser et al., 2022), (Antonanzas et al., 2020). In order to incorporate economic performance considerations, it is important to account for the financial penalty imposed on a market participant when there is a deviation between the day-ahead bid and the actual power production (Visser et al., 2022).

One may conclude that accurate predictions not only support power system operators dispatch decisions, but also help maximising vRES producers' profits by reducing imbalance penalties. It should be noted, however, vRES power forecasts with no errors are impossible to achieve based on the existing DAM designs, and not all markets penalise producers in the same way for the deviations between the energy bided and produced. Most markets penalise deviations using either single or dual pricing systems where the penalty is proportional to the deviation from the bidding/schedule (Ahmed et al., 2020), (Ntomaris et al., 2022). Typically, the power producer that produces more than the energy scheduled in the DAM receives the downward price. Under this condition, the system buys the energy produced in excess (with respect to the bidding offer) at a price below the DAM, leading to a reduction in the potential profit. However, if the energy produced is below the scheduled, power producers need to compensate the system because it was necessary to use more expensive generators to meet the undersupply energy.

As several authors have highlighted, *e.g.*, (Ruhnau et al., 2020), the relation between the vRES power forecast accuracy using metrics such as RMSE and the economic benefits is not straightforward since the energy used in balancing markets changes every hour. Large forecast errors when imbalance penalty costs are small might represent a small economic loss. On the other hand, a small forecast error when imbalance penalty costs are significant can lead to high market losses. For the Iberian electricity market (MIBEL), Bessa et al. (2011) showed that a Spanish wind power producer can economically benefit from biased forecasts, which may be in conflict with the perspective of the transmission system operator. In the case of solar power, Antonanzas et al. (2017) identified that a tendency to forecasts with lower RMSE also enables obtaining the highest economic profit following almost a linear relationship. For their case study, Antonanzas et al. identified an improvement of 1 kWh in the RMSE leads to a profit increase of 11.66 €. The authors also noted that minor differences in the RMSE can represent large economic variations and *vice-versa*. Therefore, it is crucial to optimize forecasting models, considering their intended application and how it translates into value for the vRES participants in the electricity markets, as highlighted by the previous discussion regarding the relationship between power forecast technical and economic metrics.

3. Methodology to obtain deterministic power forecasts

This study extended the methodology proposed by Couto and Estanqueiro (2022) by applying it to the specific case of HPP. In this work, an additional objective function to identify the meteorological features aiming to maximise the remuneration of the producers was incorporated. In this work, a price-taker approach was assumed, meaning that individual market participants are small when compared to the overall market, and as a result, they have no capacity to influence the market prices.

Fig. 1 presents a flowchart that outlines the main steps of the proposed methodology applied for power forecasts. Broadly speaking, the methodology starts by obtaining forecast data from a NWP model, as well as observed data for wind and solar PV production. Another relevant dataset obtained is the MIBEL data, specifically, the hourly data from the DAM and the balancing markets, as discussed in the previous section. The methodology starts with the application of principal component analysis (PCA) to each meteorological parameter used. Thus, PCA is used as a data reduction technique. Based on the PCA results, the correlation distance between the observed data and the different principal components (PCs) is determined to rank how the various PCs enter the FS algorithm. The PCs are sorted from the highest correlation value to the lowest. The FS used is a sequential forward feature selection (SFFS) algorithm, therefore, the algorithm starts with only one feature, in this case, the PC with the highest correlation. Then, each PC is tested to assess whether it enables to improve the performance of the analysed objective function. If improvement occurs, the PC is considered relevant and therefore retained; otherwise, the PC is discarded. This procedure is applied to the different case studies, scenarios and objective functions used in this work enabling to identify the most relevant meteorological parameters for use in the power forecast for wind and hybrid power plants. In the following section, a more detailed description of each step is provided. A brief description of each step is provided in the following sections. A detailed description can be found in (Couto and Estanqueiro, 2022).

3.1. Principal component analysis

A principal component analysis (PCA) is used in this work as a dimensionality reduction method, by transforming a large set of variables into a small number of features that contain most of the variance of the original data (Liu and Chen, 2019). Thus, this technique is a common procedure that enables to identify the main dominant spatial-temporal synoptic variability modes by cancelling poorly correlated local effects that exist in the principal components (PCs) that explain a smaller percentage of variance (Couto et al., 2015), (Zhang et al., 2018). Consequently, the PCA makes it possible to reduce the dataset dimension (Liu and Chen, 2019), (Garcí et al., 2022). Mathematically, a

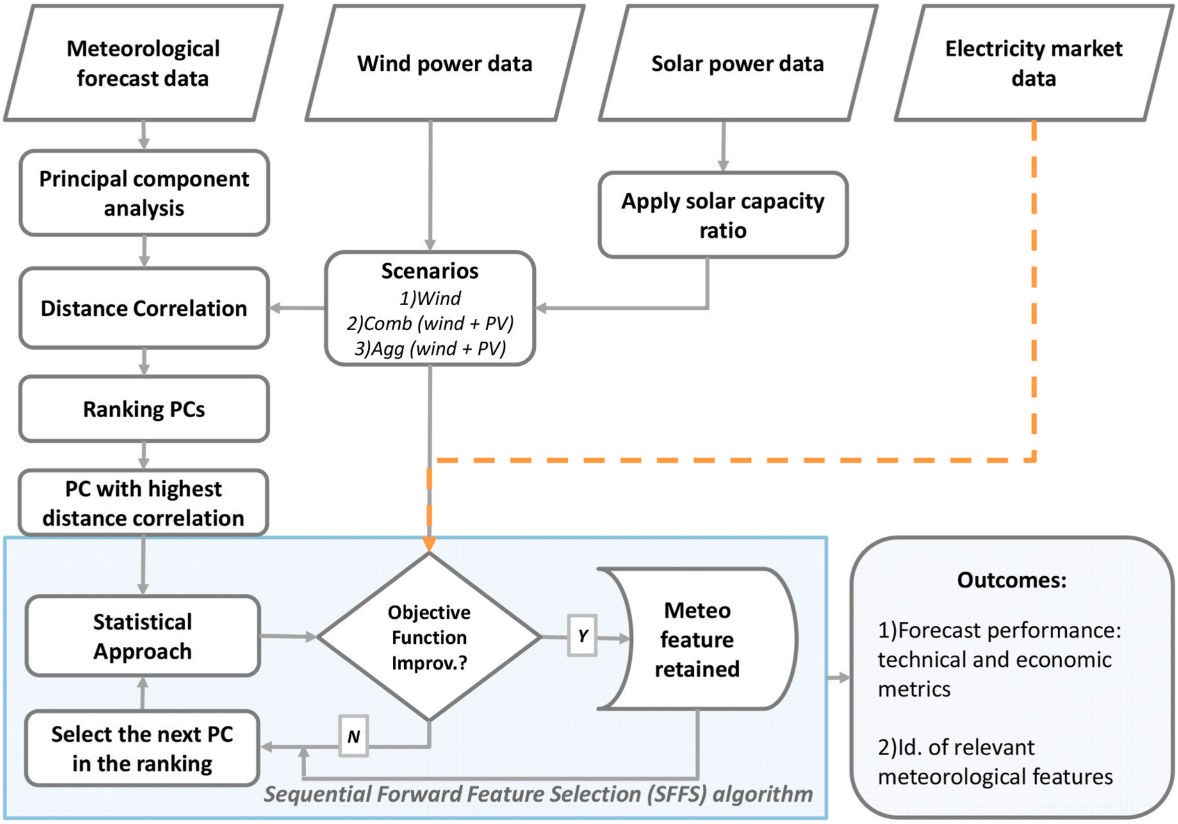

Fig. 1. Flowchart with the main steps used to generate power forecasts (adapted from (Couto and Estanqueiro, 2022)). The sequential forward feature selection algorithm is depicted within the blue box. The orange dotted arrow is utilized based on the chosen objective function. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

meteorological variable defined by j-*th* grid spatial points with a temporal dependence can be expressed as a linear combination of spatial patterns and time-dependent weights (1).

$$Z = \sum_{j=1}^{N} a_{ij} e_j \tag{1}$$

where Z represents the meteorological variable, a_{ti} represents the PC scores, and e_i represents the eigenvector of the covariance matrix. The equality is usually possible for N equal to the number of spatial points. The PCs are formed through linear combinations of observed variables, ordered based on a criterion of information explained, specifically, their variance (Liu and Chen, 2019). Thus, the first PC captures the largest amount of variance in the data. Subsequently, the second PC captures the maximum remaining variance that is uncorrelated with the first PC. The same occurs for the remaining PCs. In this work, PCA was applied individually for each meteorological parameter, and the number of retained PCs was determined according to the amount of variance they explain. Therefore, the criterion employed was to select the number of principal components (PCs) that accounted for 90% of the total variance (Dav et al., 2016). After the PCs for each meteorological parameter were obtained, the distance correlation (Szé et al., 2009) between all the PCs available and the power forecast was computed to rank the PCs for iteratively feed the forward feature selection algorithm. The PCA was applied for each meteorological variable after a z-score normalisation process. In this work, the PCA is computed using the Matlab software and the covariance matrix of Z.

3.2. *Feature selection approach*

FS methods are commonly used to achieve efficient data reduction in data pre-processing and to develop accurate data models (Lv and Wang, 2023). FS methods enable to reduce overfitting and make the model more interpretable. These algorithms can be classified into four main categories: filters, wrappers, embedded and hybrid methods (Solorio-Ferná et al., 2020).

In this work, a wrapper technique was used to identify the most pertinent meteorological features. This technique involves the evaluation of feature subsets based on the performance of an algorithm using a predefined metric. The evaluation is repeated for each subset. Variables that improve the model's accuracy are retained while the rest are discarded to construct the final model.

The sequential forward feature selection (SFFS) algorithm was chosen as a wrapper technique (Chandrashekar and Sahin, 2014). In each step, SFFS uses a greedy algorithm to search for the optimal subset of variables that minimises prediction errors. This algorithm begins with only one meteorological feature: the one with highest correlation value between the power observed and all the PCs available. At each iteration, the algorithm extends the previous set of meteorological features if the added feature enables to improve the objective function chosen; otherwise, the feature is discarded. An artificial neural network technique (see section 3.3) is then used to obtain the forecasts with different subsets of meteorological parameters. After obtaining the forecast, the next step is to verify if the objective function has been achieved. This verification process is repeated iteratively for all input variables, i.e., all PCs.

To ensure the robustness of the forecast results the methodology applied in (Couto and Estanqueiro, 2022) was used. Thus, a ten-fold cross-validation technique was implemented, dividing the calibration period into training and testing datasets –calibration dataset. Then, the validation of the methodology is performed using 6 months of data (as described in detail in section 5). For the final forecast results used in section 5 with the validation dataset, only the meteorological features that were identified in eight or more time in the SFFS are used.

3.3. Artificial neural networks: a statistical approach to obtain power forecasts

Artificial neural networks (ANNs) are a machine learning (ML) method widely applied to perform wind and solar power forecasts (Ruhnau et al., 2020), (Mellit and Kalogirou, 2008). ANNs are designed to recognise patterns and make predictions based on input data and can build complex non-linear relationships between the predictor and target variables. In contrast to deep learning techniques like recurrent neural networks or convolutional neural networks, ANNs have a generalization capacity being: i) robust to the presence of noise in the dataset, and ii) capable of identify dynamic patterns such as seasonality and trends within time-series datasets (Qiu et al., 2022).

The basic structure of an ANN is composed of three layers: input, hidden, and output. The input layer receives the data. The hidden layer processes the input data through a series of mathematical operations, which involve the use of weights and biases to transform the data into a more useful form. The output layer produces the final outcome, which can be a prediction, or a classification result based on the input data. The number of neurons in each layer and the number of layers in the network are important factors in determining the performance of the network (Bochenek et al., 2021). In addition of the using *rule of thumb* (Heaton, 2008), techniques such as neural architecture search, which automates the design of ANNs can support the identification of the most adequate architecture of ANNs since large networks with several layers and neurons can potentially learn more complex patterns, but they also require more computational resources and are more prone to overfitting, and potentially perform poorly on new data (Elsken et al., 2018).

ANNs require training to learn patterns and make accurate predictions. During training, the network is presented with paired input and output data. The network adjusts its weights and biases to minimise the difference between the predicted output and the target output (Bochenek et al., 2021), (Nazar et al., 2020). This process can be performed using backpropagation algorithms such as Levenberg-Marquardt (LM), and it is a crucial part of training ANNs (Papageorgiou and Poczęta, 2017). The LM algorithm is especially valuable for nonlinear optimisation problems where a balance of speed, stability, and accuracy is crucial. Additionally, the LM algorithm is highly suitable for optimisation problems with many parameters because it can efficiently handle high-dimensional optimisation tasks.

The architecture of the ANN adopted in this work is identical to that used by Couto and Estanqueiro (2022). Therefore, the number of hidden layers is set to one. For the first iteration of the SFFS, the number of features available is one, thus, the number of neurons in this layer is two. Afterwards, two thirds the size of the input layer plus the size of the output layer with rounding up as needed were applied to define the number of neurons in the hidden layer. The transfer function is a sigmoid function in the input layer to the hidden layer, and a linear function from the hidden layer to the output layer. Using the LM algorithm was used for training the ANN and the network was trained until one of three criteria was fulfilled: that the maximum number of iterations (1, 000) was reached, the minimum gradient (1e-7) was achieved, or the network performance measured by the RMSE reached the minimum of 0. This methodology was implemented using the neural network toolbox available in MATLAB software (Matlab, 2012).

3.4. Feature importance to increase the explainability of the results

The ANN used in this study is often referred to as a “black-box model” due to its low capability to provide directly significant insights regarding the relevance of each parameter being only evaluated according to the performance across various metrics (Molnar et al. et al., 2022). Although, the explainability of ANN results can, for instance, provide useful information for understanding the role of the meteorological features identified with the FS algorithms. Feature importance techniques can enable overcoming this limitation of the black-box models by quantifying the contribution of each feature within the context of a particular model (Molnar et al. et al., 2022), (Carriere and Kariniotakis, 2019)

In the literature, several explainability techniques are available depending on the ML model used (Carriere and Kariniotakis, 2019). In this work, a permutation feature importance technique is employed for feature importance evaluation. The main advantages of this technique are: i) simple to apply and ii) applicable after training a model, and iii) agnostic nature enabling to apply to a wide array of ML models (Carriere and Kariniotakis, 2019). The fundamental principle underlying this technique is to observe the alterations in model performance resulting from shuffling the values of a feature. In this work, the application of this technique involves the following steps: 1) identification of all pertinent features using the SFFS algorithm, 2) compute the performance of the model with all relevant features, 3) randomly permuting the values of a single feature and maintaining the remaining features unchanged and evaluating the performance of the model, 4) repeat the previous step for all relevant features; 5) compute the importance of each feature according to equation (2).

$$\varepsilon(\%) = \left(\frac{Performance_{AllRelevant}}{Performance_{Permuted}} - 1\right) \times 100 \tag{2}$$

In equation (2), ε represents the feature importance score, $Performance_{AllRelevant}$ and $Performance_{Permuted}$ represent the performance of the model obtained using all the relevant features, and for each feature permuted, respectively. In this work, the performance is assessed using the RMSE. A higher value of ε indicates a higher relevant feature. Steps *2)* to *5)* are repeated a total of 100 times for each individual feature, and the resulting importance score is determined by averaging these repetitions.

3.5. Metrics to assess the power forecast performance

The technical metrics used to assess the power forecast performance were the RMSE (see equation (3)), which is one of the most recommended metrics (Yang et al., 2021), (Yang et al., 2020) and the normalized RMSE (NRMSE - equation (4)). In this work, the NRMSE is given by the RMSE divided by the average observed power production (*AveragePower*) of each scenario (Schyska et al., 2017). The optimal value of RMSE and NRMSE is equal to zero. For economic parameters, the total remuneration (Equation (5)) and the market value (Equation (6)) were used.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}\left[P_{For.}(t) - P_{Obs.}(t)\right]^2}{T}} \tag{3}$$

$$NRMSE = \frac{RMSE}{AveragePower} \tag{4}$$

$$Renumeration = \sum_{t=0}^{T} P_{For.}(t)Price_{DAM}(t) + \begin{cases} [P_{Obs.}(t) - P_{For.}(t)]Price_{Up}(t) & for\ P_{For.}(t) < P_{Obs.}(t) \\ [P_{For.}(t) - P_{Obs.}(t)]Price_{Down}(t) & for\ P_{For.}(t) > P_{Obs.}(t) \end{cases} \tag{5}$$

$$Marketvalue = \frac{Renumeration}{\sum_{t=1}^{T} P_{Obs.}(t)} \tag{6}$$

In the previous equations, $Price_{DAM}$ is the DAM price at t-*th* hourly time step; $Price_{Up}$ and $Price_{Down}$ are the upward and downward deviation costs, respectively; and $P_{Obs.}$ and $P_{For.}$ are the observed and forecast power, respectively. The term $P_{For.}(t)Price_{DAM}(t)$ represents the remuneration obtained through the participation in DAM, while the second parcel represents the remuneration paid or received in balancing markets. For these economic metrics, the results will be only presented in comparison with the reference scenario results to quantify the benefits of the HPPs.

4. Input data: wind and solar power complementarity

For this work data covering the period from 1 January 2017 to 31 December 2019 were used. Two main sources of data were used, namely 1) NWP data from a mesoscale model and ii) observed data from three different wind parks and solar parks. The locations of solar power plants are within 5 km of the associated wind park, and it is assumed that they are affected by similar weather conditions and share the same busbar. Historical data from the MIBEL were used to compute the economic metrics (OMIE, 2022), (REN, 2022).

4.1. Forecast data: numerical weather prediction data

Today, NWP models are the core of weather forecasting and have evolved substantially in past years due to growing knowledge regarding the physical processes that govern the atmospheric dynamics and circulation as well as advancements in computational capabilities. These models are now less simplistic and incorporate more detailed physical parameterisations. They have also benefited from more efficient and representative data acquisition and assimilation systems around the globe. There are two primary categories of NWP models: global, which covers the entire planet, and regional/mesoscale, which is also called limited area models (Sweeney et al., 2020). The latter is usually fine-tuned using physical parameterisation for specific areas with high spatial and temporal resolution, resulting in reduced forecast errors.

This work utilized data provided by MeteoGalicia (2021) using the weather research and forecasting (WRF) model to obtain various historical meteorological forecast parameters. MeteoGalicia offers freely available data with spatial grid resolutions of 36, 12, and 4 km. This study used data from the 12-km domain because the 4-km domain does not cover all the locations analysed. The domain of simulation comprised longitudes from −21.33° to 6.16° and latitudes from 33.78° to 49.46° with hourly resolution and a 48-h forecast horizon. However, this study examined only the last 24 h, covering the DAM negotiation period in the MIBEL.

The study tested several outputs of the WRF model to identify meteorological characteristics that can improve the power forecast accuracy for HPPs. The parameters used in this work are listed in Table 1. From this table is possible to observe that some of selected parameters are not primarily focused on their influence on solar irradiance or wind speed but due to their impact on PV or wind power production. For instance, air temperature can help to include the effect from the PV module temperature that have a significant impact in the efficiency of the panels, while the wind speed at the surface are enable to include the convective heat transfer effect from PV panels caused by wind (Heusinger et al., 2020). In the table is also possible to identify meteorologically based parameters that influence the variability of wind speed and the process of wind power conversion that can be computed using the NWP output (Couto and Estanqueiro, 2022). These variables include the vertical temperature variation ($T_{var.}$), which is especially relevant for coastal zones that are influenced by the atmosphere-ocean interface and wind energy density.

In Table 1, sigma levels 0.994 and 0.986 correspond to nearly 50 m and 110 m above ground level, respectively. The parameter vertical temperature variation (in Kelvins) was computed based on data from the temperature at 2 m above the ground and an 850 hPa level according to equation (7).

$$T_{var.} = T_{850} - T_2 \tag{7}$$

Table 1
Meteorological parameters (and acronyms) used in this work. The parameters computed based on NWP are in italic. Adapted from (Couto and Estanqueiro, 2022).

Information obtained	Meteorological Parameter	Acronym	Units
Single level	Atmospheric visibility	Vis	M
	Cloud area fraction in high atmosphere layer	CFH	[0 1]
	Cloud area fraction in medium atmosphere layer	CFM	[0 1]
	Cloud area fraction in low atmosphere layer	CFL	[0 1]
	Cloud cover at low and medium layers	CFT	[0 1]
	Convective available potential energy	CAPE	J kg-1
	Convective inhibition	CIN	J kg-1
	Precipitation convective	$Prec_{Conv.}$	kg m-2
	Precipitation	Prec	kg m-2
	Humidity relative at 2 m	HR	[0 1]
	Mean sea level pressure	MSLP	Pa
	Planetary boundary layer height	PBL	m
	Surface downwelling longwave flux	LWF	W m-2
	Surface downwelling shortwave flux	SWF	W m-2
	Surface downward latent heat flux	LHF	W m-2
	Surface downward sensible heat flux	SHF	W m-2
	Surface wind speed gust	Gust	m/s
	Sea surface temperature	SST	K
	Mean sea level pressure gradient (Couto and Estanqueiro, 2022)	$MSLP_{Grad.}$	Pa m-1
	Wind shear (Couto and Estanqueiro, 2022)	WindShear	Dimensionless
	Wind power density at sigma levels 0.986	WPD	W m-2
	Vertical temperature variation	$T_{var.}$	K
Multiple level	Air temperature at 500, 850, and 2 m	T_{500},T_{850}, T_2	K
	Geopotential height at 500, 850 and sigma level 0.986	HGT_{500}, HGT_{850}, $HGT_{\sigma 0.986}$	m
	U-wind component at sigma levels 0.994 and 0.986, and 10 m	$U_{\sigma 0.986}$, $U_{\sigma 0.994}$, U_{10}	m s-1
	V-wind component at sigma levels 0.994 and 0.986, and 10 m	$V_{\sigma 0.986}$, $V_{\sigma 0.994}$, V_{10}	m s-1
	Wind speed at sigma levels 0.994, 0.986 and 10 m	$WS_{\sigma 0.986}$, $WS_{\sigma 0.994}$, WS_{10}	m s-1

The wind shear was computed using $WS_{\sigma 0.994}$ and $WS_{\sigma 0.986}$ data as detailed in (Couto and Estanqueiro, 2022). It is important to mention that while parameters such as solar zenith angle or hour of the day (Verbois et al., 2022) or sunshine duration (Castangia et al., 2021) can be also explored, particularly for solar PV, the focus of this work is on NWP-based features that can improve the power forecasts.

4.2. Observed power data: wind and solar photovoltaic

It was not possible to identify data publicly available from wind and solar PV hybrid parks to apply and test the developed methodology. To overcome that difficulty it was used data from existing small solar PV power plants located in proximity to existing wind parks that are affected by the same (or very similar) meteorological conditions. The combination of such wind and solar PV power plants can be configured similarly to an HPP sharing the same substation. To enable the comparison of results, the following criteria were adopted in this work to identify case studies: data availability for a period of more than 2 years, distance between wind and solar PV parks of less than 5 km, similar nominal capacity of the wind park (in this case, 20.00 MW), and the characterization of complementarity level between wind and solar PV energy. Based on these criteria, three different HPPs were identified.

Regarding the solar PV power plants, the nominal capacity - $SolarPV_{CapacityRatio}$- varied from nearly 0.14 MW in HPP1 to 2.00 MW in HPP2 and HPP3. To increase the impact of solar PV technology, the observed data of this technology was upscaled using two different nominal capacity ratios. The first ratio (*PV5*) corresponds to a peak capacity of the solar PV of 5 MW, while the second ratio (*PV10*) corresponds to a peak capacity of 10 MW. The hourly solar PV production ($SolarPV_{CapacityRatio}$) within each HPP was obtained using equation (8).

$$\mathrm{SolarPV}_{CapacityRatio}(\mathrm{t}) = \frac{CapacityRatio}{SolarCapacity}\mathrm{SolarPV}(\mathrm{t}) \quad (8)$$

where *CapacityRatio* is the assumed peak capacity of the solar PV power plant while *SolarCapacity* and $\mathrm{SolarPV}(t)$ are the peak capacity of the existing solar PV power plant and the hourly production for the solar power plant, respectively.

4.3. Case studies location and wind and solar photovoltaic complementarity characterisation

Fig. 2 displays the regions where the HPPs are located as well as the normalized daily profile for each technology. Due to confidentiality reasons, the exact locations/coordinates of the vRES plants cannot be disclosed.

The evaluation of complementarity between wind and solar PV generation typically involves the use of correlation-based metrics (Jurasz et al., 2020a), (Yan et al., 2020). Table 2 presents the Pearson correlation values at hourly and daily levels, which is a well-established metric for identifying linear relationships between wind and solar PV generation. The Pearson correlation ranges from −1, indicating strong complementarity, to 1, indicating strong similarity (see Table 2).

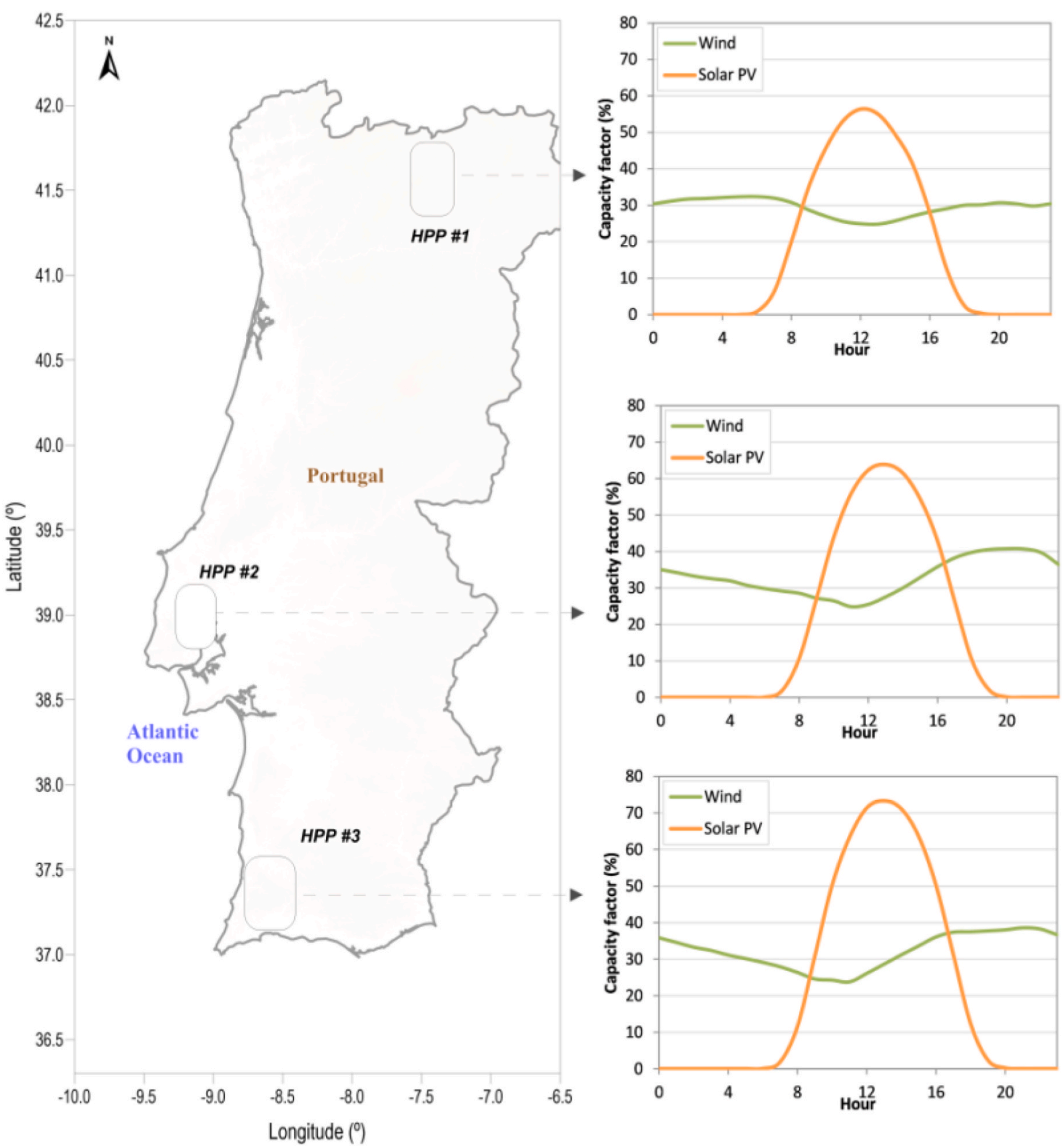

Fig. 2. a) Location of HPPs analysed and the average daily profile normalized for each location. The boxes in the map represent the region of HPP.

Table 2
Hourly and daily correlation in each HPP.

HPP	Hourly resolution	Daily resolution
1	−0.21	−0.38
2	−0.07	0.23
3	0.09	0.12

Table 2 illustrates the varying degrees of complementarity among the HPPs analysed, which are consistent with the outcomes presented in (Couto and Estanqueiro, 2021). The highest level of complementarity was observed in the inner regions centre/north of Portugal, where HPP1 is located. Conversely, HPP2 and HPP3 have correlation values near 0 at the hourly scale and weak positive values at the daily scale, indicating that wind and solar PV generation behave similarly.

5. Power forecasts

This chapter presents and discusses the power forecast results for an HPP and the potential benefits identified. Table 3 summarises the scenarios analysed in this work and their specifications. The benchmark for this work is the forecast for existing wind power plants using the minimisation of the RMSE as an objective function in the SFFS algorithm.

To gain insights into how to forecast power for HPP, two distinct sets of scenarios were conducted. In the first, the wind and solar PV time series were aggregated before the forecasts were obtained. This set of scenarios was designated as “Aggr”. According to Lindberg et al. (2023), forecasting aggregated time series has the potential effect of reducing the power production variability, thus enabling to reduce the associated errors. In the second set of scenarios, the power forecasts were obtained individually for each technology, and the results were added being limited at each time step the maximum inject capacity allowed. These scenarios were designated as “Comb”. For these scenarios, and after sensitivity tests, it was determined that the hourly solar power forecast should be constrained to the maximum hourly value observed in the calibrated dataset for the corresponding month. Therefore, if the forecast for a specific hour exceeds the historically observed value for that hour in the same month of previous years, the maximum historical value is considered. Otherwise, the forecast value obtained with the applied methodology is used.

The maximum inject capacity allowed in all scenarios is 20 MW. In any case, if the sum is above 20 MW, the hourly maximum combined power value was capped at 20 MW. For each scenario, the forecast methodology was calibrated using data from 01 January 2017 to 30 June 2019. The validation dataset consists of data from 01 July 2019 to 31 December 2019. The results shown in the next sections are for the validation dataset.

5.1. Wind power plant hybridisation: impact on power forecasts

Table 4 to Table 5 present the results of the technical and economic metrics for the different HPPs and scenarios under analysed in this work for the validation dataset.

From the previous tables, it is possible to observe that HPP1 has the highest RMSE values among the three analysed power plants, slightly above 3 MW (around 12% of the maximum capacity). However, some hybridization scenarios show lower RMSE values compared to the benchmark (*Wind_RMSE*). The NRMSE values decrease in all hybridization scenarios. The $Comb\_RMSE_{10PV}$ scenario has the lowest NRMSE (38.92%), representing a reduction of 8% compared to the benchmark. The market value improvement for HPP1 compared to the existing wind power plant ranges from 1.07% (*Wind_DAM*) to 5.31% ($Comb\_DAM_{10PV}$). This wide range can be partially attributed to the benefits of wind and solar complementarity. However, despite these benefits, HPP1 has the lowest market value and average remuneration. This is attributed to its lower capacity factor during hours of high market prices. Additionally, the highest forecast errors among the three case studies may also contribute to the lower market value of HPP1.

HPP2 shows slightly less accuracy in predicting power generation in the *Wind_DAM* scenario, with a higher RMSE value (3.07 MW) compared to the benchmark. However, several scenarios, such as Agg_RMSE_{5PV}, $Comb\_DAM_{5PV}$, Agg_RMSE_{10PV}, and $Comb\_DAM_{10PV}$, have lower RMSE values than the benchmark. When comparing NRMSE values to the *Wind_RMSE* results, the best case ($Comb\_DAM_{5PV}$) shows a decrease of nearly 5% with an additional 5 MW of solar PV, while the 10 MW hybridization scenarios show a decrease of almost 6%. The increased market value ranges from 0.27% (*Wind_DAM*) to 2.24% ($Comb\_DAM_{10PV}$) for the analysed scenarios. The maximum increase in remuneration (21.18%) is observed in the $Comb\_DAM_{10PV}$ scenario. For HPP3, the RMSE values range from 2.89 MW ($Comb\_DAM_{5PV}$) to 3.07 MW ($Aggr\_DAM_{10PV}$). *Wind_RMSE* outperformed the results for HPPs in several scenarios showing a reduction that ranged from 3 to 8% in the remaining scenarios. The only exception is the *Wind_DAM* scenario with a RMSE value above the benchmark scenario. The scenario $Comb\_DAM_{10PV}$ presents the highest remuneration increment reaching nearly 31%.

Fig. 3 summarises the results of the different scenarios considering the *Wind_RMSE* results as the benchmark. In Fig. 3a), a positive value means an underperformance compared to the benchmark.

From Fig. 3 is possible to observe an improvement in performance metrics such as negative RMSE and increased market value in most of the scenarios, especially for HPP1. HPP2 is the HPP that presents the lowest benefits with the market value ranging from 1.19 to 2.24%. Hybridising existing wind power plants with solar technology enables to reduce the RMSE in several cases, especially for low additional levels of solar PV. The result can be partially explained by the cancellation of errors obtained when integrating the two technologies in the *Comb*

Table 3
Specifications of the scenarios performed.

Technology	Scenario	Solar Capacity (MW)	Objective function in the SFFS algorithm		Observation
			Minimise NRMSE	Maximise Remuneration	
Wind	Wind_RMSE	–	x		Benchmark approach
	Wind_DAM	–		x	–
Wind and Solar PV (HPP)	Agg_RMSE_{PV5}	5	x		Forecast based on an aggregated time series of wind and solar PV generation limited to the maximum inject capacity
	Agg_DAM_{PV5}			x	
	Agg_RMSE_{PV10}	10	x		
	Agg_DAM_{PV10}			x	
	$Comb\_RMSE_{PV5}$	5	x		Forecasts are obtained for each technology and then are joined being limited to maximum inject capacity.
	$Comb\_DAM_{PV5}$			x	
	$Comb\_RMSE_{PV10}$	10	x		
	$Comb\_DAM_{PV10}$			x	

Table 4
Technical forecast performance for the different HPPs and scenarios.

Scenario	HPP 1		HPP 2		HPP 3	
	RMSE (MW)	NRMSE (%)	RMSE (MW)	NRMSE (%)	RMSE (MW)	NRMSE (%)
Wind_RMSE	3.10	47.70	2.97	38.74	2.97	41.95
Wind_DAM	3.19	49.08	3.07	40.05	3.00	42.37
Agg_RMSE$_{5PV}$	**3.00**	41.94	2.95	35.11	2.89	35.85
Agg_DAM$_{5PV}$	3.09	43.19	3.03	36.06	3.08	38.21
Comb_RMSE$_{5PV}$	3.01	42.08	3.01	35.82	**2.96**	36.72
Comb_DAM$_{5PV}$	3.17	44.31	**2.94**	34.99	2.98	36.97
Agg_RMSE$_{10PV}$	3.11	39.95	2.95	32.47	3.00	33.36
Agg_DAM$_{10PV}$	3.19	40.98	3.01	33.13	3.07	34.14
Comb_RMSE$_{10PV}$	3.03	**38.92**	2.97	32.69	2.97	**33.03**
Comb_DAM$_{10PV}$	3.18	40.85	**2.94**	**32.36**	3.00	33.36

Table 5
Economic forecast performance for the different HPPs and scenarios.

Scenario	HPP 1		HPP 2		HPP 3	
	Market value (%)	Increment of remuneration (%)	Market value (%)	Increment of remuneration (%)	Market value (%)	Increment of remuneration (%)
Wind_DAM	1.07	0.98	0.27	0.27	0.74	0.61
Agg_RMSE$_{5PV}$	2.97	13.40	1.19	10.92	2.27	15.59
Agg_DAM$_{5PV}$	3.26	13.94	1.23	10.96	2.36	15.67
Comb_RMSE$_{5PV}$	2.97	13.40	1.23	10.96	2.37	15.71
Comb_DAM$_{5PV}$	3.74	15.63	1.43	11.38	2.46	15.79
Agg_RMSE$_{10PV}$	4.94	25.82	1.99	20.88	3.24	30.14
Agg_DAM$_{10PV}$	4.99	25.91	2.04	20.95	3.47	30.45
Comb_RMSE$_{10PV}$	4.84	25.66	2.12	21.03	3.59	30.61
Comb_DAM$_{10PV}$	**5.31**	**26.42**	**2.24**	**21.18**	**3.60**	**30.63**

scenarios and the use of a series with less wind power production variability in the case of *Aggr* scenarios.

Although no consistent and significant differences were observed, the results suggest that using separate forecasts for each technology ("*Comb*" scenarios) appears to be the most suitable approach.

Regarding the objective function, results show a dependence on the objective function used. Using the market value objective function leads to the highest performance in economic metrics while minimisation of RMSE leads to better results in the technical metrics.

The results obtained indicate that the integration of HPPs into electrical power systems has a minimal impact since maintains the forecast errors in the range of the values previously observed in the existing wind park. In fact, in certain cases, it may even lead to a reduction in forecast uncertainty. Moreover, the hybridization of HPPs brings substantial economic benefits to producers operating within an electricity market, regardless of the objective function used.

5.2. Identification of the relevant meteorological features

To better understand the previous results, Table 6 presents the meteorological features identified after applying the SFFS approach for HPP1. In this table, the parameters identified in the "*Comb*" scenarios include the parameters obtained using the methodology proposed for forecasting each technology. Thus, in the *Comb_RMSE$_{PV5}$*, for example, the parameters are those achieved in the *Wind_RMSE* scenario plus the ones obtained in the solar power forecast considering a 5 MW capacity. For the sake of brevity, the focus is given to wind and HPP power forecasts. Solar power forecasts are only addressed as a step needed to obtain the forecast in the "*Comb*" scenario.

According with Table 6, the following parameters were not identified as adequate to improve the objective function results: CFH, CFT, CIN, HGT_{500}, MSLP, $Prec_{Conv.}$, SST, T_{500}, U_{10}, and WS_{10}. Unusual meteorological parameters in forecast systems, such as Gust, $MSLP_{Grad.}$, PBL, WindShear, and $T_{var.}$, were identified as important to accomplish the objective functions.

While the parameters remain generally similar across the same set of scenarios, slight differences can be observed in the parameters and number of PCs necessary to achieve the best performance. In the *Agg* and *Comb* scenarios, certain parameters such as temperature (at different levels), CFL, HR, $T_{var.}$, SHF, and SWF resulted from solar PV technology forecast results. For the same installed capacity and objective function, a highest number of parameters and PCs are required to achieve the best forecast in the *Comb* scenarios compared with the *Agg* scenarios. Similarly, the objective function of maximising producer remuneration utilizes a highest number of meteorological parameters compared with minimising the RMSE. However, considering the total combined number of PCs from all available meteorological variables was 183, it is possible to identify that less than 15% of these PCs are used. For the wind power forecast, the most significant differences in the results between the different objective functions are the utilisation of the wind U-component and information regarding the geopotential height at a highest atmospheric level in the case of *Wind DAM*, which maximises revenue in a market environment.

5.3. Impact of the meteorological features on the power forecast errors

Using HPP 1 and considering the *Agg_RMSE$_{10PV}$* scenario as example, the benefits of using the SFFS algorithm are analysed in detail in this section aiming to increase the explainability of the optimal subset of meteorological features in the final results.

Table 7 presents the PCs identified with the SFFS for this scenario, namely, the meteorological parameter and the PCs number, the ranking of each parameter (determined based on the distance correlation) and the feature importance score.

As expected, the two PCs with highest correlation are the most important for the model obtaining feature importance score values above 10%. The two first parameters $WS_{\sigma 0.986}$ PC#1 and Gust PC#1 are strongly associated with the main driving force of wind power. Results also show that despite similar distance correlation values of the wind speed in two different levels of the atmosphere ($WS_{\sigma 0.986}$ and $WS_{\sigma 0.994}$), the wind speed at the highest level (nearly 100m) has a higher importance in the performance. Until the feature T_{850} PC#1, as the distance

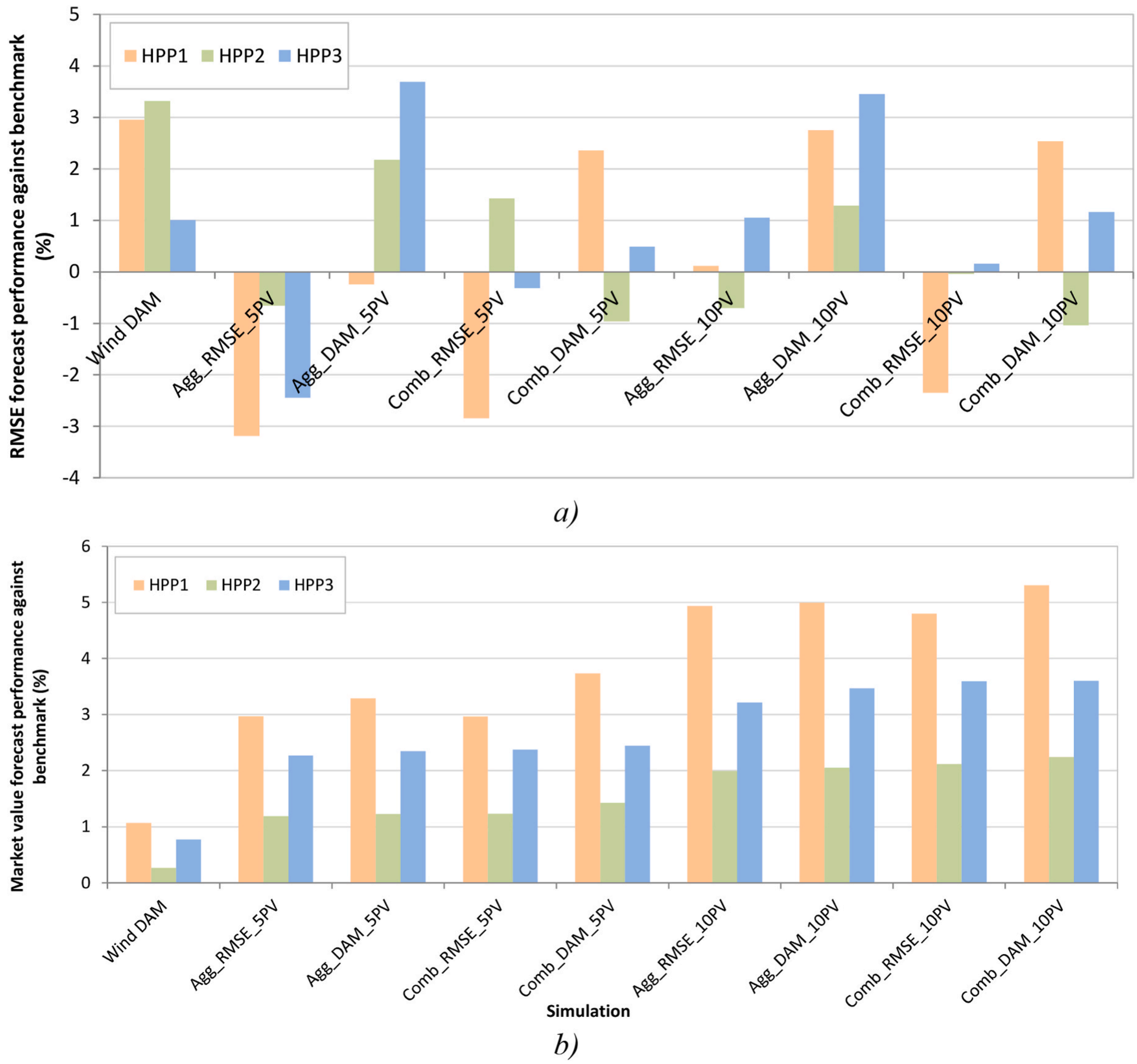

Fig. 3. Forecast performance against benchmark scenario: a) RMSE and b) market value.

correlation decreases there is a tendency to decrease the influence on the model. After, features such as $U_{\sigma0.994}$ PC#2, $V_{\sigma0.986}$ PC#2 and WPD PC#4 associated with the wind power variability show an increase in the importance score despite the low distance correlation values.

Results show that the number of features usually associated with solar power forecast variability is reduced. Nevertheless, parameters such PBL PC#1 reflect the atmosphere dynamic in lowest levels of the atmosphere. As discussed Couto and Estanqueiro (2022), the PBL height varies during the day due to factors such as daily solar radiation cycle and cloud cover. T850 PC#1 is the feature with the lower importance score. Based on the results shown in Table 6, scenarios analysed without considering hybridization, temperature may not be relevant to improve the model performance for forecast wind power production. However, for solar energy, this parameter is relevant due to the efficiency of the panels that varies linearly with the temperature.

The results indicate that the SFFS algorithm identifies relevant features ranked at position 147 (from the 183 available). This feature (WindShear) presents a distance correlation of 0.15 but it has an importance score close to 4%. Thus, the results suggest that features with low correlation can have a significant impact on the performance obtained with the ANN approach.

Fig. 4 presents the temporal time series of the observed and forecasted power. The results are presented based on the use of first 5 and 10 of the most common meteorological features obtained through the SFFS algorithm, as well as all 17 meteorological features (Table 6).

From Fig. 4 it can be observed that as the number of features used increases the forecasts tends to significantly reduce their errors, as can be seen in more detail for the all validation period in Table 8.

6. Conclusion

This study focused on power forecasts for utility hybrid power plants (HPP), which is a new area of research that presents several challenges. Specifically, this work analysed the benefits of hybridyzing wind and solar PV plants, *i.e.*, by creating HPPs, from the accuracy of power forecasts and the value of the energy generated in electricity markets perspectives. That was accomplished by considering three case studies with different levels of wind and solar PV complementarity. The methodology applied allowed for the identification of the most relevant meteorological parameters in the forecast for this type of power plant according to two objective functions: i) minimising the root mean squared error (RMSE); and ii) maximising the remuneration in a market environment.

Hybridization of wind power plants with solar PV capacity increases market value and remuneration for the three case studies analysed in this work. The highest increase in market value was observed in

Table 6
Meteorological features identified with the SFFS algorithm for the HPP1.

Parameter	Wind RMSE	Wind DAM	Agg RMSE_PV5	Agg DAM_PV5	Comb RMSE_PV5	Comb DAM_PV5	Agg RMSE_PV10	Agg DAM_PV10	Comb RMSE_PV10	Comb DAM_PV10
CAPE				1						
CFH										
CFL	1	1			1	1	1	1	2	1
CFM			1			1			1	1
CFT										
CIN										
Gust	1	1	1	1	1	1	1	2	1	1
$HGT_{\sigma0.986}$	1				1		2		1	
HGT_{850}		1	2	1		1		1		1
HGT_{500}										
HR					1	1	1	1	1	1
LHF					1				1	1
LWF			1	1	1	2			1	2
MSLP										
$MSLP_{Grad}$	1	1	1	1	1	1	1	1	1	1
PBL	1	1	1	1	1	1	1	1	1	2
Prec						1		1	1	
$Prec_{Conv.}$										
SHF				1				2		
SST										
SWF					1	1			1	1
T_2				2	1	1			2	1
T_{500}										
T_{850}							1			
T_{var}					1	1			1	2
U_{10}										
$U_{\sigma0.994}$		1	2	1		2	1	1		2
$U_{\sigma0.986}$				1				1		
V_{10}	1	1		1	1	1			1	2
$V_{\sigma0.994}$			2		2	1			1	
$V_{\sigma0.986}$	1	2	1		1	2	1	1	1	2
Vis	1	1		1	1	1			1	1
WindShear	1		1	1	1		1		1	
WPD	2	2	1	1	2	2	2	2	2	2
WS_{10}										
$WS_{\sigma0.994}$	2	2	2	2	2	2	2	2	2	2
$WS_{\sigma0.986}$	2	2	2	2	2	2	2	2	2	2
Total PCs	15	16	18	19	23	26	17	19	26	28

Nr. of PCs used	1	2

Table 7
Analysis of meteorological parameters identified with the SFFS algorithm.

Meteorological parameter and PCs number	PC ranking based on the distance correlation	Distance correlation (Dimensionless)	Feature importance score (%)
$WS_{\sigma0.986}$ PC#1	1	0.62	11.22
Gust PC#1	2	0.62	12.36
$WS_{\sigma0.994}$ PC#1	3	0.61	8.05
$MSLP_{Grad}$ PC#1	6	0.51	6.35
WPD PC#1	7	0.48	5.15
PBL PC#1	8	0.47	4.62
$WS_{\sigma0.986}$ PC#2	10	0.42	4.07
$HGT_{\sigma0.986}$ PC#2	11	0.41	3.06
$WS_{\sigma0.986}$PC#2	12	0.40	3.12
$WS_{\sigma0.994}$ PC#2	23	0.37	3.13
T_{850} PC#1	29	0.32	1.41
$HGT_{\sigma0.986}$ PC#1	34	0.30	4.27
$U_{\sigma0.994}$ PC#2	39	0.29	5.57
$V_{\sigma0.986}$ PC#2	40	0.28	5.10
WPD PC#4	70	0.23	5.25
CFL PC#1	71	0.23	3.70
WindShear	147	0.15	3.99

scenarios with high wind and solar PV complementarity (HPP1), where the market value increased between 2.97% and 5.31% compared to the reference case that uses only the existing wind park with RMSE objective function. Hybridization can also contribute to reduce the forecast RMSE in some scenarios, especially those with low levels of additional solar PV capacity, compared with forecast errors for a reference case study.

The selection of meteorological features and objective functions affected the results. Unusual meteorological parameters in forecast systems, such as wind gust, mean seal level pressure gradient, planetary boundary layer height and wind shear were identified as important for accomplishing the objective functions. Although, results from the feature importance show that not all have the same importance for the final performance of the forecast approach. The objective function of maximising the producer remuneration utilizes more meteorological parameters compared with the objective function of minimising the RMSE. Nevertheless, the findings emphasize the importance of properly selecting the meteorological features and objective functions to calibrate the forecast approach based on user requirements.

Future perspectives of this work include the investigation of different forecast methods and their explainability and enhancing the robustness

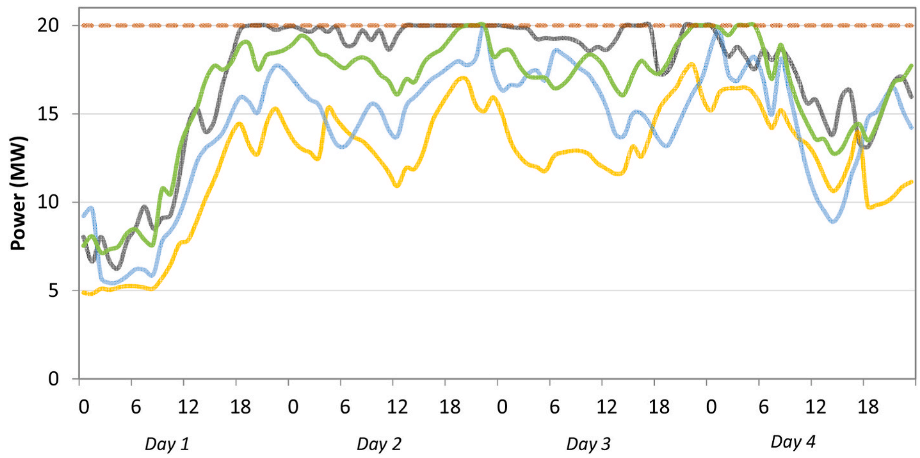

Fig. 4. Power observed and forecasted for four days in HPP1 utilizing different numbers of meteorological features: 5, 10 and 17 (all features for this scenario). Dashed orange line represents the maximum inject capacity of the HPP (20 MW). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 8
Technical metric results for the *Agg_RMSE*$_{10PV}$ scenario in HPP1 for different number of meteorological features.

Number of meteorological features	RMSE (MW)	NRMSE (%)
5	4.38	56.27
10	3.65	46.89
17 (all)	3.11	39.95

of the meteorological features identified. Artificial neural networks are one of the most widely used techniques in various power forecasting applications. However, more advanced techniques are available, so, it is important to ascertain if these techniques can provide better technical and economic performance in the forecast for HPP, bringing added value to the power producers and/or the overall power system. In addition, the robustness of the meteorological parameters identified during this work needs also to be analysed in more detail, especially in the operationalization phase, and regularly evaluated to determine if they reveal, for example, seasonal dependence.

In the short-term, hybridisation of existing wind power plants can play a relevant role in supporting the transition to nearly 100% renewable power systems due to the various advantages such as increasing the capacity factor of the power plants, which is beneficial for producers. Furthermore, the results of this study suggest that the integration of solar PV into existing wind power plants, although increasing the overall renewable capacity, it maintains the forecast errors in the range of the values previously observed in the wind power plants, and, in some cases, could enable to reduce the forecast errors.

Funding

This work has received funding from the EU Horizon 2020 research and innovation program under project TradeRES (grant agreement No 864276)

CRediT authorship contribution statement

António Couto: performed the literature review, collected, and, processed all the data, implemented all the necessary computational routines, and, wrote the first version of the manuscript. **Ana Estanqueiro:** discussed the work to be developed, supported the identification of the most relevant results to be shown, and, also reviewed the manuscript, All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Antonio Couto reports financial support was provided by Horizon 2020.

Data availability

The authors do not have permission to share data.

References

Ahmed, R., Sreeram, V., Mishra, Y., Arif, M.D., 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. Renew. Sustain. Energy Rev. 124 (109792), 26.
Alessandrini, S., McCandless, T., 2020. The Schaake Shuffle technique to combine solar and wind power probabilistic forecasting. Energies 13 (10), 2503.
Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G., 2015. An analog ensemble for short-term probabilistic solar power forecast. Appl. Energy 157, 95–110.
Algarvio, H., Couto, A., Lopes, F., Estanqueiro, A., 2019. Changing the day-ahead gate closure to wind power integration: a simulation-based study. Energies 12 (14), 2765.
Andrade, J.R., Bessa, R.J., 2017. Improving renewable energy forecasting with a grid of numerical weather predictions. IEEE Trans. Sustain. Energy 8 (4), 1571–1580.
Antonanzas, J., Pozo-Vázquez, D., Fernandez-Jimenez, L.A., Martinez-de-Pison, F.J., 2017. The value of day-ahead forecasting for photovoltaics in the Spanish electricity market. Sol. Energy 158, 140–146.
Antonanzas, J., Perpinan-Lamigueiro, O., Urraca, R., Antonanzas-Torres, F., 2020. Influence of electricity market structures on deterministic solar forecasting verification. Sol. Energy 210, 44–46.
Babaremu, K., et al., 2022. Overview of solar–wind hybrid products: prominent challenges and possible solutions. Energies 15 (16: 6014), 25.
Bessa, R.J., Miranda, V., Botterud, A., Wang, J., 2011. 'Good' or 'bad' wind power forecasts: a relative concept. Wind Energy 14 (5), 625–636.
Bochenek, B., et al., 2021. Day-ahead wind power forecasting in Poland based on numerical weather prediction. Energies 14 (2164), 18.
Carriere, T., Kariniotakis, G., 2019. An integrated approach for value-oriented energy forecasting and data-driven decision-making application to renewable energy trading. IEEE Trans. Smart Grid 10 (6), 6933–6944.
Castangia, M., Aliberti, A., Bottaccioli, L., Macii, E., Patti, E., 2021. A compound of feature selection techniques to improve solar radiation forecasting. Expert Syst. Appl. 178, 114979.
Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40 (1), 16–28.
Couto, A., Estanqueiro, A., 2021. Assessment of wind and solar PV local complementarity for the hybridization of the wind power plants installed in Portugal. J. Clean. Prod. 319 (128728), 12.
Couto, A., Estanqueiro, A., 2022. Enhancing wind power forecast accuracy using the weather research and forecasting numerical model-based features and artificial neuronal networks. Renew. Energy 201, 1076–1085.
Couto, A., Costa, P., Rodrigues, L., Lopes, V.V., Estanqueiro, A., 2015. Impact of weather regimes on the wind power ramp forecast in Portugal. IEEE Trans. Sustain. Energy 6 (3), 934–942.
Davò, F., Alessandrini, S., Sperati, S., Delle Monache, L., Airoldi, D., Vespucci, M.T., 2016. Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. Sol. Energy 134, 327–338.

Dykes, K., et al., 2020. Opportunities for Research and Development of Hybrid Power Plants Opportunities for Research and Development of Hybrid Power Plants, p. 25. *NREL/TP-5000-75026*.
Éles, A., Heckl, I., Cabezas, H., 2021. Modeling renewable energy systems in rural areas with flexible operating units. Chem. Eng. Trans. 88, 643–648.
EDP, 2023. EDPR commissions its first hybrid solar and wind energy park on the Iberian Peninsula [Online]. Available: https://www.edpr.com/en/news/2023/01/12/edpr-commissions-its-first-hybrid-solar-and-wind-energy-park-iberian-peninsula.
Elsken, T., Metzen, J.H., Hutter, F., 2018. Neural architecture search: a survey. J. Mach. Learn. Res. 20, 1–21.
Eseye, A.T., Zhang, J., Zheng, D., 2018. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. Renew. Energy 118, 357–367.
García-Cuesta, E., Aler, R., del Pózo-Vázquez, D., Galván, I.M., 2022. A combination of supervised dimensionality reduction and learning methods to forecast solar radiation. Appl. Intell. 14.
Heaton, J., 2008. Introduction to Neural Networks for Java, second ed. Heaton Research, Inc. 2nd ed.
Heusinger, J., Broadbent, A.M., Sailor, D.J., Georgescu, M., 2020. Introduction, evaluation and application of an energy balance model for photovoltaic modules. Sol. Energy 195, 382–395.
Jerez, S., et al., 2023. An action-oriented approach to make the most of the wind and solar power complementarity. Earth's Future 11 (6).
Jurasz, J., Canales, F.A., Kies, A., Guezgouz, M., Beluco, A., 2020a. A review on the complementarity of renewable energy sources: concept, metrics, application and future research directions. Sol. Energy 195, 703–724.
Jurasz, J., Kies, A., Zajac, P., 2020b. Synergetic operation of photovoltaic and hydro power stations on a day-ahead energy market. Energy 212 (118686), 12.
Kapica, J., Canales, F.A., Jurasz, J., 2021. Global atlas of solar and wind resources temporal complementarity. Energy Convers. Manag. 246 (June), 114692.
Karadöl, İ., Yıldız, C., Şekkeli, M., 2021. Determining optimal spatial and temporal complementarity between wind and hydropower. Energy 230 (120790), 18.
Klonari, V., Fraile, D., Rossi, R., Schmela, M., 2019. Exploring the viability of hybrid wind- solar power plants. In: 4th International Hybrid Power Systems Workshop, p. 7.
Lange, M., Focken, U., 2006. Physical Approach to Short-Term Wind Power Prediction. Springer, Berlin, Heidelberg.
Lindberg, O., Arnqvist, J., Munkhammar, J., Lingfors, D., 2021. Review on power-production modeling of hybrid wind and PV power parks. J. Renew. Sustain. Energy 13, 042702-1-042702–22.
Lindberg, O., Lingfors, D., Arnqvist, J., van der Meer, D., Munkhammar, J., 2023. Day-ahead probabilistic forecasting at a co-located wind and solar power park in Sweden: trading and forecast verification. Adv. Appl. Energy 9, 100120.
Liu, H., Chen, C., 2019. Data processing strategies in wind energy forecasting models and applications: a comprehensive review. Appl. Energy 249, 392–408.
Lv, S.-X., Wang, L., 2023. Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model. Energy 263, 126100.
Markovics, D., Mayer, M.J., 2022. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew. Sustain. Energy Rev. 161 (112364), 17.
Matlab, 2012. MATLAB and Neural Network Toolbox Release. The MathWorks, Inc., Natick, Massachusetts, United States.
Mellit, A., Kalogirou, S.A., 2008. Artificial intelligence techniques for photovoltaic applications: a review. Prog. Energy Combust. Sci. 34 (5), 574–632.
MeteoGalicia, 2021. MeteoGalicia THREDDS [Online]. Available: https://www.meteogalicia.gal. (Accessed 2 July 2021).
Molnar, C., et al., 2022. General pitfalls of model-agnostic interpretation methods for machine learning models. In: Goebe, R., Wahlste, W., Zhou, Z.-H., Siekmann, J. (Eds.), xxAI - beyond Explainable AI. Springer, Cham, pp. 39–68.
Nazaré, G., Castro, R., Gabriel Filho, L.R.A., 2020. Wind power forecast using neural networks: tuning with optimization techniques and error analysis. Wind Energy 23 (3), 810–824.
Ntomaris, A.V., Marneris, I.G., Biskas, P.N., Bakirtzis, A.G., 2022. Optimal participation of RES aggregators in electricity markets under main imbalance pricing schemes: price taker and price maker approach. Elec. Power Syst. Res. 206 (107786), 14.
OMIE, 2022. MIBEL market results [Online]. Available: https://www.omie.es/es/file-access-list. (Accessed 1 August 2022).
Papageorgiou, E.I., Poczęta, K., 2017. A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks. Neurocomputing 232, 113–121.
Pombo, D.V., Rincón, M.J., Bacher, P., Bindner, H.W., Spataru, S.V., Sørensen, P.E., 2022. Assessing stacked physics-informed machine learning models for co-located wind–solar power forecasting. Sustain. Energy, Grids Networks 32 (100943), 13.
Qadir, Z., et al., 2021. Predicting the energy output of hybrid PV–wind renewable energy system using feature selection technique for smart grids. Energy Rep. 7, 8465–8475.
Qiu, L., He, L., Lu, H., Liang, D., 2022. Systematic potential analysis on renewable energy centralized co-development at high altitude: a case study in Qinghai-Tibet plateau. Energy Convers. Manag. 267 (115879), 12.
REN, 2022. REN - data Hub [Online]. Available: https://datahub.ren.pt. (Accessed 8 January 2022).
Ruhnau, O., Hennig, P., Madlener, R., 2020. Economic implications of forecasting electricity generation from variable renewable energy sources. Renew. Energy 161, 1318–1327.
Salcedo-Sanz, S., Cornejo-Bueno, L., Prieto, L., Paredes, D., García-Herrera, R., 2018. Feature selection in machine learning prediction systems for renewable energy applications. Renew. Sustain. Energy Rev. 90, 728–741.
Schyska, B.U., Couto, A., von Bremen, L., Estanqueiro, A., Heinemann, D., 2017. Weather dependent estimation of continent-wide wind power generation based on spatio-temporal clustering. Adv. Sci. Res. 14 (2003), 131–138.
Shirkhorshidi, A.S., Aghabozorgi, S., Wah, T.Y., 2015. A Comparison study on similarity and dissimilarity measures in clustering continuous data. PLoS One 10 (12), e0144059, 20.
Solorio-Fernández, S., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., 2020. A review of unsupervised feature selection methods. Artif. Intell. Rev. 53 (2), 907–948.
Strbac, G., et al., 2021. Decarbonization of electricity systems in Europe: market design challenges. IEEE Power Energy Mag. 19 (1), 53–63.
Sweeney, C., Bessa, R.J., Browell, J., Pinson, P., 2020. The future of forecasting for renewable energy. WIREs Energy Environ 9 (2).
Székely, G.J., Rizzo, M.L., 2009. Brownian distance covariance. Ann. Appl. Stat. 3 (4), 1236–1265.
Verbois, H., Saint-Drenan, Y.-M., Thiery, A., Blanc, P., 2022. Statistical learning for NWP post-processing: a benchmark for solar irradiance forecasting. Sol. Energy 238, 132–149.
Visser, L., AlSkaif, T., van Sark, W., 2022. Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. Renew. Energy 183, 267–282.
Wang, K., Qi, X., Liu, H., 2019. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl. Energy 251, 113315.
Wang, Y., Millstein, D., Mills, A.D., Jeong, S., Ancell, A., 2022. The cost of day-ahead solar forecasting errors in the United States. Sol. Energy 231, 846–856.
WindEurope, 2019. Renewable hybrid power plants - exploring the benefits and market opportunities. Tech. Rep. p. 22.
Yan, J., Qu, T., Han, S., Liu, Y., Lei, X., Wang, H., 2020. Reviews on characteristic of renewables: evaluating the variability and complementarity. Int. Trans. Electr. Energy Syst. 30 (7), 1–21.
Yan, J., Möhrlen, C., Göçmen, T., Kelly, M., Wessel, A., Giebel, G., 2022. Uncovering wind power forecasting uncertainty sources and their propagation through the whole modelling chain. Renew. Sustain. Energy Rev. 165, 112519.
Yang, D., van der Meer, D., 2021. Post-processing in solar forecasting: ten overarching thinking tools. Renew. Sustain. Energy Rev. 140 (110735), 35.
Yang, D., et al., 2020. Verification of deterministic solar forecasts. Sol. Energy 210, 20–37.
Yang, B., et al., 2021. State-of-the-art one-stop handbook on wind forecasting technologies: an overview of classifications, methodologies, and analysis. J. Clean. Prod. 283, 124628.
Zhang, J., Hodge, B.-M., Florita, A., 2015. Joint probability distribution and correlation analysis of wind and solar power forecast errors in the western interconnection. J. Energy Eng. 141 (1).
Zhang, H., Cao, Y., Zhang, Y., Terzija, V., 2018. Quantitative synergy assessment of regional wind-solar energy resources based on MERRA reanalysis data. Appl. Energy 216, 172–182.

Glossary

Agg: Forecast based on an aggregated time series
ANN: Artificial neural network
AveragePower: Average observed power
Comb: Forecast based on combined time series
D: Day
DAM: Day-ahead market
FS: Feature selection
HPP: Hybrid power plant
LM: Levenberg-Marquardt
MIBEL: Iberian electricity market
ML: Machine learning
NRMSE: Normalized root mean square error
NWP: Numerical weather prediction
PC: Principal component
PCA: Principal component analysis
P_{For}: Power forecast
P_{Obs}: Power observed
$Price_{DAM}$: Price in day-ahead market
$Price_{DOWN}$: Downward deviation costs
$Price_{UP}$: Upward deviation costs
PV: Photovoltaic
RMSE: Root mean square error
SFFS: Sequential forward feature selection
vRES: Variable renewable energy sources
WRF: Weather research and forecasting

Nomenclature

AveragePower: Average observed power production for each scenario
a_{ti}: PC scores of each meteorological variable
CapacityRatio: Peak capacity of the solar PV power plant in each case study
e_i: Eigenvectors of each meteorological variable
$Performance_{AllRelevant}$: Performance of the model obtained using all the feature identified with the SFFS
$Performance_{Permuted}$: Performance of the model for each feature permuted
$P_{For.}$: Forecasted power forecast for each scenario

$P_{Obs.}$: Observed power forecast for each scenario
$Price_{DAM}$: Hourly day-ahead market price
$Price_{Down}$: Downward hourly deviation costs
$Price_{Up}$: Upward hourly deviation costs
Renumeration: Market remuneration
$SolarPV_{CapacityRatio}$: Hourly observed solar PV production
$SolarPV(t)$: Hourly production for the solar power plant in each case study
$T_{var.}$: Vertical temperature variation
Z: Meteorological variable defined by j-*th* grid spatial points with a temporal dependence