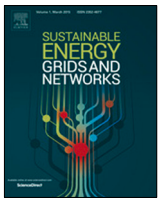
Sustainable Energy, Grids and Networks 35 (2023) 101110

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

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



Deep learning-based forecasting of the automatic Frequency Reserve Restoration band price in the Iberian electricity market

Javier Cardo-Miota *, Emilio Pérez, Hector Beltran

Department of Industrial Systems Engineering and Design, Universitat Jaume I, Castelló de la Plana, Comunitat Valenciana, Spain

ARTICLE INFO

Article history:
Received 10 January 2023
Received in revised form 29 June 2023
Accepted 11 July 2023
Available online 14 July 2023

Keywords:
Ancillary services
aFRR service
Forecasting
Electricity prices
Energy markets
Neural networks
Deep learning

ABSTRACT

The replacement of conventional and dispatchable generation technologies by intermittent renewable energy sources increases the need for ancillary services. New agents, such as batteries, may join frequency regulation markets but they require accurate information about future market prices and service demand trends in order to make their participation profitable. This paper proposes and analyzes the accuracy of various deep learning-based models to estimate the secondary reserve marginal band price in the automatic frequency restoration reserves service of the Iberian electricity market. First, a correlation analysis allows determining various subsets of market variables used as model inputs. These subsets include some highly correlated variables together with different combinations of others whose influenced is analyzed. Next, three different neural network techniques are considered: feedforward, convolutional and recurrent networks. For each of them, a random search is performed to obtain the best set of hyperparameters. The analysis of the results shows how the LSTM model returns the best performance metrics (63.22 % of mean absolute scaled error), clearly improving the state-of-the-art in the domain.

1. Introduction

In the last decade, the amount of renewable energy systems (RESs) penetration in the electricity systems has continuously increased at a global scale, Fig. 1. However, a faster deployment of RESs is required to achieve the Net Zero Emissions (NZE) targets fixed for 2050 [1]. The path to the NZE scenario implies doubling the share of RESs generation in the coming decade, as indicated in Fig. 1 extracted from [2]. The massive integration of RESs into the system introduces several challenges. On the one hand, plant operators of RESs tend to offer energy at very low prices in the day-ahead electricity market in order to ensure their participation. This pushes the marginal prices down in an effect that is being called '*cannibalization*' of the prices, as discussed in [3]. On the other hand, as indicated in [4], the intermittent and stochastic behavior of RESs due to their dependency on natural factors such as wind availability or clouds' movement adds energy fluctuations and operational uncertainty to the system that can affect the grid stability.

Ancillary services (ASs) [5], responsible for ensuring the system security and energy supply conditions within high quality standards against every unexpected event that affect grid stability, become essential for the operation of electric systems with high RESs penetration [6]. There are three categories of ASs: frequency response, voltage control and black-start capability [5].

Traditionally, only conventional generators have been entitled to provide ASs. However, several countries such as Ireland, UK or USA have recently started to qualify additional technologies as ASs providers [7,8]. Among them, battery energy storage systems (BESSs) are an outstanding alternative [9] due to their operational capabilities: rapid response, ability to provide and consume energy, and lifespan. In this context, the Iberian transmission system operator (TSO), 'Red Eléctrica de España (REE)', announced in January 2021 that BESSs would be eligible as ASs providers in the coming years. The main handicap for BESSs is the huge initial investment they require which hinders their current return on investment. Therefore, some tools are required to optimize the battery scheduling with the goal to maximize the revenue and turn the project economically viable. In this sense, electricity price forecasting (EPF) is one of the main strategies being explored to help the operators of BESSs to take advantage of future information so that operational decisions are taken considering which market is more profitable to participate in.

Many works about EPF are found in the literature. For instance, authors in [10] propose a convolutional neural network (CNN) model to forecast the 24 h-ahead marginal price of a specific node in a zone at the PJM power market using other zonal prices. For this purpose, a correlation study among the 21 zonal prices in the PJM market is developed to determine the inputs of the model. Several machine and deep learning approaches are compared and

* Corresponding author.
E-mail address: jcardo@uji.es (J. Cardo-Miota).

https://doi.org/10.1016/j.segan.2023.101110

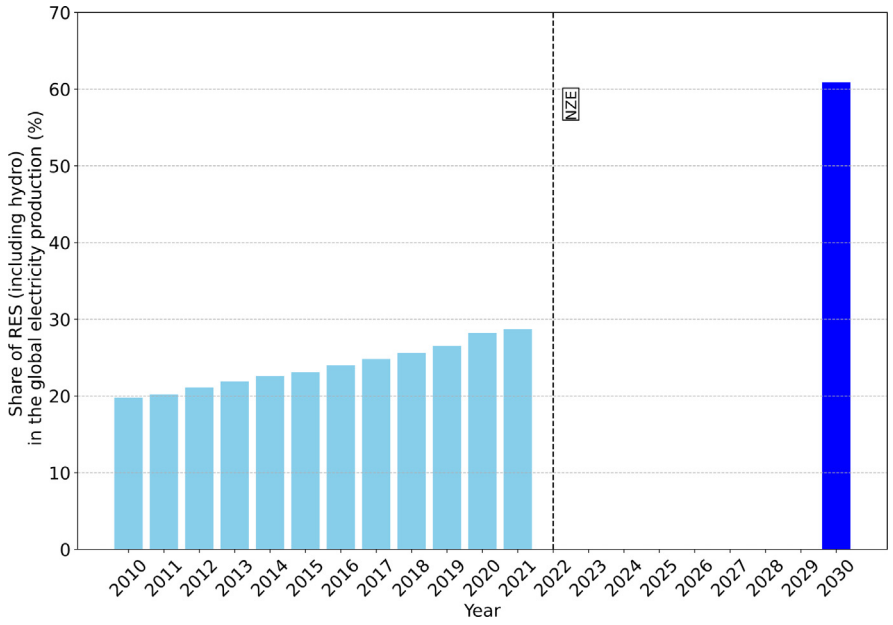


Fig. 1. Global renewables share of power generation and 2030 target in the roadmap to the Net Zero Emissions (NZE) scenario [2].

the proposed model returns a significantly higher accuracy than the other methods. A methodology for EPF in the Lithuanian price zone of the Nord Pool electricity market has also been developed in [11]. Several statistical and artificial neural network (ANN) approaches have been analyzed, with the feedforward model being the technique with the best performance metrics.

In [12] a long short-term memory (LSTM) model has been proposed to forecast electricity prices in three different case studies: PJM electricity market for 2006 and 2018 data, and the Spanish day-ahead market. Before introducing the data inputs into the LSTM model, a wavelet transformation and a MI-based feature selection were used, obtaining a robust model that has a strong capability to forecast electricity prices in different energy markets with high performance.

Two papers have been identified to analyze the EPF in the ASs market, particularly in the automatic fast frequency restoration (aFRR) service. Authors in [13] present a methodology for estimating the optimal amount of aFRR band provided by an aggregation of renewable power plants. To do so, they propose the deterministic forecasting of RESs generation, energy and reserve prices based on statistical forecasting techniques, and the probabilistic forecasting of the price spreads. Conversely, authors in [14] propose several statistical and deep learning forecasting approaches to estimate three variables of the German aFRR market, including the marginal mixed price.

Focusing on the Iberian electricity market, only one proposal has been found in the literature exploring the EPF of its aFRR service. Authors in [15] propose a mixed integer quadratic programming model to maximize the profits of a pump storage plant participating in both the day-ahead and secondary reserve market. For this purpose, some aFRR market variables, including the band price called the intercept of the residual reserve curve, have been forecast using classical approaches such as SARIMA and GARCH-ARIMAX. However, other forecasting techniques, such as those based on deep learning, have not been analyzed to the authors' knowledge. This paper aims to fill that gap and presents a methodology for estimating the secondary reserve marginal band price in the Iberian aFRR service market using a variety of the aforementioned approaches. In particular, three different deep learning-based techniques are analyzed with that goal: feedforward neural networks, CNNs and recurrent neural networks (RNN). The resulting accuracy of the different techniques is compared with that obtained by previous works developed for other markets in order to evaluate the validity of the proposal.

The paper is organized as follows: Section 2 introduces a complete explanation of the Iberian aFRR market. Section 3 presents the data sources, how they are grouped into three different datasets, and the performance metrics used to evaluate the models. Then, the model selection from each deep-learning forecasting approach is described in Section 4. Section 5 presents the results obtained and analyzes the goodness of the models. Finally, Section 6 introduces some conclusions and future works.

2. Automatic Frequency Regulation Reserve (aFRR): Spanish case

For quality and safety conditions, the frequency of the electrical system must remain within certain limits. Consequently, electricity systems count on a series of ASs devoted to maintain the frequency around the reference value (50 Hz in Europe) and, hence, avoiding large frequency oscillations that might cause faulty operation of the equipment connected to the grid.

TSOs usually implement four different types of control reserves to restore the grid frequency upon imbalance, Oureilidis et al. [5]:

- **Frequency Containment Reserves (FCRs)**: responsible for eliminating within 30 s the instantaneous power mismatches experienced between generation and consumption.
- **automatic Frequency Restoration Reserves (aFRRs)**: intended to restore in a time interval between 30 s and 5 min the power exchange balance among neighboring electric systems and responsible for eliminating the frequency deviation that arises from an imbalance.
- **manually Frequency Restoration Reserves (mFRRs)**: ready to be on within 15 min after any imbalance in order to free the aFRR.
- **Replacement Reserves (RRs)**: the typical activation time for RRs range from 15 min to hours after the imbalance occurs. Their function is to restore the capability of the aFRR and mFRR providers.

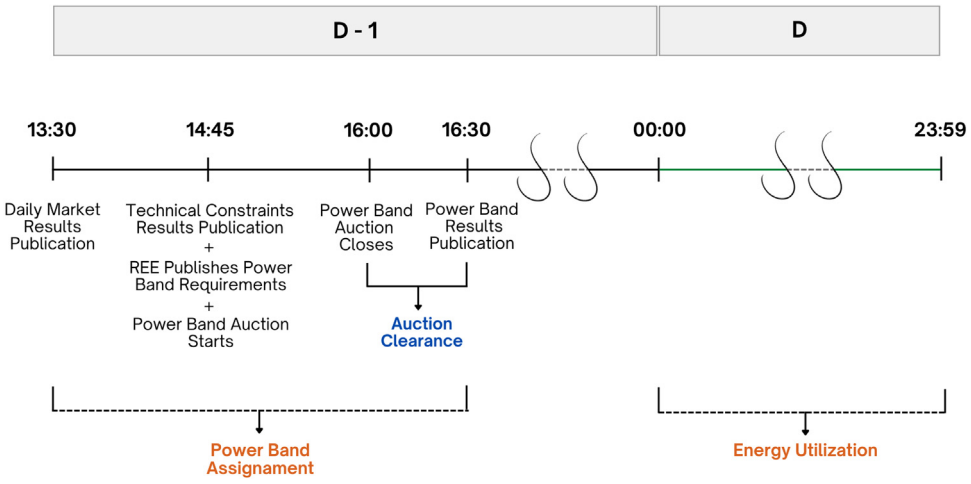


Fig. 2. aFRR market: time line.

Among them, the aFRR being the second fastest control is the key service to eliminate the frequency deviation. Therefore, aFRR providers must present fast response and the ability to both inject and consume and/or stop providing energy into/from the system.

Although only conventional generators have been traditionally entitled in most of the countries to provide this service, recent changes in the regulations in markets such as that of the UK or some in the USA have entitled BESSs and flexible loads to participate. This possibility seems will be open within the European context for the Iberian electricity system in the coming years.

2.1. Iberian aFRR market design

The aFRR market is managed by REE in the Iberian electricity system. It encompasses 2 different mechanisms: the power band assignment and the energy utilization. Both imply an *upward* service, when an increase in the supply or a decrease in the consumption of energy is required (identified as positive energy value), or a *downward* service, when an increase in the consumption or a decrease in the supply of energy is required (negative energy value).

2.1.1. Power band assignment

This mechanism defines the maximum capacity the aFRR participants (called '*regulation zones (RZs)*') might provide every hour in case their service would be required. The hourly resources committed as aFRR reserves are called '*regulation bands*'.

Every day, just after the day-ahead market resolution, the Spanish TSO publishes the hourly upward and downward band requirements for the following day together with the hourly maximum and minimum admissible power band size per bid.

Subsequently, the RZs submit from 2:45 pm to 4 pm their corresponding offers for each available hourly period. Offers may be simple, consisting of a bidirectional (upward and downward) band reserve bid (in MW) at a given price (in €/MW), or composite, with various power bands at increasing prices. Then, a marginal auction takes place, Fig. 2.

Once the auction closes, REE checks if the offers satisfy the band requirements and sorts all those valid for each period, from lowest to highest. The marginal band clearing price is that at the point where the reserve requirement and the RZ bidding curves agree, Fig. 3. Every offer below the clearing price enters the aFRR service during the corresponding period and is paid at that price. Finally, the market results are published at 4:30 pm.

2.1.2. Energy utilization

This mechanism corresponds to the overall energy exchanged every hour by all the RZs, with power band assigned in the auction for that hour, to bring the frequency deviation to zero while following the instructions delivered by REE, Fig. 2. These instructions respond to the continuous evolution of the energy exchanges among Spain, France and Portugal, compared with the scheduled exchanges. The arising deviation signal is called *PRR*. It is divided among the RZs proportionally to their assigned power band, each receiving its corresponding *CRR* signal. Then, the service providers must use that signal together with the frequency deviation signal of the system (Δf) to build every 4 s their corresponding Area Control Error (ACE_i) reference. This reference also depends on the secondary control contribution (NID_i) that each RZ is providing in real time, and it is calculated as shown in Eq. (1):

$$ACE_i = CRR_i + \frac{1}{G} \cdot NID_i - \beta \Delta f, \qquad i = 1, \ldots, N_{RZ} \tag{1}$$

where 'β' *(Hz/W)* and 'G' are constants and 'N_{RZ}' is the number of RZs.

Once the ACE signal is built, each RZ tries to bring it to zero by modifying its generation or consumption, in such a way that ACE implies an upward service and an ACE implies a downward service. The net energy supplied every hour by the RZ, calculated as the energy difference between upward and downward services, is remunerated at the mFRR price that would have been required to activate to provide the same service during that hour.

The non-compliance of the service is penalized with the 50% of the band price defined at the corresponding period.

3. Data sources, datasets, and metrics

Data from various sources have been compiled and grouped into different datasets designed to test the potential influence of the multiple input variables. The performance of the developed models is evaluated through various metrics that take a persistent model as a baseline. Sources, datasets and metrics are detailed in the following.

3.1. Data sources and variables of interest

Two online databases have been consulted to extract data to be fed into the models. These are the websites for the Spanish TSO, REE [16], and the Iberian electricity market operator, OMIE

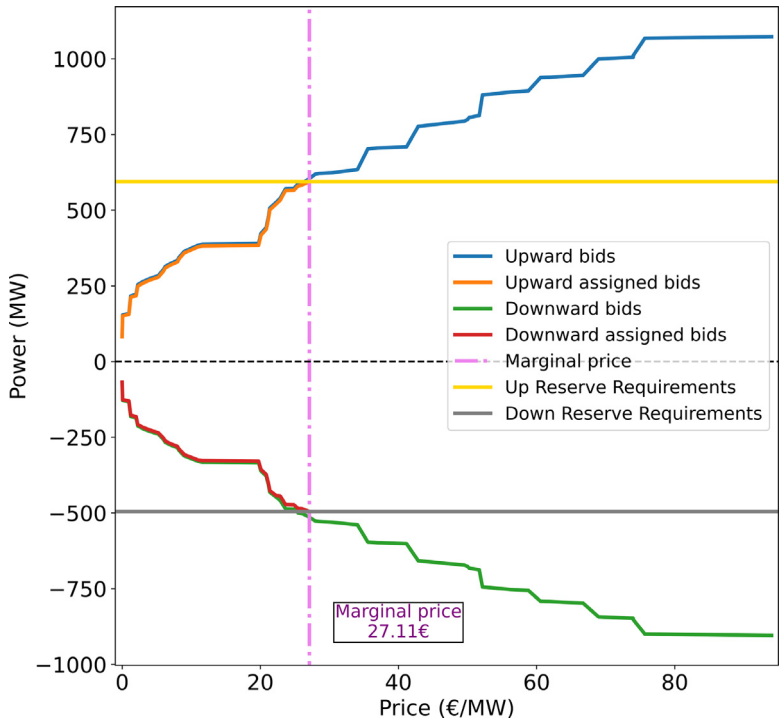


Fig. 3. aFRR market: auction clearance.

Table 1
Delayed variables. (C.C. GT stands for combined-cycle gas turbines; P48 is the resulting exchange program at the closure of the daily schedule in the Iberian electricity system market).

Data	Set 1	Set 2	Set 3	2019 Correlation with target (t −24 h)	2019 Correlation with target (t −48 h)
Secondary reserve marginal band price	x	x	x	**0.48**	0.28
Marginal price day-ahead market	x	x	x	**−0.34**	−0.23
Energy allocated in day-ahead market	x	x	x	−0.10	−0.05
Real-time generation wind	x	x	x	**0.26**	0.15
Real-time generation solar PV	x	x	x	−0.15	−0.15
Real demand	x	x	x	**−0.22**	−0.15
Scheduled demand	x			−0.22	−0.15
Viable daily schedule PVP wind	x			0.27	0.17
Viable daily schedule PVP solar PV	x			−0.15	−0.14
Net secondary reserve energy used (upward – downward)		x	x	0.13	0.08
Real-time generation C.C. GT		x	x	**−0.36**	−0.27
Real-time generation coal		x	x	−0.08	−0.03
Real-time generation hydro		x	x	0.02	0.11
Real-time generation exchanges			x	−0.12	−0.06
Operating hourly program P48 total balance interconnections			x	−0.12	−0.06

[17] (for its Spanish acronym), which provided historical and prediction hourly system and market data.

Tables 1 and 2 list the 32 variables of interest extracted from these databases and used in this work. These variables are classified into two different types of inputs as a function of their characteristics:

- *Delayed data variables* (listed in Table 1): 15 variables whose future values are not known at the time when the prediction is initiated (3 pm) but whose previous 24 hourly samples are public and used in our models. An example of time series belonging to this group is the '*real-time generation solar PV*'.
- *Known data variables* (listed in Table 2): 21 variables whose current and/or future values are either already set and known, or forecast, at the time when the prediction is initiated (3 pm). These future values are used in our models. Examples of time series belonging to this group are the '*peninsular wind power generation forecast*' or the '*marginal price in the day-ahead market*'.

Note that four of the variables are included in both tables: '*Marginal price day-ahead market*', '*Energy allocated in day-ahead market*', '*Viable daily schedule PVP wind*', and '*Viable daily schedule PVP solar PV*', where PVP is the exchange program established at the closure of the day-ahead market.

With regard to the '*Marginal price day-ahead market*' and the '*Energy allocated in day-ahead market*', they belong to the '*Known data variables*' group because they are available at the prediction running time and their future values are required in our models. However, their past values are also of interest to the models and, therefore, they are listed in the '*Delayed*' type. Regarding the '*Viable daily schedule PVP wind*' and '*Viable daily schedule PVP solar PV*' time series, although they are within the '*Known variables*' type, their past behavior in relation to their forecast values may also influence the current and future evolution of the secondary reserve marginal band price for the aFRR service in the Iberian electricity market. Hence, they are equally included in the '*Delayed*' type.

Table 2
Known variables. (UGH (Hydro Management Unit) represents a set of hydropower stations belonging to the same hydroelectric basin and the same individual energy market participant.)

Data	Set 1	Set 2	Set 3	2019 Correlation with target
Marginal price day-ahead market	x	x	x	**−0.53**
Energy allocated in day-ahead market	x	x	x	−0.11
Solar FV generation forecast	x	x	x	−0.16
Peninsular wind power generation forecast	x	x	x	**0.42**
Secondary reserve requirements upward	x	x	x	0.18
Secondary reserve requirements downward	x	x	x	0.04
Energy capacity matched by combined cycle plants in the day-ahead market	x	x	x	**−0.43**
Energy capacity unmatched by combined cycle in the day-ahead market	x	x	x	0.39
Energy capacity matched by hydroelectric in the day-ahead market	x	x	x	0.04
Energy capacity unmatched by hydroelectric in the day-ahead market	x	x	x	0.13
Energy capacity matched by coal in the day-ahead market	x	x	x	−0.18
Energy capacity unmatched by coal in the day-ahead market	x	x	x	0.16
Forecast demand	x			**−0.29**
Scheduled demand PVP total		x	x	−0.26
Total viable daily scheduled generation PVP		x	x	−0.10
Viable daily schedule PVP UGH + non UGH (hydroelectric)			x	0.10
Viable daily schedule PVP combined cycle GT			x	−0.41
Viable daily schedule PVP France balance			x	−0.18
Viable daily schedule PVP Portugal balance			x	−0.02
Viable daily schedule PVP wind			x	0.44
Viable daily schedule PVP solar PV			x	−0.15

Although other variables can be considered to forecast the aFRR band price, these 32 different variables have been selected due to their proven significance in terms of correlation, calculated as defined in Eq. (2) by means of the Pearson's correlation coefficient, *r*.

$$r_{\widehat{y}y} = \frac{\mathrm{cov}(\widehat{Y}Y)}{\sigma_{\widehat{Y}}\sigma_Y}, \tag{2}$$

where *cov* is the covariance between variables and $\sigma_{\widehat{Y}}$, σ_Y are the corresponding standard deviations for each time series.

The resulting correlations for each variable, with regard to the target variable, are defined in Tables 1 and 2. Table 1 contains two correlation columns: '*2019 Correlation with target (t - 24 h)*' and '*2019 Correlation with target (t - 48 h)*' that introduce the correlation between the target variable and the value that each *Delayed* variable presented 24 and 48 h before the instant at which the prediction is made. Accordingly, Table 2 contains '*2019 Correlation with target*' column that introduces the corresponding correlation between the target variable and the value that each *Known variable* is defined to present during the 24 h after the price prediction is made.

Results from Table 2 show that the target variable, the aFRR band price, is strongly correlated with many of the input variables, among which we highlight: *forecast demand*, *marginal price in the day-ahead market*, *peninsular wind power generation forecast*, and *energy capacity matched by combined cycle plants in the day-ahead market*. Note that all of these correlations are negative except for the *wind generation forecast* time series. This is probably due to the fact that higher values of demand and marginal price of the day-ahead market involve or indicate a higher share of conventional generation in the system, mainly gas in the Iberian electricity market, i.e. higher energy capacity matched by combined cycle plants in the day-ahead market. These technologies are dispatchable over time and they hardly commit service failures, minimizing the amount of secondary reserves required and, consequently, the marginal price after the market clearance. On the contrary, wind generation, being the largest energy contributor in annual terms to the Iberian market, depends on natural resources which entail production deviation levels higher and more frequent than those registered for conventional technologies. Therefore, higher wind power generation implies higher deviations that force an increase in secondary reserves' requirements and price. Note that some other inputs from Table 2 are also strongly correlated with aFRR band price due to the fact that they are very similar with the previous four. For instance, consider *viable daily schedule PVP wind* time series which is very similar to *peninsular wind power generation forecast*, or *energy capacity unmatched by combined cycle in the day-ahead market* time series, which is essentially complementary to *energy capacity matched by combined cycle plants in the day-ahead market*.

The correlation for the rest of the variables in Table 2 is significant (above 10%), except for some exceptions such as *exchanges with Portugal* and *hydro matched energy in day-ahead market*. Nonetheless, these have been included as inputs because they have a strong correlation with other input variables.

Finally, very similar conclusions can be drawn from the correlation analysis for the delayed inputs in Table 1. The target variable presents the strongest correlations with the past values of the aFRR price band itself, the *marginal price in the day-ahead market*, *real-time generation wind*, *real demand* and *real-time generation by combined cycle gas turbines*, for analogous reasons to those previously discussed.

3.2. Datasets definition

The 32 variables of interest are split into three different datasets to determine which of their combinations present a higher influence in the forecasting models. The three datasets share a number of time series that are considered to be relevant enough to be always included. These are related to: energy allocated; reserve requirements and prices in the secondary and day-ahead markets; energy capacity matched and unmatched by the different types of generation sources; real demand; and generation by RES. Moreover, each one includes some additional variables of interest to evaluate their effect on the prediction model.

- *Dataset 1* - Its goal is to evaluate the effect of the demand deviations as well as power deviations derived from the scheduled programs registered by RES generation. Therefore, it includes the demand and the scheduled RES generation variables from both variable groups: *delayed* and *known*.
- *Dataset 2* - It focuses on adding to the base set of time series: the real-time generation from conventional sources and the net secondary reserve energy used (upward – downward)

from Table 1, as well as the total scheduled generation and the demand after the day-ahead resolution (PVP program) from Table 2.
- *Dataset 3* - The goal in this case is to evaluate if results for Dataset 2 can be improved by complementing its inputs with the variables related to the energy exchanged with the neighboring control areas (France and Portugal) from Table 1 and the disaggregated scheduled generation from each source after the day-ahead resolution (PVP program) from Table 2.

Finally, two years of data are available for each input variable. These comprehend the whole years 2019 (70% used for training and 30% used for validation) and 2020 (used as test data), taking into account that the period between 14th March 2020 and 30th June 2020 has been removed from test data due to the exceptional and unforeseeable market conditions caused by the COVID-19.

3.3. Metrics

Different performance metrics are used in the literature to evaluate and compare forecasting models. Among them, this study implements 5 options that are: the mean absolute error (MAE), the root mean squared error (RMSE), the mean absolute percentage error (MAPE), the mean absolute scaled error (MASE) and the root mean square scaled error (RMSSE). These are defined through Eqs. (3) to (7).

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|\widehat{y_t} - y_k| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\widehat{y_t} - y_k)^2} \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{k=1}^{N}\left|\frac{\widehat{y_t} - y_k}{y_k}\right| \cdot 100 \tag{5}$$

$$MASE = \frac{MAE}{MAE_{persistent}} \tag{6}$$

$$RMSSE = \frac{RMSE}{RMSE_{persistent}} \tag{7}$$

The MAE and RMSE metrics are used because they are the most common measure errors used in the electricity price forecasting literature [18]. However, as explained in [19], electricity profits and cost usually have a linear dependence regarding the electricity price. For that reason, absolute metrics such as MAE are more representative than RMSE in these cases. In addition, the squared term nature of RMSE penalizes large differences between (y_k) and $(\widehat{y_t})$ in a time step. Hence, this metric works better for short-term than for medium-term predictions, which is the case in this work.

The MASE, RMSSE and MAPE metrics are also used in this work to compare the results with those introduced in different references.

Finally, a Diebold–Mariano test [20] is used to check if the difference in MAE observed among the various models is statistically significant or not. This test consists of a hypothesis contrast where the null hypothesis means that the two forecasts have the same accuracy while the alternative hypothesis indicates a difference of performance.

4. Forecasting methods: model selection

Deep learning is a machine learning technique that has steadily increased in many research fields in the last decades and particularly in the electricity price forecasting domain [21]. With the goal to predict the secondary reserve marginal price of the power band assignment process in the Spanish aFRR service, various approaches have been considered in this work. These are based on: feedforward, convolutional and LSTM neural networks.

Within the daily trading window for the aFRR service market shown in Fig. 2, the price prediction is performed every day at 3 pm. This forecast provides the 24 hourly prices for the following day. Therefore, the prediction horizon for the different tested models ranges from $t + 9h$ to $t + 33h$ (where t is the instant when the prediction is made — every day at 3 pm). i.e., from 0 am to 11:59 pm on the day D+1 (24 prediction outputs).

Fig. 4 shows the data processing diagram defining the data organization and flow through each of the models. Note the two groups of input variables are used: '*Delayed*' and '*Known*'. A different model is analyzed for each of the three input datasets explained in Section 3 to evaluate how the different combinations of inputs influence the prediction.

Finally, the criterion to find the optimal hyperparameters for each model is to minimize, using the Adam optimizer, the MAE as a validation metric. Consequently, a random search is performed with the package Keras tuner [22].

4.1. Persistent approach

This type of model is based on the characteristic daily periodicity of the aFRR band price. It consists in assigning the value of the time series for the previous day at the same time as the predicted value for the target day.

This simple forecasting technique is used as a baseline in this work. Consequently, the MAE and RMSE obtained with it are used as a reference in order to compare the results obtained with the other approaches.

4.2. Feedforward neural network approach

Feedforward neural networks are ANN composed of a set of stacked dense layers that apply non-linear operations, learned in the training process, to the model inputs to obtain the outputs. They have already been used to predict prices in electricity markets such as the day-ahead [23–25].

Therefore, feedforward models are analyzed in this work. For this approach, the inputs are initially normalized and flattened. Then, they are concatenated to be introduced in the, up to 5, fully-connected layers. Following these dense layers (also known as hidden layers), a dropout is implemented during the weight optimization to avoid the over-fitting. Finally, since 24 price prediction values are desired, it only present 1 output layer with 24 output values and a rectified linear unit (ReLU) activation function, because negative prices were not an option for the years analyzed. The whole hyperparameters configuration and the validation metrics obtained for each tested model are collected in the feedforward section of Table 4. This model, as well as all the following ones, is programmed using Python 3.8 and Tensorflow 2.3.0 and using a laptop with an Intel Core i7 processor and NVIDIA GeForce RTX 3060 GPU. Observe that the best performance metric during the validation process is achieved using dataset 2.

Fig. 5 shows the complete structure of the best performing neural network obtained for with the validation data.

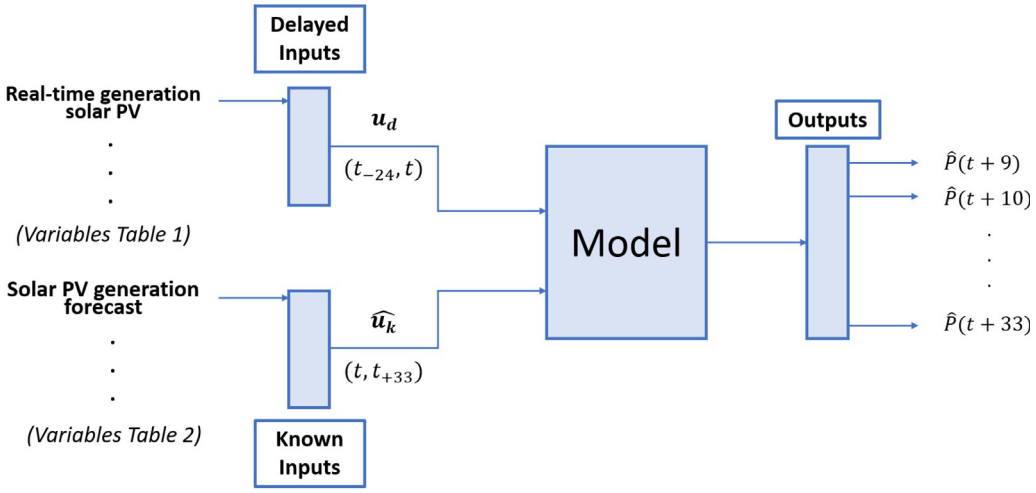


Fig. 4. Data processing diagram.

Table 3
Random search hyperparameters.

Hyperparameter	Type	Search values
1-D Conv-layers (only for convolutional models)	Choice	0, 1, 2
Convolutional filters (only for convolutional models)	Continuous	10 to 100 (step 10)
Convolutional activation (only for convolutional models)	Choice	ReLU, Linear
LSTM-layers (only for LSTM models)	Choice	0, 1, 2
LSTM units (only for LSTM models)	Continuous	10 to 200 (step 10)
LSTM activation (only for LSTM models)	Choice	ReLU, Linear
LSTM dropout (only for LSTM models)	Choice	0.0, 0.1, 0.2, 0.3, 0.4
Dense-layers	Continuous	1 to 6 (step 1)
Dropout	Choice	0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5
Layer 1,…,5 neurons	Continuous	10 to 250 (step 10)
Layer 1,…,5 activation	Choice	ReLU, Linear
Learning rate	Choice	10^{-2}, 10^{-3}, 10^{-4}

Table 4
Neural network hyperparameters and validation metrics.

	Feedforward			1-D convolutional			LSTM		
	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3
MAE validation	3.60	3.65	3.65	3.61	3.56	3.62	3.51	3.47	3.57
1-D CONV. Delayed	–	–	–	40 linear	40 linear	90 relu, 50 relu	–	–	–
1-D CONV. Known	–	–	–	20 linear	20 linear, 90 linear	40 linear	–	–	–
LSTM Delayed	–	–	–	–	–	–	70 relu	60 relu	70 relu
LSTM Known	–	–	–	–	–	–	–	–	–
LSTM Dropout	–	–	–	–	–	–	0.2	0.2	0.4
Dense-Layers	1	3	2	2	1	1	2	1	1
Dropout	0.2	0.3	0.2	0.1	0.2	0.0	0.0	0.1	0.1
Layer 1	220 relu	120 linear	40 linear	180 relu	80 relu	180 relu	140 linear	160 relu	100 relu
Layer 2	–	50 linear	130 relu	120 relu	–	–	100 relu	–	–
Layer 3	–	100 relu	–	–	–	–	–	–	–
Layer 4	–	–	–	–	–	–	–	–	–
Layer 5	–	–	–	–	–	–	–	–	–
Learning-rate	10^{-3}	10^{-2}	10^{-2}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-2}	10^{-3}

4.3. Convolutional neural network approach

Convolutional Neural Networks (CNNs) consist of a sequence of convolutional layers that apply a series of filters to the inputs to identify patterns. Moreover, they achieve good performances on noisy time series by extracting only the important and useful features of the correspondent input [26]. For this reason, they have been used in several works related with EPF in recent years [27,28].

The complete structure of the CNN model used for this work is represented in Fig. 6(a). Observe in this case how each type of input is introduced into a series of 1-D convolutional layers in order to identify the main patterns that influence the prediction. Once the features are extracted, they are introduced into a max pooling layer in order to reduce the number of parameters to be learned and the computational burden. After that, they are flattened and concatenated prior to feed them into a series of hidden layers used to assign the corresponding weights before obtaining the final prediction.

Analogously to the feedforward models, a random search is run to select the best feasible hyperparameters. The set of candidates are included in Table 3. Furthermore, a model per input dataset is also built and the best hyperparameters obtained together with the final validation MAE per model are reflected in the convolutional section of Table 4. Observe that the best performance metric for this type of neural network is obtained using the input dataset 1.

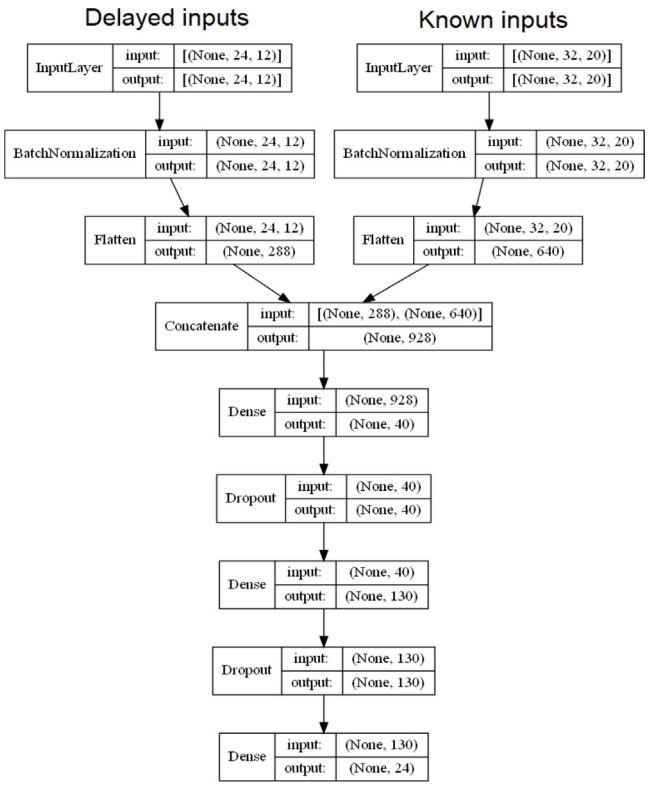


Fig. 5. ANN architecture for the feedforward model.

4.4. Recurrent neural network approach

Another third type of ANNs are the recurrent neural networks (RNN) [29]. This class of neural network is characterized by a series of internal loops that allow information to persist within the model [30]. Thus, their outputs depend on previous inputs and computations at each step, developing a 'memory' over previous events [31].

Among the different types of RNNs, the long short-term memory (LSTM) networks are the most widely deployed in the field of EPF because they can forget non-relevant information from the previous inputs and computations, increasing the training speed and storing important dependencies [32–34].

A LSTM approach is considered in this work. The resulting optimal architecture is summarized in Fig. 6(b). As in the other approaches, Table 3 collects the hyperparameter candidates for the random search. In this case, the search verified that the best network configuration only required one LSTM layer for the *Delayed* inputs, while none is required for the *Known* inputs. Once both inputs are flattened and concatenated, the resulting array is introduced in a series of fully-connected layers before the output matrix is obtained. After tuning, the best model configuration for each input set along with the final validation MAE are collected in Table 4. The best LSTM model for the validation data is obtained using the dataset 2.

5. Results and discussion

This section presents the results obtained after testing the different deep learning approaches using the test data with each dataset and allows comparing the accuracy of the various models. To do so, a Diebold–Mariano test is developed: first, comparing results among models develop for each approach; then, comparing approaches via their best registered models. The latter opens a final discussion among the different selected models to decide the best option to predict the marginal price of the secondary reserve in the Spanish aFRR market.

The comparative performance of the different approaches with each dataset is summarized in Table 5, which shows the results for the Diebold–Mariano test and the corresponding MAE values for each model. Assuming a confidence level of 95%, all p-values higher than 0.05 correspond to the acceptance of the null hypothesis (indicating similar performance between the compared models), while a p-value lower than 0.05 suggests different performance between models. Then, when the p-value is lower than 0.05, the model with the lower MAE is assumed to perform better and, if this is the dataset defined in the row, it is bold-highlighted in the table.

According to that, most of the p-values in Table 5 indicate different performances between models, with only 2 exceptions (0.06 between models with sets 2 and 3 in the convolutional approach and 0.7 also between models with sets 1 and 3 in the LSTM approach). In summary, the best model for the feedforward approach is the one using dataset 3; for the LSTM, that with the dataset 2. Finally, regarding the convolutional approach, both models with dataset 2 and 3 work better than that using dataset 1 while presenting similar accuracy between them. Among these two, the model with dataset 2 is selected because of its fewer trainable parameters, due to a lower number of inputs. Not surprisingly, the selected models coincide with those best performing with the validation data.

Once the best models for each approach are selected, Fig. 7 compares them by representing the real price on the X-axis and the forecast price on the Y-axis. The red diagonal line corresponds to the perfect prediction, while the blue circles represent the real predictions of the corresponding model, under/overestimating the price when they are below/above the red line. As expected, the persistent model is the worst predictor. The rest present very similar performances, as can be also derived from the proximity among MAE values in Table 5. The main conclusion from Fig. 7 is that all the models reduce the forecast accuracy as the aFRR service price increases, being this limitation especially visible for the persistent model. Since only 1% of the actual 2020 prices were above 30 €/MW, this fact does not significantly impact the performance metrics' values though.

Fig. 8 offers another way to compare the performance of the models. This figure shows the actual price (blue line) and the predicted price (orange line) for six consecutive days in December 2020. Observe the daily dependence of the prices that helps the persistent model to obtain reasonable predictions along those days, except for the high prices period registered on December 8th. Its resulting MAE for the period is 9.54 €/MW. Regarding the deep-learning approaches, all models predict low prices (below 15 €/MW) better than high prices (above 15 €/MW). Note that the largest prediction errors appear for prices above 40 €/MW experienced on December 8, while all models predict practically without error prices around 8 €/MW experienced on December 6. When comparing the MAE metrics obtained by each model for these consecutive days, hardly any differences can be seen. The LSTM model produces the lowest forecasting error, offering a MAE of 3.27 €/MW, while the feedforward and convolutional models yield 3.40 €/MW and 3.65 €/MW MAE errors, respectively.

The final comparison among models is done by means of Tables 6 and 7. On the one hand, Table 6 introduces the results of Diebold–Mariano test comparing the three approaches via their best (selected) models. The main conclusion drawn is that no model has a significantly superior performance in terms of MAE, since all the p-values obtained are above the 5% significance level.

On the other hand, Table 7 collects the performance metrics computed for each model together with their corresponding

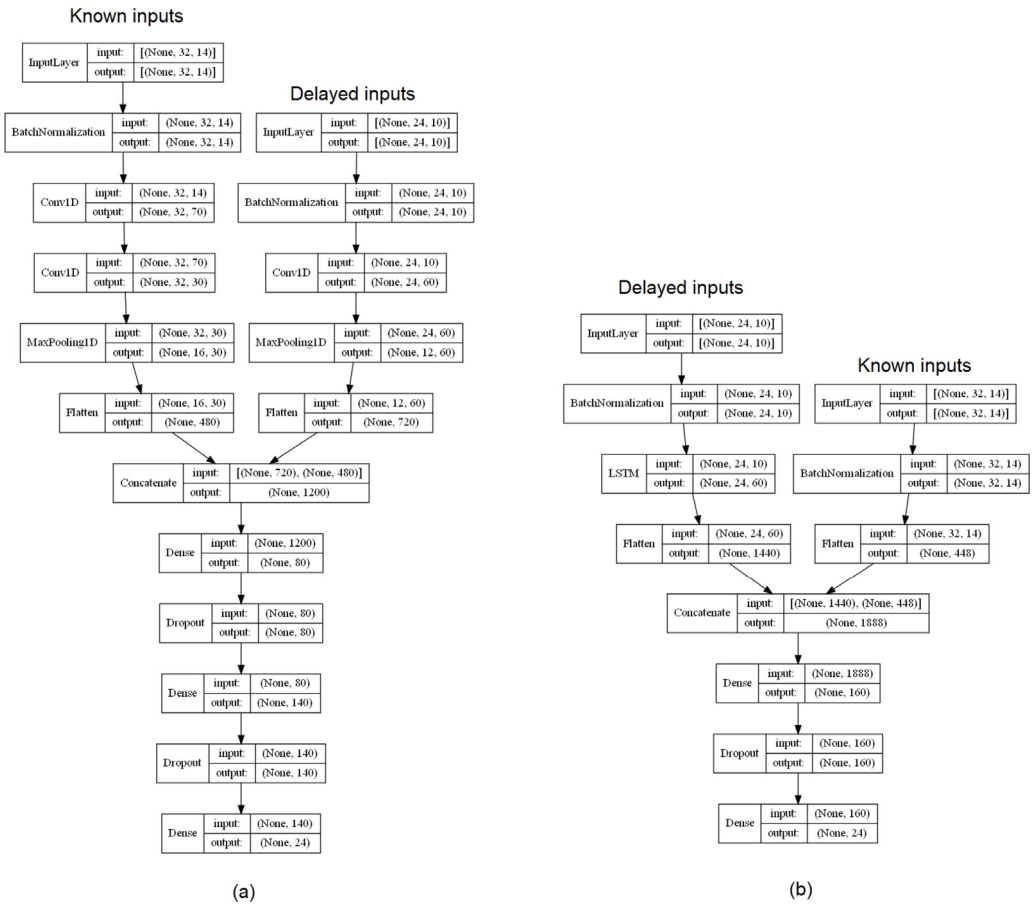


Fig. 6. ANN architectures for the 1-D Convolutional (a) and LSTM (b) models.

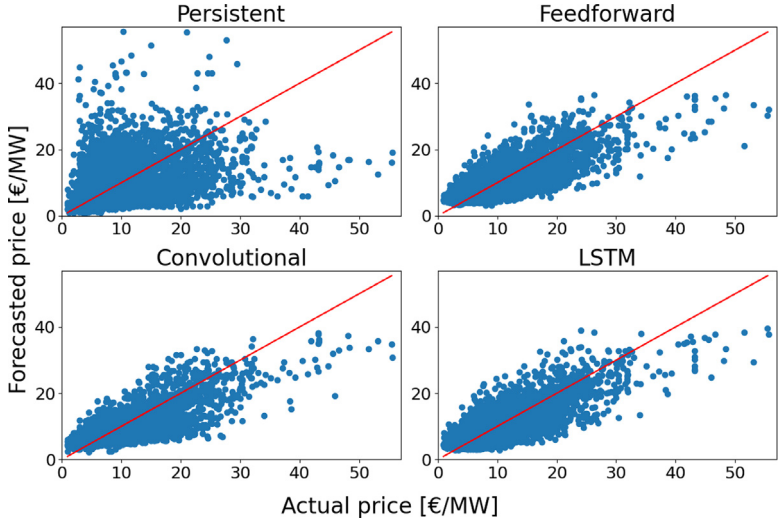


Fig. 7. Performance models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Diebold–Mariano test, p-values obtained when comparing the performance of the different models.

	Feedforward			Convolutional			LSTM		
	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3
Set 1	–	**0.002**	0.008	–	<0.001	0.006	–	<0.001	0.7
Set 2	0.002	–	<0.001	**<0.001**	–	0.06	**<0.001**	–	**<0.001**
Set 3	**0.008**	**<0.001**	–	**0.006**	0.06	–	0.7	<0.001	–
MAE	2.70	2.78	**2.64**	2.75	**2.66**	2.70	2.85	**2.63**	2.84

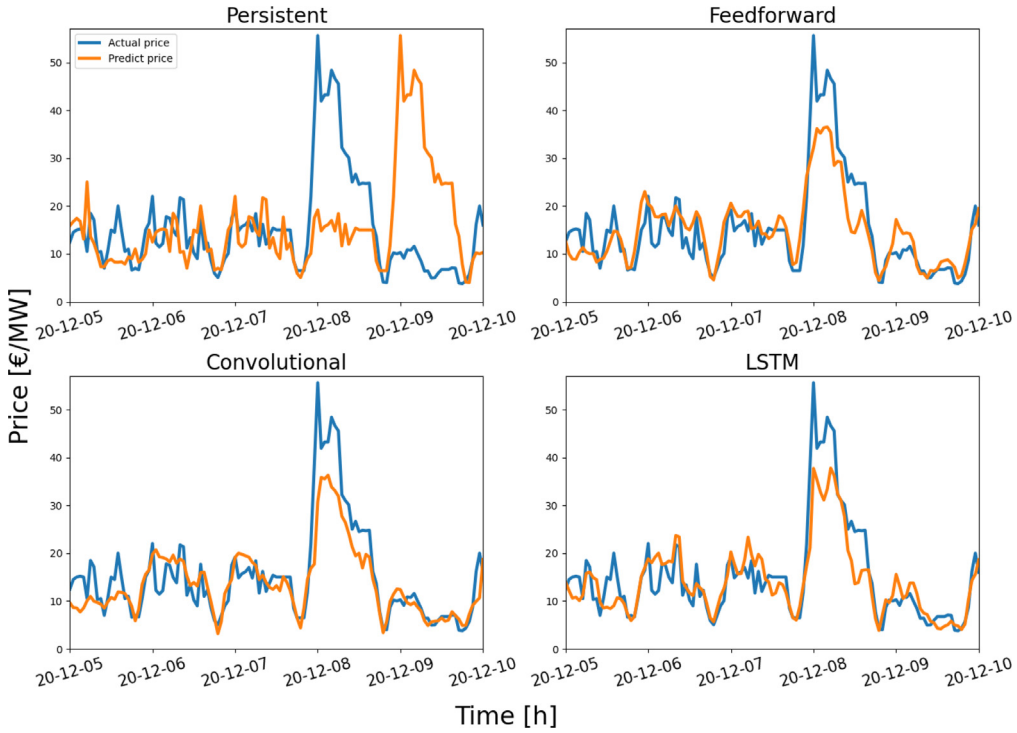


Fig. 8. Time evolution of actual and predicted price. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Diebold–Mariano test, p-values obtained when comparing the performance of the selected models from each approach.

	Feedforward	Convolutional	LSTM
Feedforward	–	0.30	0.77
Convolutional	0.30	–	0.25
LSTM	0.77	0.25	–
MAE	2.64	2.66	2.63

training times. Note that these metrics are calculated on the time series resulting from concatenating the hourly predictions made everyday at 3 pm for the whole test period. These results are summarized in Table 7 and confirm the previous conclusions:

- The persistent model returns the worst performance metrics.
- The deep learning models experience similar performance metrics among them, being the LSTM approach the one with slightly lower error metrics (63.22% of MASE, 55.28% of RMSSE and 35.50% of MAPE).
- There are hardly any differences in terms of training times. The LSTM model takes a longer time to complete the training process (4 min), but both the feedforward and convolutional models also employ training times around 3 min (3 min 05 s and 3 min 19 s, respectively). Therefore, training time is not a critical aspect to conclude which is the best model.

Finally, the goodness in the performance of the models is analyzed by comparing the resulting values of the metrics with those published by similar works found in the literature:

- Chazarra et al. [15], in their paper discussed in the introduction, calculate the accuracy of their models by means of the MAPE. The resulting values these authors present are lower than the MAPE results achieved with the models proposed in this paper. However, comparison among proposals is somehow risky and difficult to complete owing to the huge market evolution experienced in the last decade in Iberia. The paper by Chazarra et al. [15] used data from years 2013 and 2014 when the main technologies affecting the aFRR band price represented a much reduced share of the generation mix. Hence, the aFRR band price variability was much lower at the time. When persistent models are compared for years 2014 and 2020, MAPE results are very different (22.60% Vs. 50.90%), what confirms the significantly higher predictability of the data in 2014.
- Camal et al. [13] use a Support Vector Regression (SVR) forecasting technique to predict aFRR prices from the German aFRR market. They obtain a MASE metric of 84% for the *aFRR upward and downward average price* variable. Our work outperforms their results using the described deep learning models that provide MASE values significantly lower than those published by these authors.
- Merten et al. [14] analyze several forecasting techniques (including feedforward and LSTM models) to predict marginal mixed prices in the German aFRR market. This market is slightly different from the Spanish market because it allows only six time slices of 4 h per day and both energy generation (positive) and consumption (negative) are tendered separately. Thus, the best MASE obtained in their work (calculated as the division of each MAE by the MAE metric of their persistent model) is 76% for the product '*NEG_12_16*' when using a RNN model. Therefore, our work improves their results probably due to: the amount of data analyzed, the neural network architectures proposed, and the type and resolution of the exogenous data used.

6. Conclusion

This work analyzed the capability of three deep learning-based forecasting approaches (feedforward, convolutional, and LSTM),

Table 7
Results summary for all models on the secondary reserve marginal price.

Model	Set	MAE	RMSE	MASE (%)	RMSSE (%)	MAPE (%)	Training times
Persistent	–	4.16	6.53	100%	100%	50.90	–
Feedforward	Set 3	2.64	3.67	63.46%	56.20%	36.70%	3′05″
Convolutional	Set 2	2.66	3.70	63.94%	56.66%	36.00%	3′19″
LSTM	Set 2	2.63	3.61	63.22%	55.28%	35.50%	4′00″

adding up nine different models, to predict the secondary reserve marginal price of the power band assignment process in the Spanish aFRR market. The data provided to the models was divided into three datasets, which combined 32 variables of interest recorded by the Spanish TSO and the Iberian market operator. The hyperparameters configurations of the models was set up through a random search. Various validation metrics together with the Diebold–Mariano test were used to test the accuracy and choose the best model.

Although using so many different market variables, most of them presenting significant correlations with the aFRR band price, none of the three grouping input datasets allows providing forecast results with much improved accuracy. According to the Diebold–Mariano test, results show that dataset 2 comprehends the most significant combination of inputs for both convolutional and LSTM models, but dataset 3 helps the feedforward model to provide marginally better forecasts. In practice, this implies no information captured by the different datasets is clearly superior to the rest for the aFRR band price prediction. In fact, adding more variables to the dataset does not imply getting better results, as it is derived from the resulting MAE observed for the convolutional and LSTM models.

Out of the best models selected for each approach, results obtained with the test data confirm that the three models perform similarly, being in the range going from 63% to 64% in terms of MASE. The Diebold–Mariano test confirms these results, demonstrating that no model has a significant higher performance.

Finally, further work on this topic will focus on the implementation of these same forecasting techniques on other market variables, such as aFRR energy requirements. The final goal is to develop new decision-making algorithms that will allow analyzing the potential and optimal participation of batteries in the Spanish aFRR market. Also, the future market reorganization to be experienced in the coming months within the European context, supposedly targeting an increase in the granularity of the energy products traded within the ENTSO-E region, will open the possibility to test and improve these models within the new framework and beyond the current results. Other machine learning algorithms beyond ANN, such as SVR, KNN, or Random Forest techniques will equally be considered for comparison in future works.

CRediT authorship contribution statement

Javier Cardo-Miota: Methodology, Software, Resources, Writing – original draft, Writing – review & editing, Data curation. **Emilio Pérez:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization, Validation. **Hector Beltran:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization, Validation.

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.

Acknowledgment

This work was supported by the Spanish State Research Agency under grants PID2021-125634OB-I00 and TED2021-130120B-C22, ERDF, EU; and the Universitat Jaume I through project number UJI-B2021-35 and grant number PREDOC/2020/35.

References

[1] IEA, Net Zero by 2050, Technical Report, IEA, Paris, 2021.
[2] IEA, Renewables share of power generation in the Net Zero Scenario, 2010-2030, Technical Report, IEA, Paris, 2022.
[3] J.L. Prol, K.W. Steininger, D. Zilberman, The cannibalization effect of wind and solar in the california wholesale electricity market, Energy Econ. 85 (2020) 104552, http://dx.doi.org/10.1016/j.eneco.2019.104552.
[4] S. Homan, S. Brown, An analysis of frequency events in great britain, Energy Rep. 6 (2020) 63–69, http://dx.doi.org/10.1016/j.egyr.2020.02.028.
[5] K. Oureilidis, K.N. Malamaki, K. Gallos, A. Tsitsimelis, C. Dikaiakos, S. Gkavanoudis, M. Cvetkovic, J.M. Mauricio, J.M.M. Ortega, J.L.M. Ramos, G. Papaioannou, C. Demoulias, Ancillary services market design in distribution networks: review and identification of barriers, Energies 13 (2020) http://dx.doi.org/10.3390/en13040917.
[6] M.R. Rapizza, S.M. Canevese, Fast frequency regulation and synthetic inertia in a power system with high penetration of renewable energy sources: optimal design of the required quantities, Sustain. Energy, Grids Netw. 24 (2020) 100407, http://dx.doi.org/10.1016/j.segan.2020.100407.
[7] M. Bahloul, M. Daoud, S.K. Khadem, A bottom-up approach for techno-economic analysis of battery energy storage system for Irish grid DS3 service provision, Energy 245 (2022) 123229, http://dx.doi.org/10.1016/j.energy.2022.123229.
[8] D.F.-M. noz, J.I. Pérez-Díaz, I. Guisández, M. Chazarra, A. Fernández-Espina, Fast frequency control ancillary services: an international review, Renew. Sustain. Energy Rev. 120 (2020) 109662, http://dx.doi.org/10.1016/j.rser.2019.109662.
[9] C. Brivio, S. Mandelli, M. Merlo, Battery energy storage system for primary control reserve and energy arbitrage, Sustain. Energy, Grids Netw. 6 (2016) 152–165, http://dx.doi.org/10.1016/j.segan.2016.03.004.
[10] Y.-Y. Hong, J.V. Taylar, A.C. Fajardo, Locational marginal price forecasting in a day-ahead power market using spatiotemporal deep learning network, Sustain. Energy, Grids Netw. 24 (2020) 100406, http://dx.doi.org/10.1016/j.segan.2020.100406.
[11] R. Beigaite, T. Krilavičius, K.L. Man, Electricity price forecasting for nord pool data, in: 2018 International Conference on Platform Technology and Service (PlatCon), 2018, pp. 1–6, http://dx.doi.org/10.1109/PlatCon.2018.8472762.
[12] G. Memarzadeh, F. Keynia, Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm, Electr. Power Syst. Res. 192 (2021) 106995, http://dx.doi.org/10.1016/j.epsr.2020.106995.
[13] S. Camal, A. Michiorri, G. Kariniotakis, Optimal offer of automatic frequency restoration reserve from a combined PV/Wind virtual power plant, IEEE Trans. Power Syst. 33 (2018) 6155–6170, http://dx.doi.org/10.1109/TPWRS.2018.2847239.
[14] M. Merten, F. Rücker, I. Schoeneberger, D.U. Sauer, Automatic frequency restoration reserve market prediction: Methodology and comparison of various approaches, Appl. Energy 268 (2020) 114978, http://dx.doi.org/10.1016/j.apenergy.2020.114978.
[15] M. Chazarra, J.I. Pérez-Díaz, J. García-González, A. Helseth, Economic effects of forecasting inaccuracies in the automatic frequency restoration service for the day-ahead energy and reserve scheduling of pumped storage plants, Electr. Power Syst. Res. 174 (2019) 105850, http://dx.doi.org/10.1016/j.epsr.2019.04.028.
[16] REE, ESIOS, https://www.esios.ree.es/es.
[17] OMIE, Omie, https://www.omie.es/.
[18] R. Weron, Electricity price forecasting: a review of the state-of-the-art with a look into the future, Int. J. Forecast. 30 (2014) 1030–1081, http://dx.doi.org/10.1016/j.ijforecast.2014.08.008.

[19] J. Lago, G. Marcjasz, B.D. Schutter, R. Weron, Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark, Appl. Energy 293 (2021) 116983, http://dx.doi.org/10.1016/j.apenergy.2021.116983.
[20] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econ. Stat. 20 (1) (2002) 134–144.
[21] J. Lago, F. De Ridder, B. De Schutter, Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms, Appl. Energy 221 (2018) 386–405, http://dx.doi.org/10.1016/j.apenergy.2018.02.069.
[22] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, et al., Keras Tuner, 2019, https://github.com/keras-team/keras-tuner.
[23] H.-T. Pao, Forecasting electricity market pricing using artificial neural networks, Energy Convers. Manag. 48 (2007) 907–912, http://dx.doi.org/10.1016/j.enconman.2006.08.016.
[24] P.M. Bento, J.A. Pombo, S.J. Mariano, M.R. Calado, Short-term price forecasting in the iberian electricity market: sensitivity assessment of the exogenous variables influence, in: 2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 2022, pp. 1–7, http://dx.doi.org/10.1109/EEEIC/ICPSEurope54979.2022.9854716.
[25] I.P. Panapakidis, A.S. Dagoumas, Day-ahead electricity price forecasting via the application of artificial neural network based models, Appl. Energy 172 (2016) 132–151, http://dx.doi.org/10.1016/j.apenergy.2016.03.089.
[26] A. Borovykh, S. Bohte, C.W. Oosterlee, Conditional Time Series Forecasting with Convolutional Neural Networks, 2018, arXiv:1703.04691.
[27] M. Zahid, F. Ahmed, N. Javaid, R.A. Abbasi, H.S. Zainab Kazmi, A. Javaid, M. Bilal, M. Akbar, M. Ilahi, Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids, Electronics 8 (2) (2019) http://dx.doi.org/10.3390/electronics8020122.
[28] C. Zhang, R. Li, H. Shi, F. Li, Deep learning for day-ahead electricity price forecasting, IET Smart Grid 3 (2020) 462–469, http://dx.doi.org/10.1049/iet-stg.2019.0258.
[29] H. Hewamalage, C. Bergmeir, K. Bandara, Recurrent neural networks for time series forecasting: current status and future directions, Int. J. Forecast. 37 (2021) 388–427, http://dx.doi.org/10.1016/j.ijforecast.2020.06.008.
[30] C. Olah, Understanding LSTM networks, 2015, http://colah.github.io/posts/2015-08-Understanding-LSTMs/.
[31] F.M. Bianchi, E. Maiorino, M.C. Kampffmeyer, A. Rizzi, R. Jenssen, Recurrent Neural Networks for Short-Term Load Forecasting, Springer International Publishing, 2017, http://dx.doi.org/10.1007/978-3-319-70338-1.
[32] L. Jiang, G. Hu, Day-ahead price forecasting for electricity market using long-short term memory recurrent neural network, in: 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2018, pp. 949–954, http://dx.doi.org/10.1109/ICARCV.2018.8581235.
[33] A. Meng, P. Wang, G. Zhai, C. Zeng, S. Chen, X. Yang, H. Yin, Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization, Energy 254 (2022) 124212, http://dx.doi.org/10.1016/j.energy.2022.124212.
[34] Z. Chang, Y. Zhang, W. Chen, Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform, Energy 187 (2019) 115804, http://dx.doi.org/10.1016/j.energy.2019.07.134.