UNIVERSIDAD PONTIFICIA COMILLAS

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

OFFICIAL MASTER'S DEGREE IN THE ELECTRIC POWER INDUSTRY

MASTER'S THESIS

Spanish electricity Wholesale market analysis of price sensitivity, based on the consideration of aggregated matching curves, with respect to the main potential drivers.

Autor: Marta del Pozo Santolaya
Supervisor: Juan Bogas Gálvez

Madrid, 16 Julio 2021

Official Master's Degree in the Electric Power Industry (MEPI)

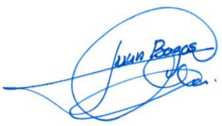
Master’s Thesis Presentation Authorization

THE STUDENT

Marta del Pozo Santolaya

THE SUPERVISOR

Juan Bogas Gálvez



Signed: Date: 16/07/ 2021

Authorization of the Master’s Thesis Coordinator

Dr. Luis Olmos Camacho

Signed: Date: 16/ 07/ 2021

Summary

Due to the new circumstances that the Spanish electricity system is undergoing, and it is expected to face in the future, hourly prices may vary substantially. The goal of this master thesis is to deeply analyze the main drivers of the price sensitivity in the last three years. Some of these potential drivers, are the different generation technologies that are having more impact in the price fluctuations, or other factors like carbon price which it seems, based on the actual trends, that it will be a very important variable in the electricity market.

The effect of these variables in the price variation is going to be analyzed for the different periods of the year due to the fact that some of them have strong seasonality. Furthermore, zero emission technologies, with capability of energy storage will play an important role in the near future. They will help to face the uncertainty that a system with high integration of renewables will have. The different factors that may affect the dispatch of these kind of energy sources will be discussed.

Once the price drivers effect is demonstrated, another method will be used in order to joint together the impact of all the variables in one equation. Multiple linear regression models (MLRM) will be computed to evaluate the adjustment of the variables to the price variation and an accuracy analysis will be done to detect the lack of adjustments. Finally, the multiple linear regression models will be tested with external data of the current year.

This energy price sensitivity analysis will help to understand the current situation of the Spanish electricity market. On the other hand, we will discuss how the expected changes based on the goals related to the transition towards a more clean and efficient system may change the market trends in the upcoming years. Furthermore, another contribution of this project is to introduce some new variables that have been computed and different ways to differentiate the data along the year in order to improve the adjustment of the forecast.

Table of content

List of tables

List of figures

1. Introduction

1.1 Motivation

The day ahead market also called single day-ahead coupling in Spain (SDAC), is the most representative electricity market in Spain because of the amount traded. It aims to perform electrical energy transactions. A supply and demand curve are constructed for each hour based on the supply and demand bids, the matching point between the sell and purchase bids of market agents for the next 24 hours of the following day will be the price for each specific hour[1].
The price and volume set for each hour of the following day it is set by a methodology done by most of European countries. Based on the Euphemia algorithm the price is set, considering the aggregated matching curves of demand and supply constructed based on the market agents' bids. The algorithm does all this process in a way to achieve the maximization of the social welfare.

The supply agent's bids are based on the operational cost of the technology that they will used to generate the electricity, and the price will change depending on it because of the price criteria that will be explain later. Because of this, the recent changes in the technology mix and the future variations in order to achieve the objectives set in the PNIEC (Plan Nacional de Energía y Clima), will trigger several important effects in the electricity market functioning.
The introduction of more low variable cost energy sources, the increased of cross border exchanges in order to fulfil the objectives of the EU policies about security of supply, interconnected energy networks and the achievement of a full renewable energy electricity system, will make the market to suffer significative changes[2]. Also, the development of storage technologies will lead to a need of a better understanding of the price dynamics.

Technologies with the capacity of store energy will play an important role. The reason it is because pumped hydro or hydro reservoirs will provide the security of supply that will be needed in the short term. Therefore, knowing the variables that affect the agent's behavior for the dispatching of these technologies will help in the way to understand the price formation.

---

[1] Omie.es. 2021. *Electricity market | OMIE*. [online] Available at: https://www.omie.es/en/mercado-de-electricidad.

[2] IEA. 2021. *European Union 2020 – Analysis - IEA*. [online] Available at: https://www.iea.org/reports/european-union-2020.

It is known that renewables sources have an effect of price reduction, but the question of up to how many MWh these technologies are having an effect in the market is an interesting question that would be consider in this project. Some studies have shown that variables like solar photovoltaic has a different effect on prices once it crosses a specific value. Also, it is said that wind has larger merit order effect than solar.[3]
In addition, carbon price will be also a potential driver of the electricity price based on the current situation of this emission trading market. All of these assumptions are going to be study in this master thesis for the Spanish system.

1.2 Objectives and structure

The objectives of this master thesis are the following:

- Developed a correlation analysis of the potential drivers of the price sensitivity in the Spanish case.
- Analyze the specific case of each of the technologies present in the system based on its technical characteristics. The factors that affect the dispatch and the variation of them throughout the year.
- Based on the estimated variations of the renewable technologies install capacity in the system, compare the actual effect with the expected future one.
- Finally, demonstrate that the variables that have been analyzed in the correlation study and the ones that have been computed are really the price formation drivers.

The project will start with a description of the Spanish electricity Wholesale Market in order to understand how the spot electricity prices that are used in this mater thesis are determine.
Secondly, an explanation of how and why the price change due to the increased of the share of some technologies in the system will be included. Moreover, some actual information of the marginal price- setting technologies that set the price in Spain will be provided.
Afterwards, a description of the problem and the methodology that has been selected will be explain as well as the mathematical fundamentals of the analysis.
The analysis will start with the main driver's correlation analysis of the hourly data from 2018 to 2020, to be followed by the multiple linear regression equations results and the simplification study of these equations.

---

[3] Mosquera-López, S. and Nursimulu, A., 2019. Drivers of electricity price dynamics: Comparative analysis of spot and futures markets. *Energy Policy*, 126, pp.76-87.

At the end, an accuracy analysis it is going to be computed to evaluate if the equations are able to explain the price and, we will conclude with the test of the multiple regression models with data of the year 2021.
The data it is in hourly values so there are around 26000 values of each of the variables.

2. State of the art

2.1 General description of the actual market functioning

In this chapter the market functioning of the Iberian electricity market it is going to be explained.
The electricity market is the wholesale market in which the electricity producing agents sell electricity to consumers. This market includes both the energy that is sold through physical bilateral contracts and the energy that is sold in the organized or pool market, where practically all the transactions are carried out.
The organized or pool market is structured on the basis of a series of successive markets in which the supply and demand of electricity is adjusted, day ahead, intraday and continuous market.
The Iberian electricity market is an aggregation of different markets, contracts and process:

- Different markets:
  - Day-ahead market
  - Intraday market
  - Balancing markets
- Bilateral contracts
- Processes: technical constraints, deviation managements

The most representative one because the amount of energy negotiated is the Day - ahead, which is a physical market where energy is traded in hourly basis for the next day. In this master thesis the study is going to be done for the prices of the day - ahead, specifically the PDBF which is the acronym of Programa Diario Base de Funcionamiento.

At the same time the intraday market it is performed in order to enable the demand and supply to adjust their positions and also the technical restrictions and complementary services, carried out by the system operator. Thanks to these adjustment markets, the system can accommodate the unbalances between actual and forecast demand and generators can optimize their operating programs given the technical characteristics of their plants. In addition, the bilateral contracts signed in long term future must be taken into account.
The time sequence of operation of the daily market can be seen in the following figure, obtained from OMIE, where D-1 is the day on which the price is set for day D.

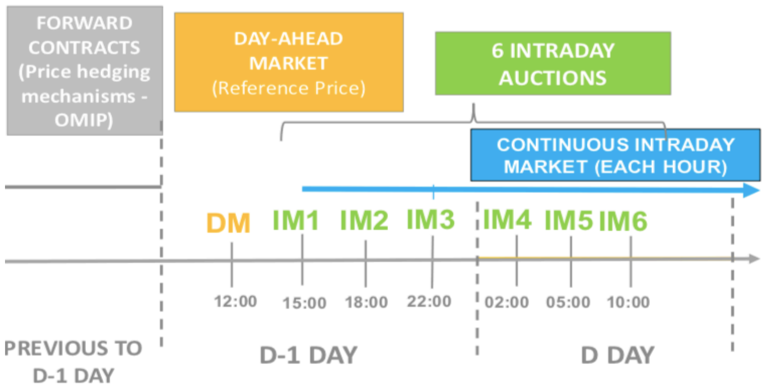


Figure 1. Iberian wholesale market structure. (Source: OMIE)

The objective of the daily market is to define the price and quantities of energy that producers are going to inject into the electricity grid and consumers are going to absorb from it during a certain hour.

The price and volume of energy in a given hour are established by the crossing point between supply and demand, following the marginalist model approved by the European Union, based on an algorithm adopted by all EU markets (EUPHEMIA). The point where both curves match it is the market price.

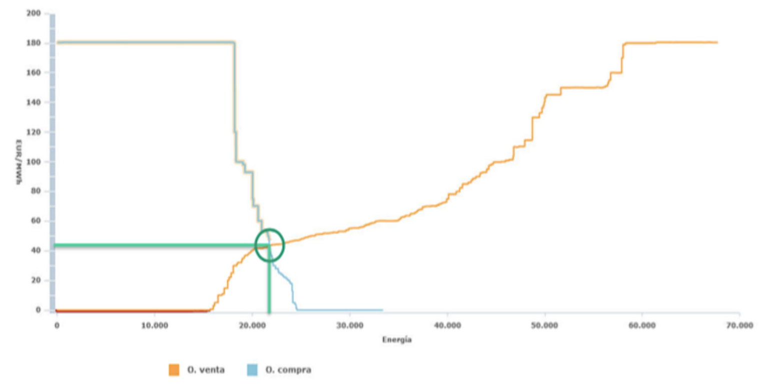


Figure 2. Example of the supply and demand curve in one hour of the Spanish electricity market. (Source: OMIE).

Once OMIE has received all the offers from the producers and consumers, it performs, for each hour, the aggregate supply and demand curves, ordering, by sections from lowest to highest, all generation offers and by sections from highest to lowest all takeover bids.

By observing an aggregate supply curve, the sections of offers at similar prices can be differentiated, which it usually corresponds to certain technologies.

The first section corresponds to technologies that offer their production at zero Euro/MWh. These technologies are usually nuclear, wind, photovoltaic and

hydroelectric reservoir plant with capability of generation schedule. The reason why these plants offer at zero price is different from nuclear and renewables.
Nuclear power plant technology does not have the capacity to vary its level of production over time. These types of plants must operate at their nominal power all times (except in periods of uranium replacement, maintenance or emergency shutdowns). For this reason, they offer their energy at zero price to ensure that they match and to keep their production level constant, allowing the price they receive in return to be set by other technologies that can bid at higher prices because they have the ability to be able to decide when to produce. Nuclear dispatch for the next day can be known easily as the maintenance or shutdowns are schedule.
In the case of renewable energy, their fuels cost is zero (wind solar or water does not have real cost) and their operational cost it is very low, therefore, they bid at zero most of the times and in some situations, when there is a very high renewable generation they bid at their operational cost.
With large reservoir with enough capacity to storage the water, this situation does not occur but it is important to pointed out that hydro power plants have to fulfil a several minimum hydrological requirements so they will have to mandatory schedule a number of MWh at instrumental price. If the forecast for the next hour is favorable for wind production, the wind farm agents will bid at zero in order to ensure that they enter into the market and as a result they make profits. In the case of solar energy and small hydroelectric plant like run of the river the same thing happens.
The second section is mainly made up of combined cycle plants. These types of power plants determine their offer prices taking into account various factors: the price of their fuel (gas, oil, coal), the available stock, the evolution future prices forecast and also some other operational costs such the start-up and shut-down cost.

Finally, the third section is dominated by adjustable hydroelectric plants with enough reservoir capacity to storge water. They will bid at the combine cycle price in order to maximize it benefits. It can be added to this part as they can decided when to produce based on their opportunity cost that can be calculated taking into account the expected bid, mostly based on MIBGAS and carbon price, of the combine cycle plants, cogeneration and some carbon power plants. The price of adjustable hydroelectric plants is determined by weather conditions, the capacity of the reservoir and future prices of OMIP. It is important to consider that there is an amount that has to be mandatory schedule in order to fulfill the minimum hydrological requirement.
The way how it works it is the following: if the reservoir is at the limit of its capacity and it is necessary to evacuate water, the plant will make offers at a very low price in order to ensure "matching" into the market. On the order hand, if the level of reserves is low, they will make offers at a very high price to ensure that it used the water in exchange for a large remuneration (water value), so it can afford not to produce while waiting for

a future situation of higher price. When the reservoir is at intermediate levels, it will take into account all the factors to maximize profits.
Because of this most of the hours the technologies that set the price are either CCGT or hydroelectric plants.

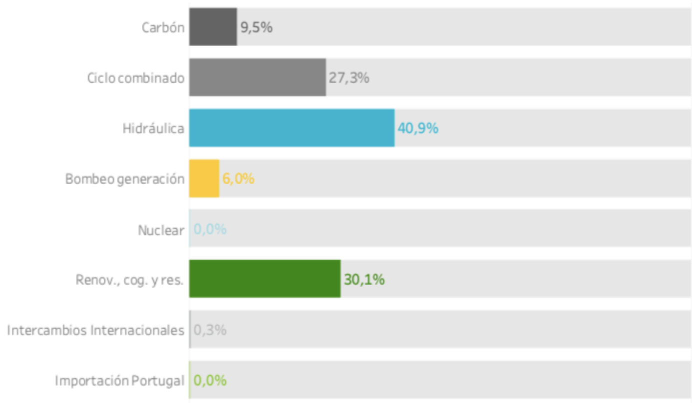


Figure 3. Percentage of hours that each technology is setting the price in 2019. (Source: OMIE)

2.2 Price formation: The Merit Order Effect

In order to match in the market, agents bid a lower price than the expected clearing one in order to ensure that they are going to be scheduled. In Europe, the spot electricity price of all the electricity markets is joint through a market coupling mechanism, were bidding areas with lower prices export electricity and areas with higher prices import. If the interconnection capacity is enough to exchange energy at competitive prices and there is not limit, the price in every region will be the same.
In the generator side what happens is the merit -order effect. Suppliers depend on the marginal cost that each agent is bidding in the spot market. This marginal cost it is characteristic of each type of technology, it depends on the operational cost. The operational cost can present different components depending on the energy source, most of them have components like fuel price, maintenance cost or emission and supply cost.[4]
As their bids depend on the operational cost, the cheapest units will be in the bottom of the curve like renewable and non-manageable technologies like nuclear or run of river.
To sum up, in the bottom of the curve we will find the bids corresponding to technologies with the lowest operational cost and also the ones that have to be committed independently from the marginal cost, like nuclear or cogeneration, which

[4] Figueiredo, N. C. and Silva, P. P., 2018. “The price of wind power generation in Iberia and the merit-order effect”, *International Journal of Sustainable Energy Planning and Management*, 15, pp. 21–30.

have to fulfill operational requirements of production. Therefore, the highest the energy coming from these low variable cost technologies the lowest will be the price, the curve will shift to the right. In the figure below the change can be observed.

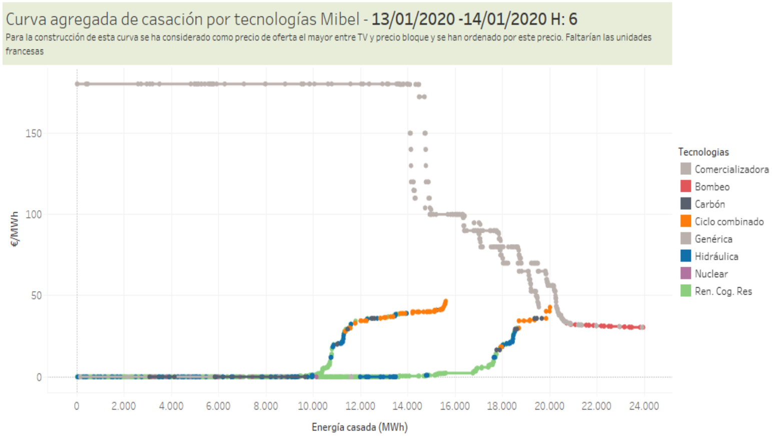


Figure 4. Example of supply curve shift to the right due to the introduction of renewables.

As we can see the introduction of more renewable which correspond to the green dots make the supply curve to move to the right. The figure also illustrates the decrease of the combine cycle plant in the system (orange dots), having as a result a decrease of the price.

The integration of more renewables in the system, in order to fulfill the objectives of the PNIEC (74% of RES-E generation share in the system by 2030[5]), will have apart from the technical impacts it will also have market design impacts such electricity market integration, cost allocation of transmission grid or capacity support mechanisms. Moreover, the depressed spot electricity prices together with what it was just mention will lead to a difficult situation for utilities.

In this context a deep study of the effect of the variables that are going to be more representative in the system will be very useful.

[5] España, Ley Orgánica 21/2013, de 9 de diciembre, de evaluación ambiental. *Boletín Oficial del Estado*, 11 de enero de 2021, núm. 9, p. 2681 a 2741. (https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-421).

3. Problem setting and description of the problem

Electricity price have some characteristic that make the forecast very difficult such not stationarity and high volatility. Because of this, it is also important to make a comprehensive analysis of the variables that are going to be used and that can explain even partially the electricity prices.
The existing studies about the variables that may be able to predict the dynamic of electricity prices is huge. Some examples of variables that have been used are power consumption and load profiles[6] or the effect of technologies, market power network congestion and demand.[7]

Different types of forecast techniques have been raised in the last years like Nash - Cournot framework, models based on computational intelligence such recurrent neural network and statistical based models such GRACH or regression models[8] which is the one that will be used. We have decided to use this statistical technique due to its capability to detect the relation between the predictor variables and the dependent variable. Moreover, another advantage that has been found in the literature was that it allows to interpret physical characteristics of the price formation.[8]

In this project we are going focus on variables than can be easily obtain the day before of the price clearing. Wind, solar, demand forecast and the expected gap which will include the difference between the forecast demand and the solar, wind and nuclear generation which can be known in advance because the variation of its production it is due to schedule maintenance or shutdowns. Also, energy exported and imported to France of the day before (D-1) as the direction of the exchanges does not suddenly change, it usually follows a smooth behavior. Other prices such MIBGAS, CO2 and future prices in order to include in some way the effect of hydroelectric generation as well as the capacity of the reservoirs as it is going to be explained later on. The following table summarize the variables known in the day D-1 before the market clearing in the day D that are going to be used in the study.

[6] Weron, R. and Misiorek, A., 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24(4), pp.744-763.

[7]Gianfreda, A. and Grossi, L., 2012. Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, 34(6), pp.2228-2239.

[8] Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), pp.1030-1081.

The following table summarize the variables that are going to be studied, the units and the sources.

Table 1. Variables used in the study.

Variable	Unit	Data Source
Wind generation forecast	MWh	REE
Solar generation forecast	MWh	REE
Demand forecast	MWh	REE
France import (D-1)	MWh	OMIE
France export (D-1)	MWh	OMIE
Thermal gap	MWh	OMIE
MIBGAS	Euro/MWh	MIBGAS
CO_2 price	Euro/ton	Refitinv
Capacity Index	none	MITECO/ own elaboration
Y+1	Euro/MWh	OMIP
Q+1	Euro/MWh	OMIP
M+1	Euro/MWh	OMIP

Before deciding the used of the variables presented in *Table 1*, a preliminary work was done. Some potential drivers like solar and wind generation of the day before (D-1) and coal prices were discarded. The first one because the relation with the generation of the day D was not very strong and it was given inconsistent results. On the other hand, coal price was not used due to the very low and decreasing presence of this technology in the system.

4. Proposed methodology and mathematical fundamentals

The analysis of the characteristic of each variable and its relationship with electricity prices it is going to be done using correlations. In this chapter the mathematical fundamentals are going to be explained. All the calculation has been done using the software RStudio and Microsoft Excel.

The linear relation between variables it is usually obtained through the Pearson Correlation value. It is a method that enable to know the association, the degree at which to variables variate together. This coefficient, designed for quantitative variables, is an index that measures the degree of covariation between different linearly related variables. The Pearson correlation coefficient is a value that is easy to perform and also easy to interpret. Its absolute values oscillate between 0 and 1. This means that if we have two variables X and Y, and we define the Pearson correlation coefficient between these two variables as r_{xy} then, $0 \leq r_{xy} \leq 1$.
If the sign is considered, the Pearson correlation coefficient oscillates between –1 and +1. However, it should be noted that the magnitude of the relationship is specified by the numerical value of the coefficient, the sign reflecting the direction of such value. In this sense, a relationship of +1 is as strong as -1.
It can be said that two variables have a perfectly positive correlation when one increase in the same magnitude as the other.

The following equation describes how it is calculated:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{XY}}}$$

Where S_{xy} is the sample covariance, and S_x *and* S_y*,* are sample standard deviations.[9]

The forecast methodology in this master thesis was based on a statistical approach. This methodology tries to determine the actual price based on data from the past the prices and also the value of the different independent variables. An advantage of using this type of model based on exogenous or independent variables is that it allows to

[9] Weher, E., 1977. Edwards, Allen, L.: An introduction to linear regression and correlation. (A series of books in psychology.) W. H. Freeman and Comp., San Francisco 1976. 213 S., Tafelanh., s 7.00. *Biometrical Journal*, 19(1), pp.83-84.

understand the effect of some physical features of the price setting.[10] As we have said in the previous section, linear regression model for prices forecasting is still widely used. The equation that this type of model follows is the following:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

Where:

- Y: predicted value, in this particular case the price
- β_0: y- intercept
- β_n: regression coefficient of the n independent variable
- e: error term

In order to know it the variables were significant in the model we have used some coefficients that were given by R studio. For each variable the program establishes a value of the t student and the p-value, which will be used in order to know the significance of the variable solving the contrast hypotheses:

$H_0 \equiv \beta_0 = 0$ vs. $H_1 \equiv \beta_0 \neq 0$

$H_0 \equiv \beta_1 = 0$ vs. $H_1 \equiv \beta_1 \neq 0$

With this hypothesis we can determine if the effect of the independent variables is significant for explaining the behavior of the dependent variable which is the price.

Considering a significance level of 0.05 %, if p-value is lower than 0.05 the null hypothesis in rejected so the variable is significant. If it is larger than 0.05, the hypothesis is accepted, and it can be considered that the variable it is not significant to predict the dependent variable. This is done by the software RStudio and it assigned different level of significance based on it, it is represented like this:

[10] Ferreira, Â., Ramos, J. and Fernandes, P., 2019. A linear regression pattern for electricity price forecasting in the Iberian electricity market. *Revista Facultad de Ingeniería Universidad de Antioquia*, (93), pp.117-127.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.320e+01  6.815e-01 -19.362  < 2e-16 ***
ForWind     -1.547e-03  2.275e-05 -67.994  < 2e-16 ***
M            4.297e-01  1.522e-02  28.234  < 2e-16 ***
ForSolar    -1.148e-03  4.896e-05 -23.449  < 2e-16 ***
Hueco       -3.505e-04  1.678e-05 -20.889  < 2e-16 ***
GAS          8.434e-01  2.652e-02  31.800  < 2e-16 ***
Prevdemand   1.658e-03  2.527e-05  65.599  < 2e-16 ***
Imp         -5.658e-04  8.209e-05  -6.893 5.99e-12 ***
Exp         -8.824e-04  1.123e-04  -7.859 4.50e-15 ***
CO2         -2.241e-01  2.393e-02  -9.364  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.287 on 6443 degrees of freedom
Multiple R-squared:  0.8008,    Adjusted R-squared:  0.8005
F-statistic:  2877 on 9 and 6443 DF,  p-value: < 2.2e-16
```

Figure 5. Example of R code for multiple linear regression.

Three asterisks mean that the variable is very significant in the model, it is giving useful information, and none means that it does not give any information. This criterion has been used in order to determine whether or not used a variable in the final equation.

Finally, we can find at the end of the R code the Multiple R squared (R^2) and the R^2 adjusted which are indicators of the goodness of the fit of our model to the data. The R^2 ranges are between 0 and 1, so that R^2 values close to 1 indicate a good fit of the linear model to the data.

5. Correlation analysis of the main electricity price drivers

In the following chapter, the variables that cause more effect in the wholesale electricity price in Spain are going to be analyzed. Mainly based on correlations and linear regression equation we will evaluate the effect in the market clearing throughout three years (2018, 2019 and 2020). Also, in some cases, it is going to be analyze depending on the season of the years as there is strong seasonality in most of them.

5.1 Carbon prices

In order to mitigate the greenhouse gas emissions in Europe, in 2005 was establish an emission trading market, the EU-wide emission trading scheme (ETS)[11]. The European emission allowance (EUA) it is the first and biggest European carbon market and it is, and it will be the cornerstone in order to reduce carbon dioxide in the energy system. It works as a "cap and trade" system[12], it is established maximum amount of greenhouse emission that a plant can emit. This allowances it is reduced over time so that the emissions are reduced, the installations buy or received allowances that can be traded in the market.

Furthermore, it has been seen that some of the principal energy sources prices like Brent crude oil, gas or oil have suffered a significant impact because of the introduction of this market (EUA futures)[13]. As power plants introduce in their operational cost the fuel cost, the electricity price will be affected directly from the CO_2 prices. Therefore, it will be important to take into consideration the fluctuations of the carbon market in order to have an idea of the electricity prices dynamics.

It is expected than carbon prices are going to be one the most important price drivers due to the increasing trend of the prices in order to give the signal to the system to shift from carbon emitting technologies to renewables or low carbon emission sources.

We have analyzed the carbon prices of the year 2018, 2019 and 2020.

---

[11] Frondel, M., Schmidt, C. and Vance, C., 2012. Emissions trading: Impact on electricity prices and energy-intensive industries. *Intereconomics*, 47(2), pp.104-111.

[12] Climate Action - European Commission. 2021. *EU Emissions Trading System (EU ETS) - Climate Action - European Commission*. [online] Available at: https://ec.europa.eu/clima/policies/ets_en.

[13] Tan, X. and Wang, X., 2017. Dependence changes between the carbon price and its fundamentals: A quantile regression approach. *Applied Energy*, 190, pp.306-325.

Table 2. Descriptive statistical analysis of CO_2 price in 2018, 2019 and 2020

	2018	2019	2020
Correlation CO_2 – Spot price	0.529	-0.220	0.557
Mean	15.92	24.86	24.68
Max	25.19	29.79	32.86
Min	7.620	18.72	15.23

We can see that the carbon prices have an increasing trend over the last three years, reaching a maximum of 32.86 euros/ton in 2020. The correlation in 2018 and 2020 is almost the same, having a positive relation, which means that the higher the CO_2 prices the higher electricity prices as we have said before. However, it is interest to notice that with similar values as 2020, the correlation in 2019 was negative.
By doing the graphical representation the following figures have been obtained:

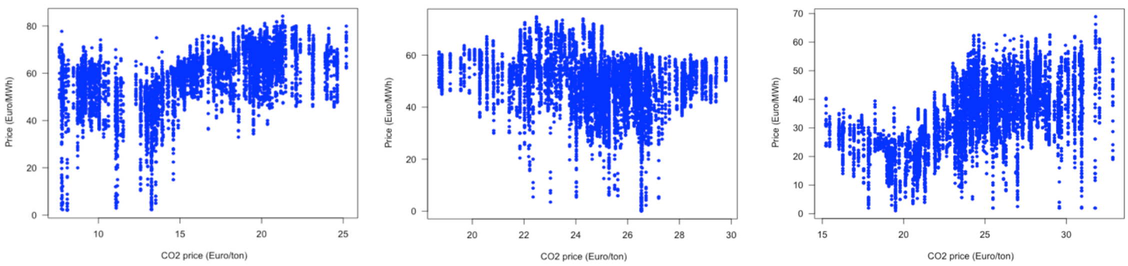


Figure 6. Spot electricity market price and CO2 price in 2018, 2019 and 2020 (from left to right).

The year 2018 and 2020 the effect of price increased that it was expected based on the theory, on the order hand, in 2019, another behave it is found. In order to understand this behavior, we have plot MIBGAS and CO_2 prices of 2019 as both prices are part of the bid of the combine cycle plant which set the price a significant number of hours in the Spanish system.

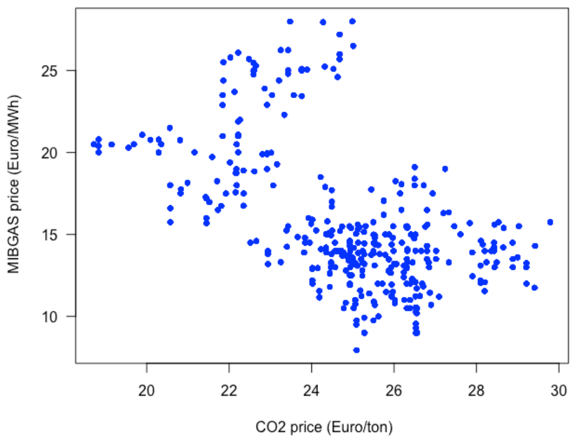


Figure 7. MIBGAS price and CO_2 price in year 2019.

In the graphic representation above we can see what the correlation value of the year 2019 described in *Table 2.* It is showing that the higher CO_2 prices are not related with higher MIBGAS price and therefore not correlated with higher electricity prices.
When going deeper in this result we have found that this weird behavior of CO_2 and electricity prices was because in the fourth quarter of the year 2019, there was an unusual increased of renewables in the system for that period. This ends up in a decreased of prices also helped for a decrease in the demand. The 64.5% of the generation of the system was obtained from zero CO_2 sources.[14]
Apart for the unexpected event the rest of the year carbon prices were having an effect of increasing prices. However, the event in November 2019 can be an example of what could happened in the future market structure with more renewables in the system. We may see that carbon prices would have less effect the larger the presence of zero carbon emission sources in the generation mix.
Nowadays, carbon prices are reaching historical maximum values, exceeding 50 Euro/ton.[15]

[14] Ree.es. 2021. *Demand for electricity in Spain falls by 0.4% in November | Red Eléctrica de España*. [online] Available at: https://www.ree.es/en/press-office/news/press-releases/2019/12/demand-electricity-spain-falls-04-november.

[15] Elperiodicodelaenergia.com. 2021. *El precio del CO2 se dispara ya por encima de los 50 euros por tonelada: Ribera, preocupada, anuncia medidas para paliar este coste de transición*. [online] Available at: <https://elperiodicodelaenergia.com/el-precio-del-co2-se-dispara-ya-por-encima-de-los-50-euros-por-tonelada-ribera-preocupada-anuncia-medidas-para-paliar-este-coste-de-transicion/> [Accessed 15 July 2021].

5.2 Gas price

Gas prices has a direct effect in the electricity system, combine cycle gas plants represent a 23.6% of the installed capacity in Spain. Moreover, it is known that they internalize the fuel costs in their bids therefore the price of the Spanish gas market (MIBGAS) will determine in a significant manner the wholesale electricity prices.
As the gas price can be known in advance, agents will know the D+1 gas price to have an idea of the market price for the following day.

Table 3. Descriptive statistical summary of MIBGAS prices in year 2020.

	MIBGAS (Euro/MWh)
Mean	17.4
Max	35.0
Min	5.1

In the *Figure 8*, it can be seen the linear regression representation of gas price and electricity price. We can observe that both variables followed a linear trend, the higher the MIBGAS price the higher the electricity price. The correlation between the two variables is 0.710.

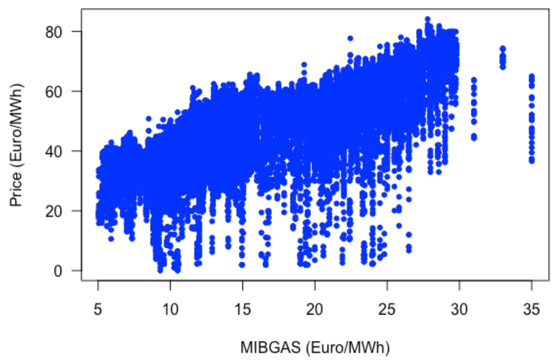


Figure 8. Spot price vs MIBGAS price.

By plotting both prices we can conclude that electricity prices are very correlated with MIBGAS as they follow the same trend over the year, even MIBGAS been much less volatile. These two variables probably have the strongest relation that we are going to see in this study.

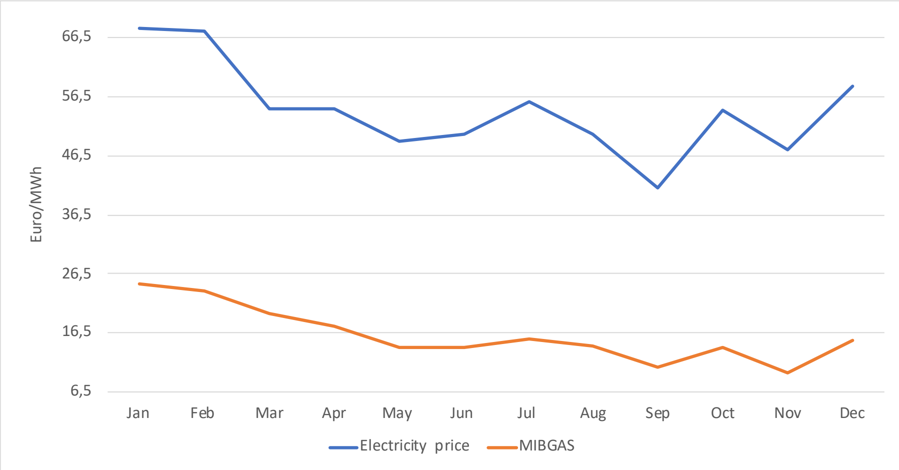


Figure 9. Comparison between electricity price and MIBGAS price in 2019.

The reason of this strong correlation is mainly because as we have seen in chapter 2, combine cycle plant, which use gas as fuel to generate electricity, have set the price 27.3 % of the hours in the year 2019. Moreover, other technologies like large hydroelectric reservoir, will bid at the same price as combine cycle plants based on MIBGAS futures in order to maximize its benefit and ensure to be dispatch.

5.3 Hydroelectric generation

Spain has an important hydroelectric production, for example the year 2020 the share of hydro in the Spanish system was 10.1 %.

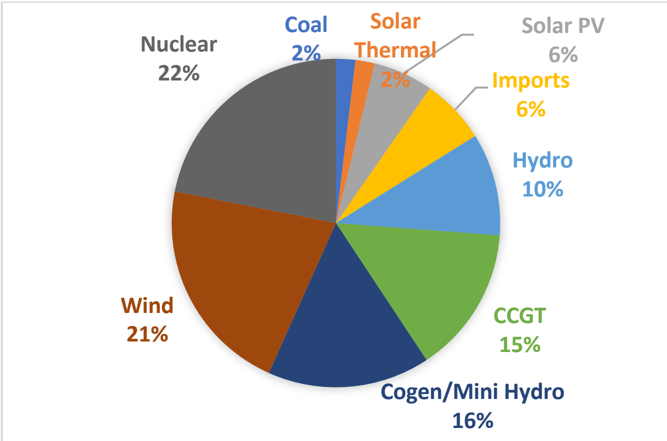


Figure 10. Share of technologies that have meet the demand in the Spanish system in 2020.

There are two types of hydroelectric generation technologies, the ones that cannot be manageable as they don't have the possibility of storing water, like run of the river plants and the ones with enough capacity to store the water. The last type has the interesting characteristics to have a water value, which is the opportunity cost that hydro agents sees depending on the time when they decide to produce in order to maximize its welfare.

Hydraulic power has a very low operating cost and great flexibility but is subject to water availability. They will take into account expected spot market price, mainly looking at MIBGAS and carbon prices due to the fact that combine cycle plants are setting the market price around 25% of the hours[16].

In addition, future market prices are usually taken into account in order to evaluate when is more profitable to produce, now or later in the future. Future markets for electricity are defined as markets in which the purchase or sale of energy can be negotiated in advance of the physical delivery. These terms are usually years, quarters, months, weeks, or days. In this case we are going to use yearly (Y+1), quarter (Q+1) and monthly (M+1) futures prices. The functioning is the following: if the agent buys one of these futures to sell energy, he will hedge the price. This means that he will pay the price of the spot market and afterwards, he will receive the difference between the spot and the future that he bought the year, month or quarter before.

Generation agents would do this when the spot price is lower than the future prices that he bought, as they want to maximize their profit. If spot price is larger than future price

[16] Omie.es. 2021. [online] Available at: https://www.omie.es/sites/default/files/2021-01/informe_anual_2020_es.pdf.

he may decide to produce now. On the order hand, if the future price is higher, he could think to wait in order to make higher profit in the future. The difference of both spot and future price is the opportunity cost, apart from other factors that may also affect in the decisions like the capacity of the reservoir or the season of the year. The Spanish future market has followed the following trend in the years that we are studying:

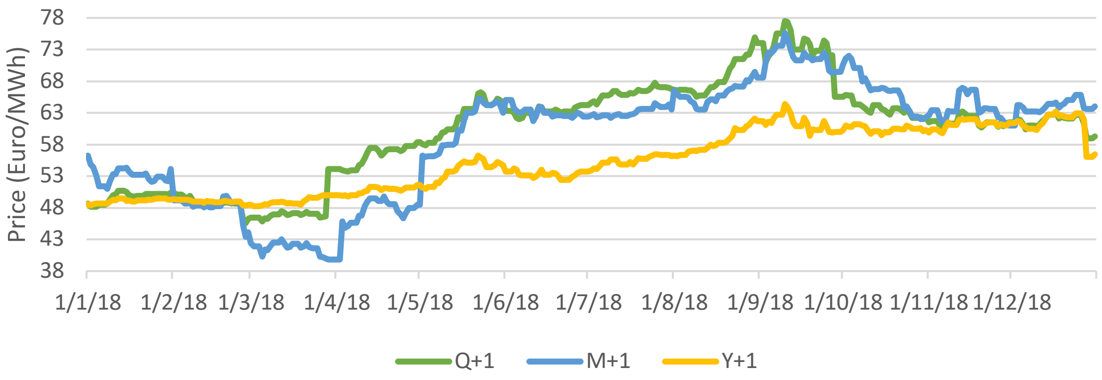


Figure 11. Future market prices in 2018.

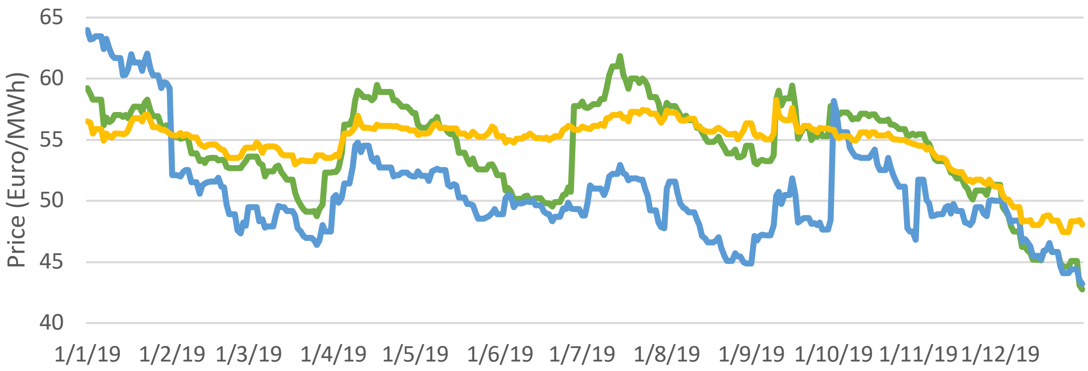


Figure 12. Future market prices in 2019.

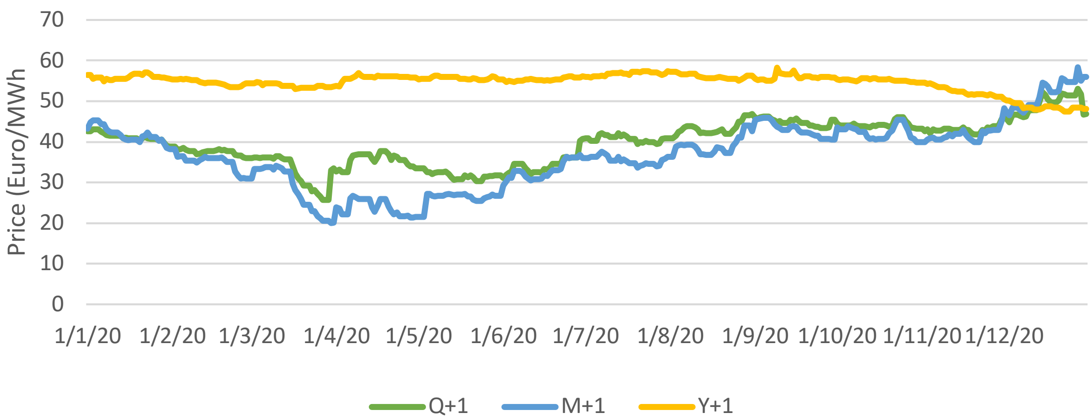


Figure 13. Future market prices in 2020.

A summary of future market prices values has been computed:

Table 4. Descriptive statistical summary of forward prices in Spain in 2018, 2019 and 2020.

	M+1(Euro/MWh)	Q+1 (Euro/MWh)	Y+1 (Euro/MWh)
Mean	49.4	50.2	53.4
Max	72.0	64.4	75.6
Min	20.0	25.7	33.6

Hydro generation along the year has the trend that it is shown in the figure below. As it can be observed the same trend it is followed throughout the three years that have been studied.

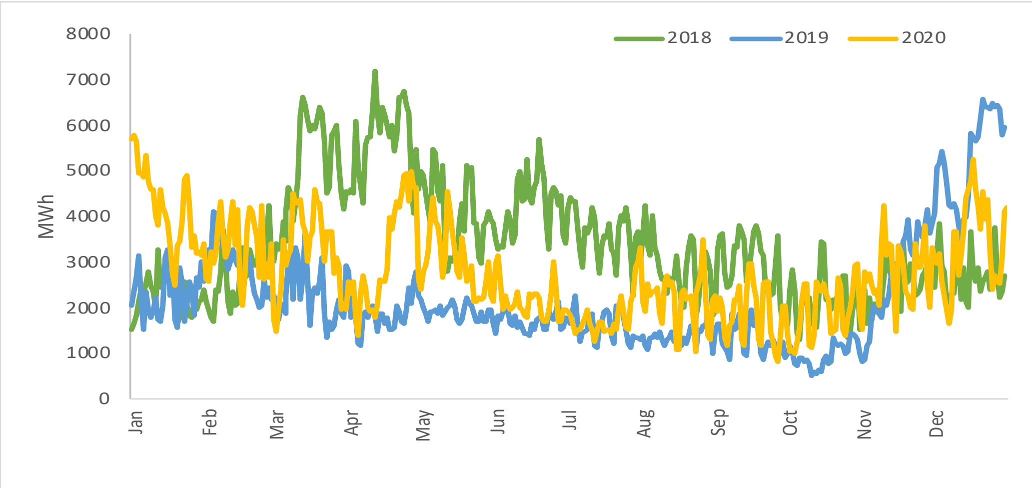


Figure 14. Hydro generation in Spain in 2018,2019 and 2019.

The capacity of the reservoir it is an important factor in the hydroelectric generation, the trend that has been followed in the last 5 years is the one represented in *Figure 15*.

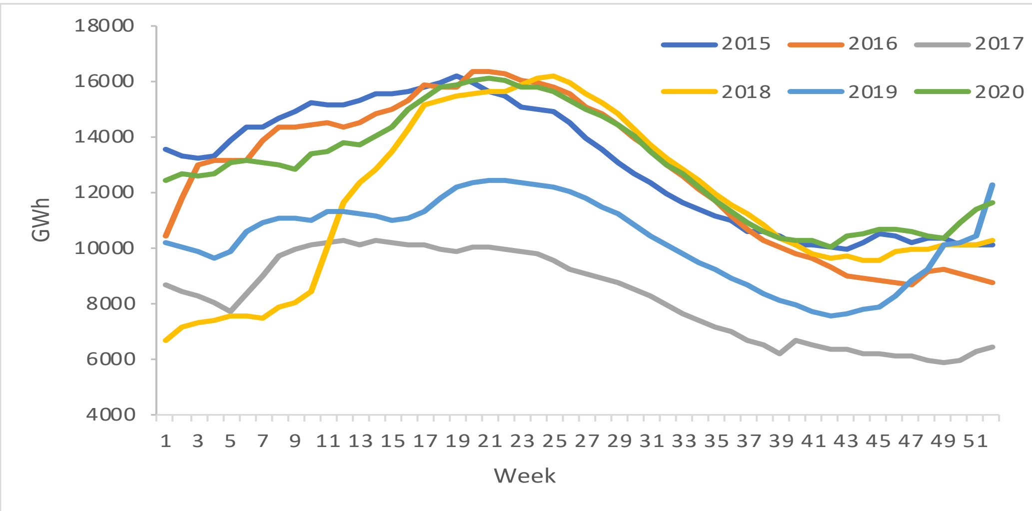


Figure 15. Capacity level of the reservoirs in the last 6 years in Spain. (Data source: OMIE).

To evaluate the relation between hydro production and the capacity, a capacity index has been computed. Data from 2002 to 2018 has been taken and the maximum level of the capacity of the reservoirs of each week during the 16 years has been selected as the maximum level of each week. The calculation of this new variable, the capacity index, aims to capture the effect that the capacity has in each period of the year but based on the historical maximum for each week. It is not the same 10000 GWh for a period where the average is 13000 MWh like in April, or for a period with 9000 GWh as average like in October (see Table 26 in the Annex).

Additionally, another index has been computed in order to capture the effect of the variation between future prices of OMIP and spot prices. The difference between them may make the agents to decide whether produce or save the water. Large reservoirs could potentially storage for a long time period up to several year.

The index that has been computed is:

$$P-(M+1)\% = \frac{M+1-Spot\ price}{Spot\ price} \cdot 100\%$$

$$P-(Q+1)\% = \frac{Q+1-Spot\ price}{Spot\ price} \cdot 100\%$$

$$P-(Y+1)\% = \frac{Y+1-Spot\ price}{Spot\ price} \cdot 100\%$$

In this study we have calculate the average price and the average hydro generation of each day as the futures and the capacity of the reservoirs are not in hourly basis.
As hydro sources are a technology with zero variable cost (cost of water is zero) like renewable, it would be expected to see a decrease in spot prices when the hydro the generation increase.

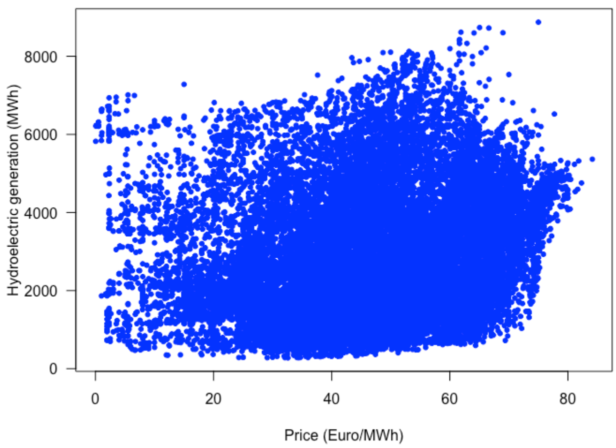


Figure 16. Hourly hydro generation (MWh) vs. price (Euro/MWh hourly).

Firstly, we are going to analyze the correlation of the three years that we are studying (2018, 2019 and 2020).
At first sight, what we can see in *Figure 17* is that the variables that are affecting more the hydro production are the yearly futures and the capacity of the reservoir. In the short term it is expected that hydroelectric generation has a very positive correlation with high prices because of the opportunity cost of the hydro plants with respect to the combine cycle plants, we will see this later. On the other than, in the long term, due to the opportunity cost to sell in the future, the correlation between future prices and hydroelectric generation will result negative.

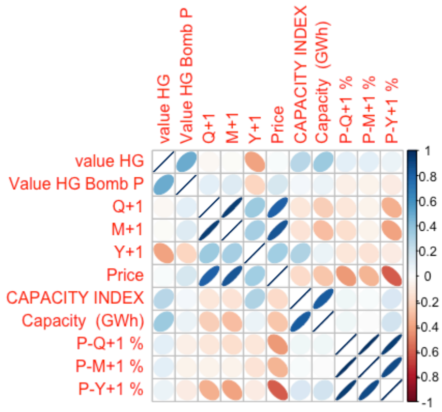


Figure 17. Correlation of hydro generation variable of 2018, 2019 and 2020.

As we can observe in the figures below, the futures Q+1 and M+1 and spot electricity price follow a similar trend, on the order hand, yearly futures behave in a much more disperse way with respect to the spot price over the range of study.

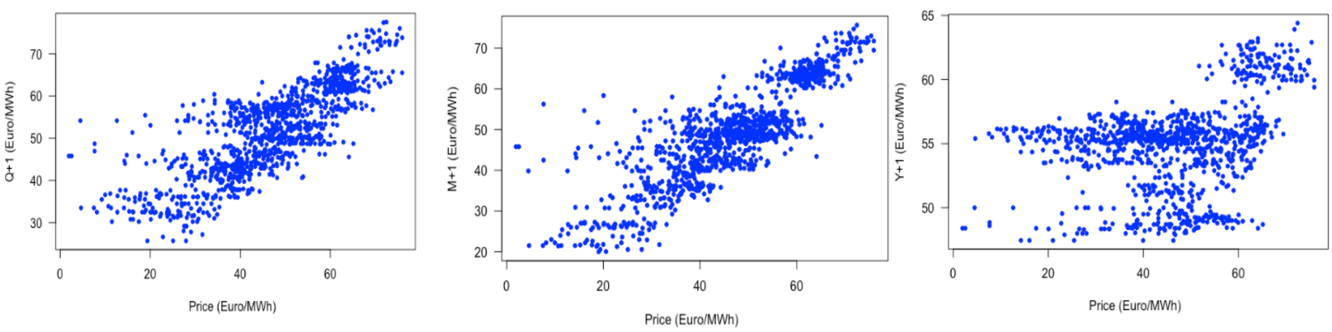


Figure 18. Market price (PBBF) and futures (Q+1, M+1 and Y+1).

Plotting the daily capacity of the reservoirs and the spot prices as it is shown in *Figure 18 it* can be observed that they follow a relation with negative correlation (-0.2733), the higher capacity the lower prices, but it is still not a very strong correlation so further studies will be need in order to understand this behavior.

In addition, as the capacity of the reservoirs levels is propionate in weekly basis by the MITECO, the correlation between the capacity index calculated before and the average weekly electricity prices has been calculated. The correlation that has been obtained is - 0.2042.

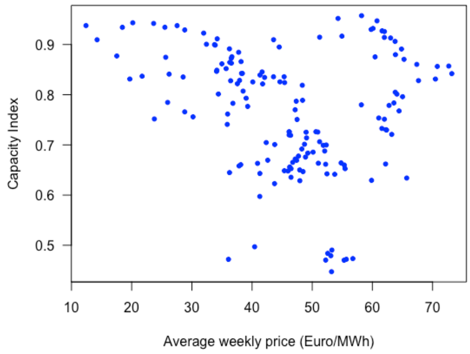


Figure 19. Capacity index and average weekly prices of 2018, 2019 and 2020 (Euro/MWh).

As in the study we are going to work with hourly data, the weekly capacity index has been duplicated in order to have it in hourly basis.

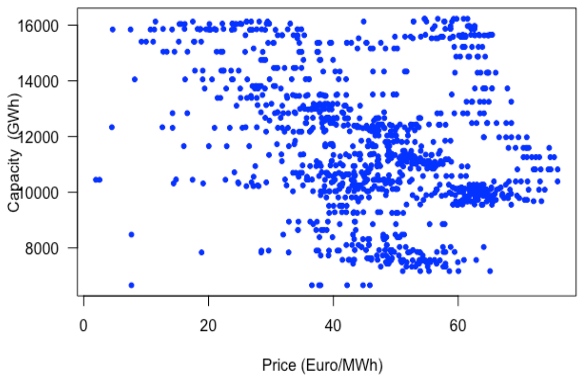


Figure 20. Capacity of the reservoirs (GWh) and prices (€/MWh) in 2018, 2019 and 2020.

Moreover, we have plotted the daily hydroelectric generation and the capacity of the reservoir and again there is not a clear correlation, in this case it is a positive relation between the variables (0.3606), the higher the capacity will end up in higher hydro generation.

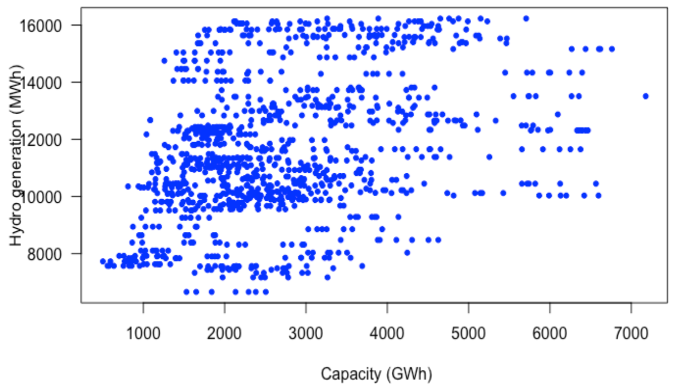


Figure 21. Hydro generation (MWh) vs capacity (GWh) of 2018, 2019, 2020.

Based on these results, where we cannot conclude in any clear relation between the variables, another analysis is going to carried out.
Knowing the strong seasonality of hydro generation, we are going to divide the year in quarters in order to capture the different behavior that agents can have during the different periods of the year.

Table 5. Division of the year for hydroelectric generation study.

1st quarter	2nd quarter	3rd quarter	4th quarter
January	April	July	October
February	May	August	November
March	June	September	December

- **First quarter**

Figure 11, Figure 12 and Figure 13 shown that the trend during this first quarter it is a decrease in the future prices M+1, Q+1 and Y+1 in the three years of the study. Moreover, it is a period were the capacity of the reservoir start increasing.

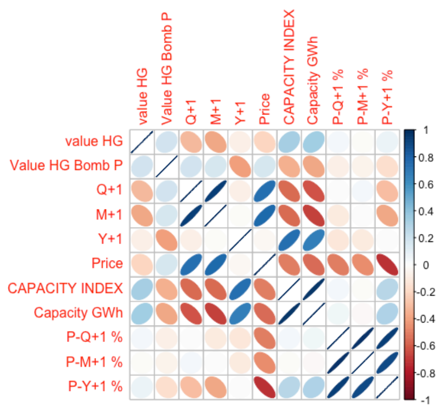


Figure 22. Correlation between variables in the first quarter of the year.

As we can see, hydro production is correlated with lower spot prices and with high-capacity indexes. The correlation analysis also shows that the difference between spot price and future market price does not have a strong effect in the hydro generation. Some graphic representation is going to be plot in order to understand better this data.

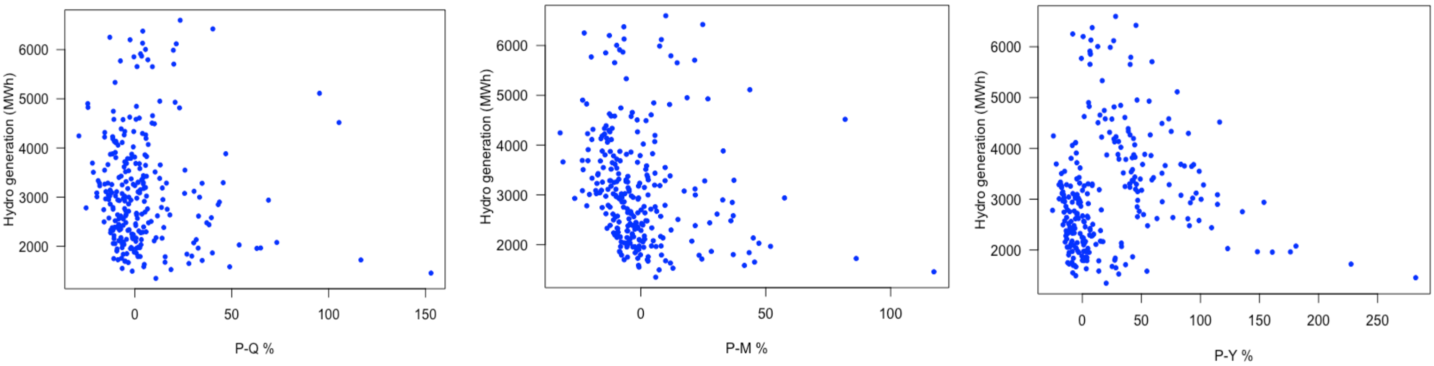


Figure 23. Hydroelectric generation (MWh) vs P-Q%, P-M% and P-Y%.

We can see more clearly in the graphs above that the variation of future prices with respect to spot price is not following a linear behavior with respect to the hydro dispatch.

Furthermore, the correlation between hydro generation and prices during the first quarter of the year is - 0.2072.

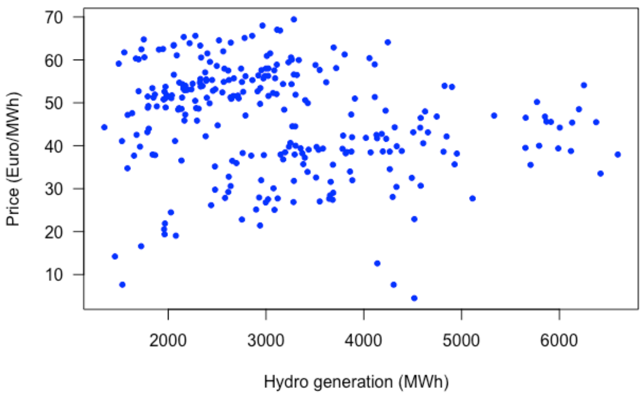


Figure 24. Daily average price (Euro/MWh) and average hydro generation (MWh) during Q1.

This correlation means that hydro can be somehow correlated with a decrease in spot prices. As in this period the capacity is considerable high the production can achieve a considerable decrease of prices, but it is not strong enough correlation to assure it.

In *Figure 25,* we can observe how hydro generation only depend on the available capacity of the reservoirs with low-capacity indexes (lower than 0.7), above that, the generation can go from 2000 to 6000 MWh.

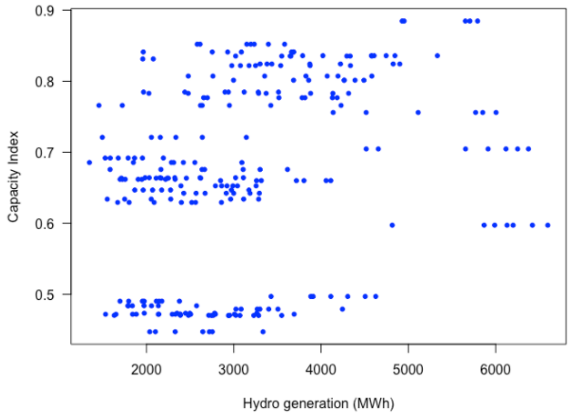


Figure 25. Hydro generation and capacity index.

The relationship between spot prices and capacity it is very similar as *Figure 26* illustrates, when the capacity is low, the prices are between 40 to 70, and above 0.7 which corresponds with around 10000 MWh, the variables are follow a more linear trend, the higher the capacity the lower the prices but also reaching high prices like in the case of lower capacity.

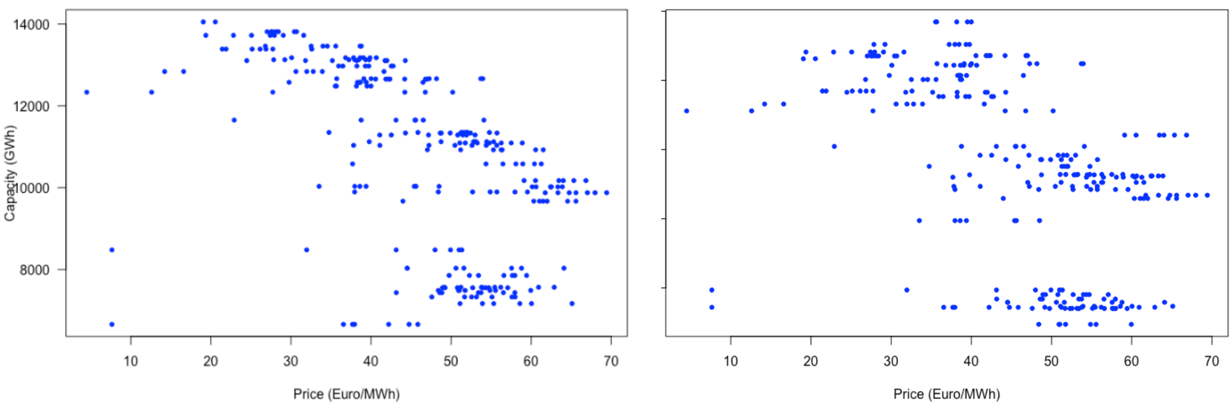


Figure 26. Daily average price vs capacity (GWh) and capacity index during Q1.

- **Second quarter**

April, May and June are months characterize with higher capacity index. Moreover, the fact that the upcoming months correspond with summer can mean that agents will try to save water in order to have enough energy to produce during these months were future prices are high as we could see in *Figure 12, 13 and 14.*

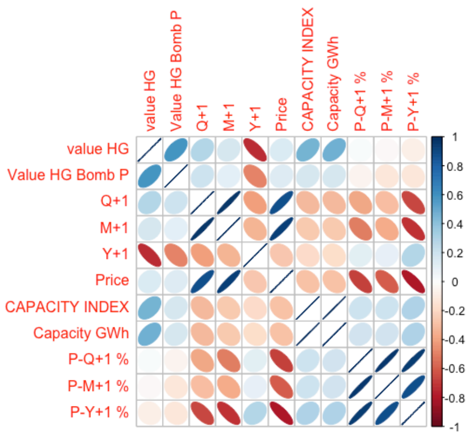


Figure 27. Correlation during Q2.

In this case we can observe that yearly futures and the values of capacity of the reservoirs are the variables with the strongest correlation with hydro generation.
Spot prices have not a clear trend with respect to hydroelectric generation. Seems to be related with an increase of prices. In *Figure 28* we can see what we have obtained in the correlation analysis.

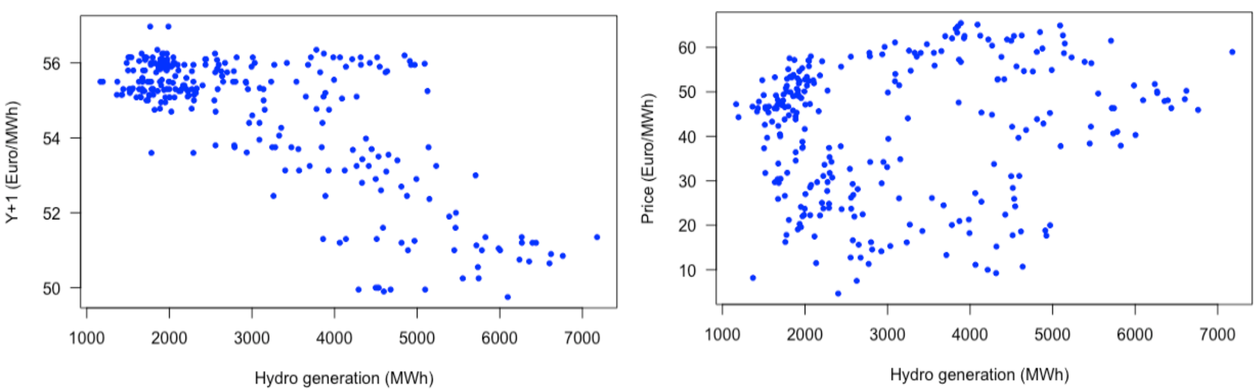


Figure 28. Hydro generation (MWh) vs Price and hydro generation (MWh) vs Y+1.

The correlation between future Y+1 and hydro generation is -0.7362. The linear regression equation is the following:

$$Hydro\ generation\ (MWh) = -593.07 \cdot Y + 1\left({}^{Euro}/_{MWh} \right) + 35464$$

With spot prices the relation is much more disperse than in the case of Y+1 future price. We can also see in the correlation study that the variation between yearly futures and prices can have some impact on the generation of hydro in the system.

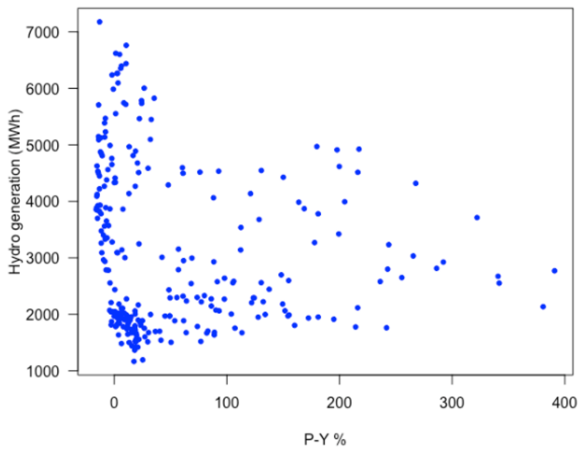


Figure 29. Hydroelectric generation (MWh) vs P-Y%.

It can be observed that the trend is that higher the variation (higher future price) the lower generation.

The capacity effect is less significant than in the first quarter, we do not see any linear relation between prices and hydro generation. Lower capacity indexes result in lower generation, but high levels of capacity index have as a result very disperse values of generation not necessarily high.

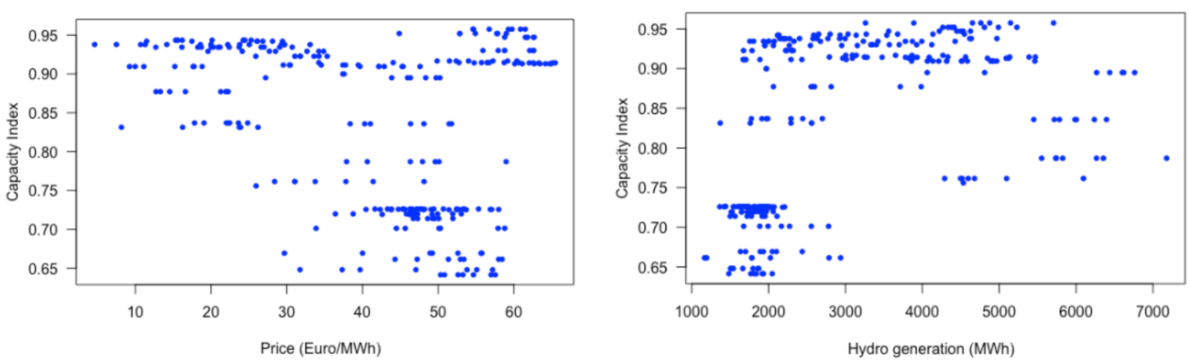


Figure 30. Capacity Index and hydro generation (MWh) and capacity Index vs price during Q2.

- **Third quarter**

In this quarter which correspond to summer months the capacity it of expected to decrease along the months. The correlation between the variables have been computed as in the previous cases.

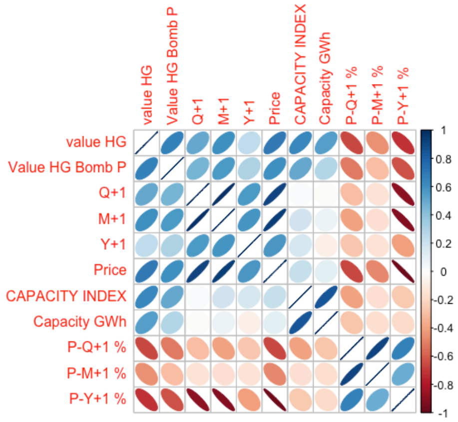


Figure 31. Correlation between variables in the third quarter.

In this case the correlation with spot prices is quite strong as we can see in the *Figure 32,* the higher the price the higher hydro generation in the system.

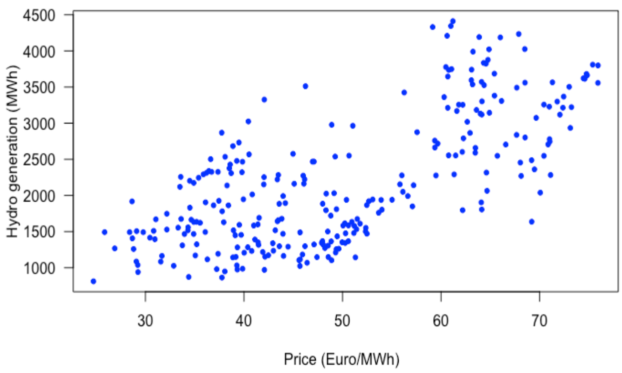


Figure 32. Price (Euro/MWh) vs hydro generation (MWh).

Correlation between hydroelectric generation and prices is 0.699. In this case as there is low capacity in the system, hydroelectric plant agents are only going to produce in hours where they can make the maximum profit as they don't have enough water, they have to see where they can make more money to recover the cost of the hours when they not produce.

It can also be observed that future prices are very correlated with spot prices so in order to analyze if there is an effect of the spread of these two prices in the hydro generation the index that has been explain previously it is very useful in order to understand this effect. A it is shown in *Figure 33*, the higher the percentage means higher future prices, and this results in lower hydro production.

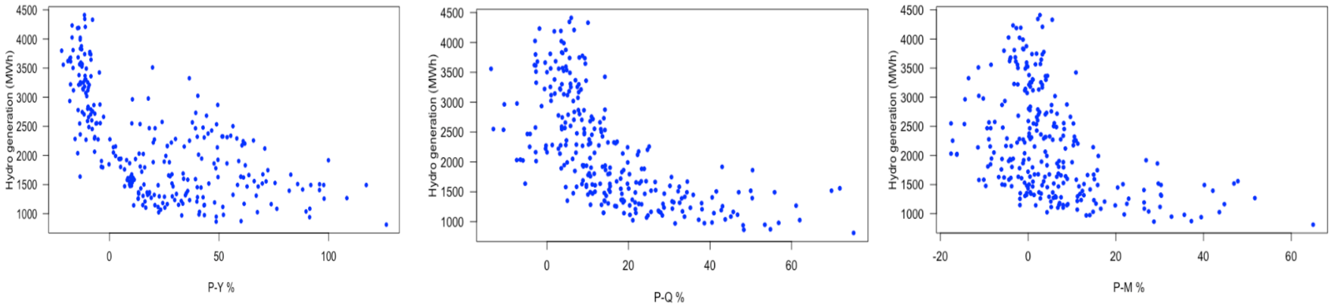


Figure 33. Hydro generation (MWh) vs P-Futures %.

In terms of the level of the reservoirs, the same trend it is followed as in the rest of the months, values lower than 0.7 do not have an effect on prices nor in the dispatch. Moreover, the period of the right-hand side of the graph correspond to September 2018 where there was historical high CO_2 and MIBGAS prices, that's why the high prices that do not correspond to the high-capacity index.

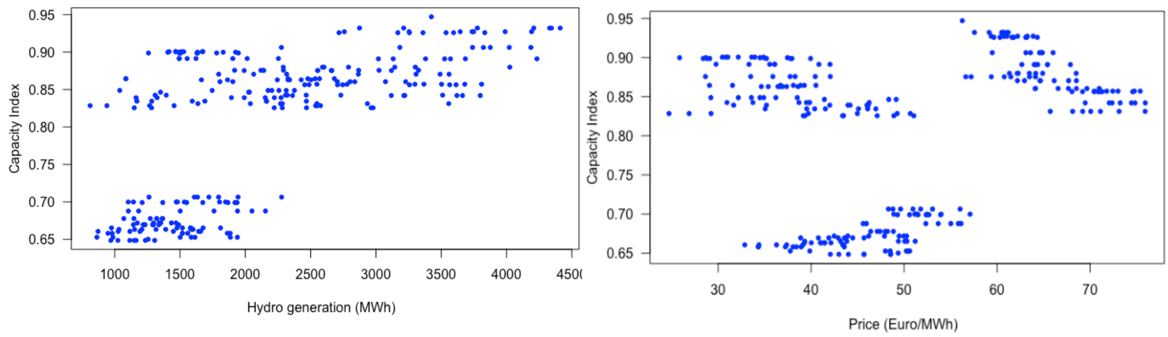


Figure 34. Hydro generation (MWh) vs capacity index during Q3.

- **Fourth quarter**

To end up with this analysis the correlation of the variables of the last period of the year has been plotted.

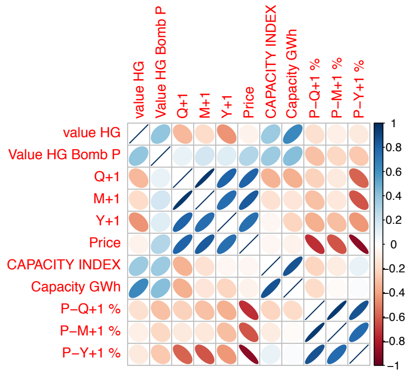


Figure 35. Correlation between variables in the fourth quarter.

In this period the capacity start increasing after summer so lower prices seems to have a relation with higher MWh of hydro energy in the system. As it can be seen in *Figure 36* up to 3000 MWh of hydroelectric generation, the increase of the share of hydro does not have an effect of price decreased. Above 3000 MWh the prices start decreasing as the generation increased.

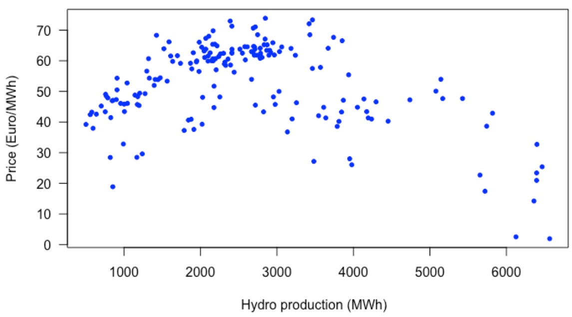


Figure 36. Correlation price (Euro/MWh) and hydro generation (MWh) during Q4.

The percentage of variation between future prices and spot prices has an effect in the decision of dispatching or not. When the percentage it is more positive (higher value of future prices with respect to spot prices) the result is lower hydro generation.

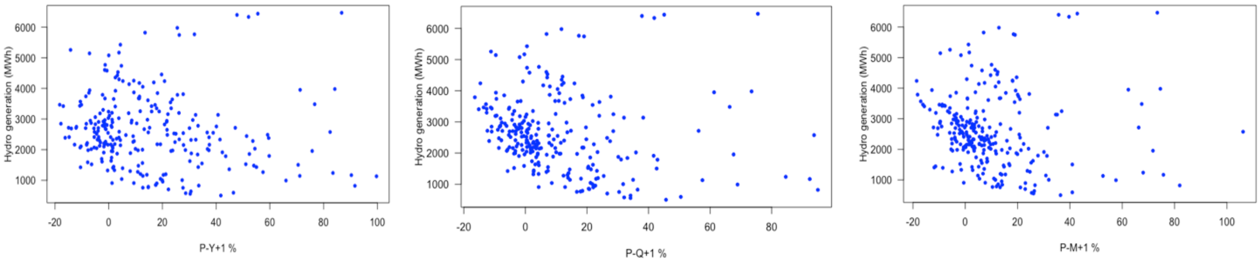


Figure 37. Hydro generation (MWh) vs P-Y %, P-Q% and P-M % during Q4.

In this case the capacity behaves in a more linear way with respect to the hydro production in the first part of the linear regression graph, up to 0.75 the linearity it is lost.

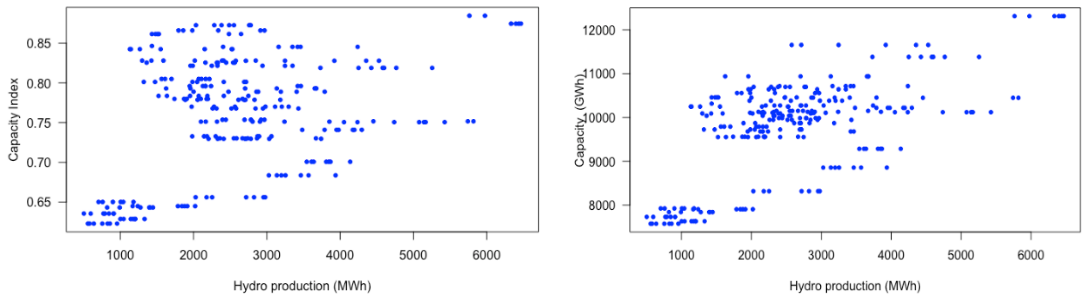


Figure 38. Hydro reservoir capacity (GWh) and hydro index vs hydroelectric production (MWh) during Q4.

To summarize, we have figured out the following facts:

1. During the periods with intermediate capacity of the reservoirs, first and fourth quarter, hydro it is correlated with a decrease of spot prices. During these periods the capacity it is also has more impact in the hydro dispatch, as we have seen they follow a more lineal trend than in the other quarters.
2. During the second quarter, high prices are correlated with higher hydro generation but not as clear as in the third quarter, probability due to the higher presence of wind in the system that will set the price more hours that in summer. On the order hand, generation is not following an increasing trend with respect to the capacity index. Futures Y+1 is very correlated with the dispatch, the higher the future with respect to the spot price the lower the generation.
3. In the third quarter the correlation between spot price and hydro is even more positive (0.699), it is the period with the highest relation. Moreover, the effect of the variations of the spot prices and the futures is the most relevant over the year and the capacity has not an effect in the decision of producing or not.

To sum up, we have found that the capacity index has an effect on the dispatch but depending on the level of the reservoir and on the period of the year. We cannot conclude that high capacity means higher production in all cases. Moreover, the effect of future market depends on the period of the year, in the third quarter is where this is more relevant.

We will put in practice this conclusion in the last part of the thesis where we computed multiple linear regression models (MLRM) for each period taking into account the capacity indexes and future prices.

5.4 Pumped hydro generation

Nowadays, pumped storage hydro has 2,7 GW of storage capacity and it is expected to increase up to 4,2 GW in 2025 and to 6,8 GW in 2030 to achieve the forecast scenario of having 6837 MWh in 2030.[17]

Pumped energy storage is one of the most mature storage technologies and, thanks to its efficiency and flexibility, is widely deployed throughout Europe. It currently represents more than 90% of the installed storage power at European level. Because of this, understand how this technology is dispatch in the market will be very useful and also it could give an idea of how batteries will enter into de market.

Pumping plants store energy in the from of the potential energy of water, lifting water from a lower reservoir to a higher one. During periods of high demand, the water stored in the upper reservoir is released by turbines to a lower reservoir to produce electricity, while in periods of low demand the water is pumped back to the upper reservoir and stored again. Pumped energy storage is essential for the reliable and safe transition to an electrical system based on renewable energies.

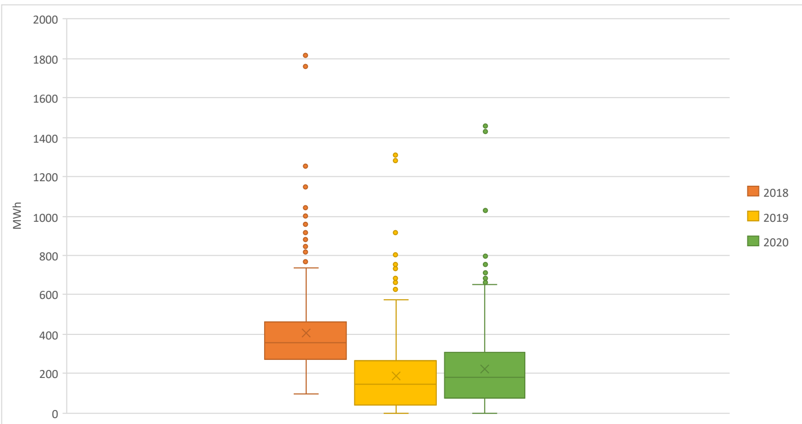


Figure 39. Average of pumped hydro generation (MWh) per day in 2018,2019 and 2020.

In Spain the average pumped hydro generation it is around 300 MWh.

As it has said before, the water is pumped in periods with low demand and release in periods with higher demand, if we translate this to the daily basis means that agents will be pumped during the valley hours and they will produce during peak hours. In the following figure the difference in prices during peak and valley consumption hours.

---

[17] Miteco.gob.es. 2021. *Plan Nacional Integrado de Energía y Clima (PNIEC) 2021-2030*. [online] Available at: https://www.miteco.gob.es/es/prensa/pniec.aspx.

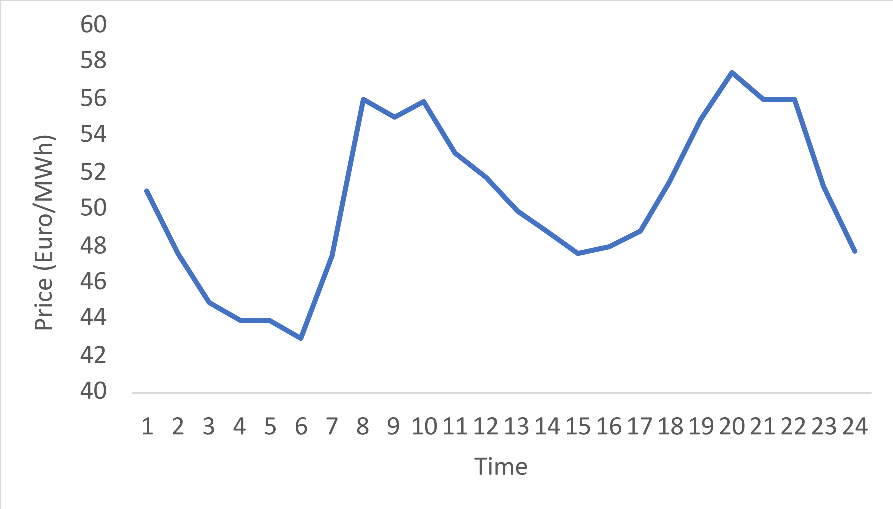


Figure 40. Electricity price profile during a day.

As we can see in *Figure* 40 the hours the hours from 4 to 7 are the ones with lowest prices and the ones from 19 to 22 are the ones with the highest prices. The difference between them is the opportunity cost that the pumped hydro plants have in order to generate energy.

The difference of prices between peak and valley hours has been computed in order to evaluate the effect in the hydroelectric production coming from pumped power plants.

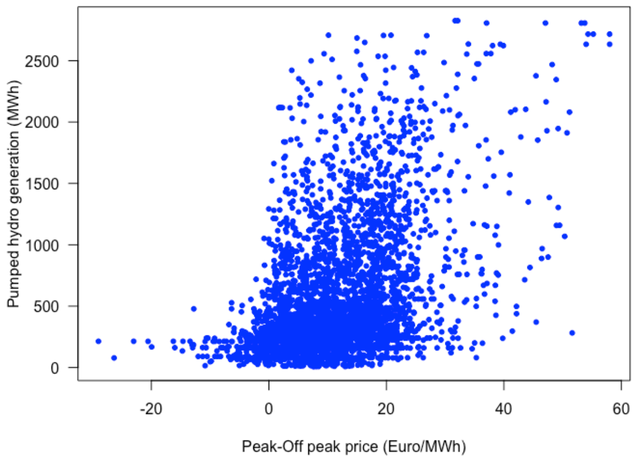


Figure 41. Pumped hydro generation (MWh) vs difference in prices (peak-off peak) in Euro/MWh.

In this case, the correlation between the spread of prices and the pumped hydro generation is 0.440. The linear regression equation that explains this behavior is the following:

$$Pumped\ hydro\ generation\ = 25.0689 \cdot (Peak - Offpeak\ price) + 282.654$$

However, it is interesting to see that with a spread of 30 €/MWh they are always going to schedule less than 500 MWh. The correlation in this case must be lower, precisely, 0.216, as they are not changing the pumped hydro generation with the change of the spread in all the range, from almost 0 to 500 MWh. Only choosing the values lower than 500 MWh the graph that it is obtained is the following:

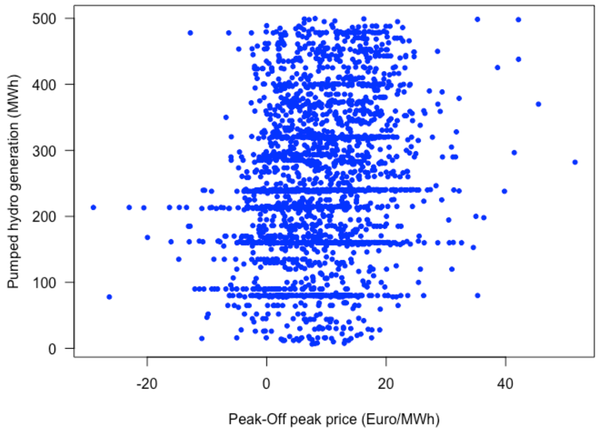


Figure 42. Pumped hydro generation (MWh) vs difference in prices (peak-off peak) in Euro/MWh below 500 MWh.

Figure 42 illustrated very clearly the decrease in the linearity between both variables. In the case of the relation with futures, comparing with the case of convectional hydro have found that the effect of the variation of spot and future prices is not significant, the values of correlation that have been obtained were less than - 0.1.

5.5 Wind generation

Wind power has increased its share in the electricity market in the last decades. The install capacity of wind (onshore and offshore) in the objective scenario for 2030 establish in the PNIEC is 48550 MW which suppose a significant increase from the value of 27446 MWh in 2020. This increase is due to the need to achieve the objective of having a net zero emission energy system by 2050. Due to the negligible variable generation cost of this technology, the increment of the penetration will lead to change in the electricity market. It has been seen in some countries negative prices in periods with large share of wind generation.[18] Spain has switch into this new format of electricity prices having a range from - 500 €/MWh y + 3000 €/MWh as it is establish in «BOE» num. 339, 29 of December 2020, pages 122417 to 122419.
In this context an analysis of the effect of wind generation in electricity prices is going to be developed in order to understand and have an idea of the future impact on prices.

Wind power forecasted time series is provided by Red Eléctrica, the Spanish TSO, who use a wind predicting model called SIPREOLICO which predicts the total hourly wind production for the following 24 h.[19]
In this study, we are going to study the effect of wind in the different month of the years 2018, 2019 and 2020 using hourly data.

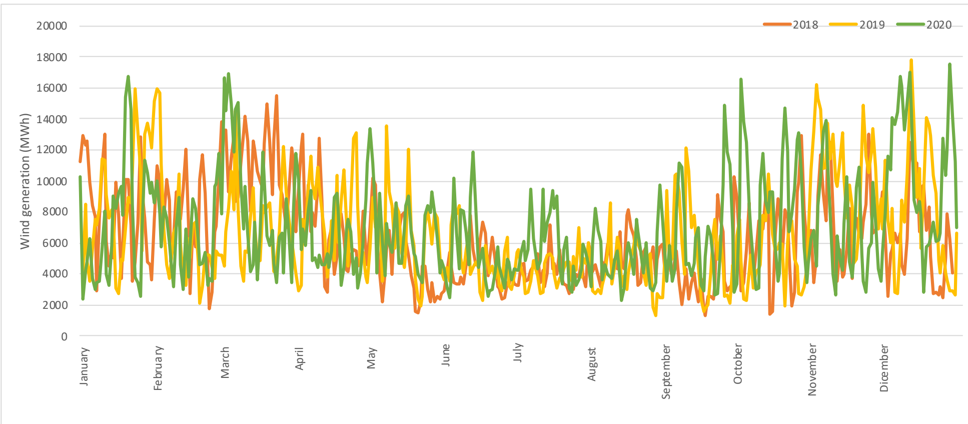


Figure 43. Wind generation (MWh) per day during 2018, 2019 and 2020

---

[18] Morris, C., 2021. *German power prices negative over weekend | Energy Transition*. [online] Energy Transition. Available at: https://energytransition.org/2014/05/german-power-prices-negative-over-weekend/.

[19] Lledó, L., Torralba, V., Soret, A., Ramon, J. and Doblas-Reyes, F., 2019. Seasonal forecasts of wind power generation. *Renewable Energy*, 143, pp.91-100.

As we can see the generation variate significantly along the year.

Table 6. Descriptive statistical summary.

	Wind forecast (MWh)
Mean	5913
Max	17996
Min	246

Wind power normally has a low marginal cost (zero fuel costs) and therefore enters near the bottom of the supply curve. This shifts the supply curve to the right, resulting in a lower power price, depending on the price elasticity of the power demand. In general, the price of power is expected to be lower during periods with high wind than in periods with low wind. In the following figure we can see this effect, the higher wind results in lower price with a correlation of - 0.340.

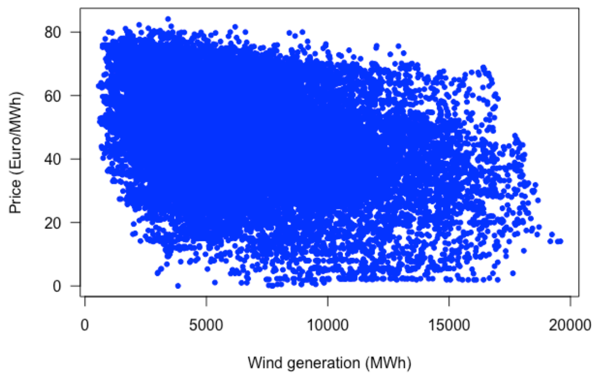


Figure 44. Wind generation (MWh) vs spot price (MWh).

$$Price = -0.0015 \cdot Wind\ generation(MWh) + 56.115$$

In order to have a better figure of this effect a correlation analysis has been developed for each month.

Table 7. Matrix correlation of spot price and wind generation in the years 2018, 2019 and 2020.

Jan	Feb	Mar	Apr	Jun	Jul	Aug	Sep	Oct	Nov	Dec
- 0.207	-0.267	-0.318	-0.501	-0.451	-0.438	- 0.127	-0.562	-0.472	-0.463	-0.379

By analyzing the results of the correlation, we have seen that the only month with a different behave its August, which is characterize for low wind generation and therefore we can see the lowest correlation between wind generation and spot prices.
We have represented both variables and we have seen that there is an effect on price reduction but there are a lot of period were the wind generation is very low as we can see in the following figure.

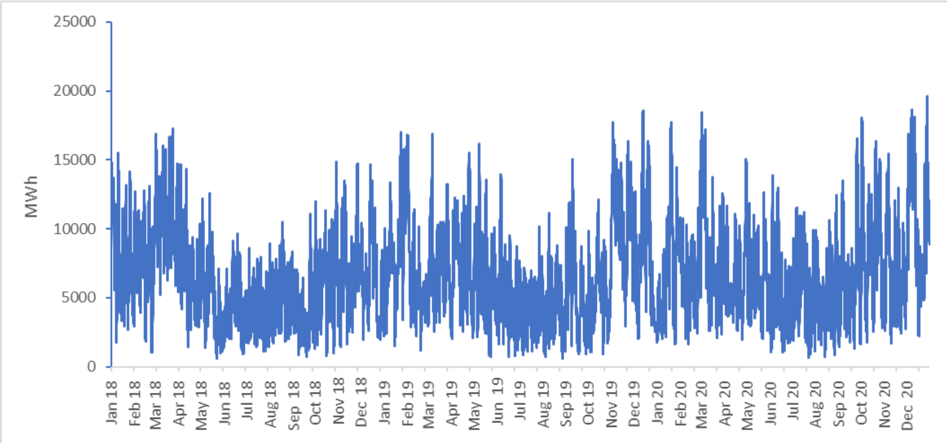


Figure 45. Wind generation and electricity price in August 2018,2019 and 2020.

After studying the data of this month, it has been seen that with lower values than 1700 MWh of wind generation there is not a clear effect of reduction the spot price.
We have analyzed the rest of the months and we came out with the fact that when there is lower generation than 1700 MWh there is not an effect of reduction of the energy price. This only represents a 2.52 % of the hours of the three years that we are studying.

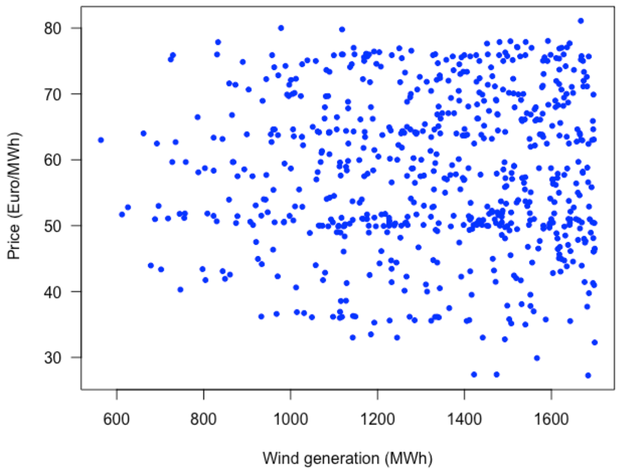


Figure 46. Price and wind generation lower than or equal to 1700 MWh.

$$Price = 0.0008 \cdot Wind\ generation(MWh) + 56.952$$

With a person correlation coefficient of 0.019.

We can conclude that wind has an effect in price reduction during all the year, however, in the case that we are studying, August had 7.08 % of the hours with wind generation lower than 1700 MWh, higher than the annual value of percentage of hours with lower wind generation in the system and this result in a lower effect on prices reduction.

5.6 Solar photovoltaic and solar thermal generation

In this section both solar technologies, photovoltaic and solar are going to be study. Its relationship with prices and the recent changes due to the increased in it share in the Spanish electricity system.

Table 8. Descriptive statistical summary.

	Solar (PV+ thermal) forecast (MWh)
Mean	1751
Max	9173
Min	0

- **Solar photovoltaic**

Spain is one of the European countries with more favorable conditions for photovoltaic energy generation as it is the country with highest number of hours of solar radiation. The increasing share of this technology in the Spanish system it is very significant in the last year as we can see in *Figure 47.*

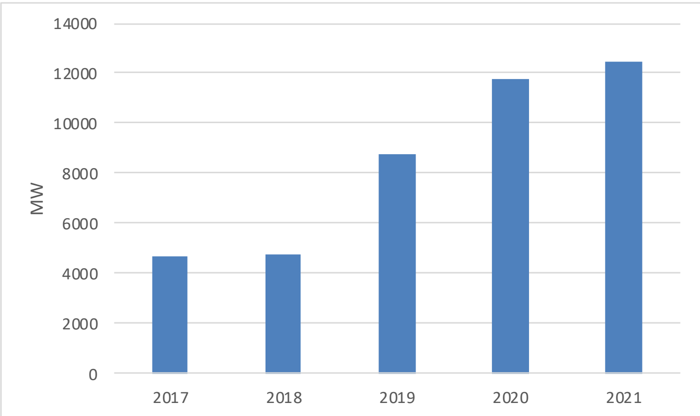


Figure 47. Evolution of solar PV installed capacity (2017-2021). (Source: REE)

Moreover, the objectives for this technology are very ambitious and as is establish in the PNIEC, solar PV installed capacity will reach 38404 MW by 2030 which suppose a huge increase from the 12444 MW that are actually installed.

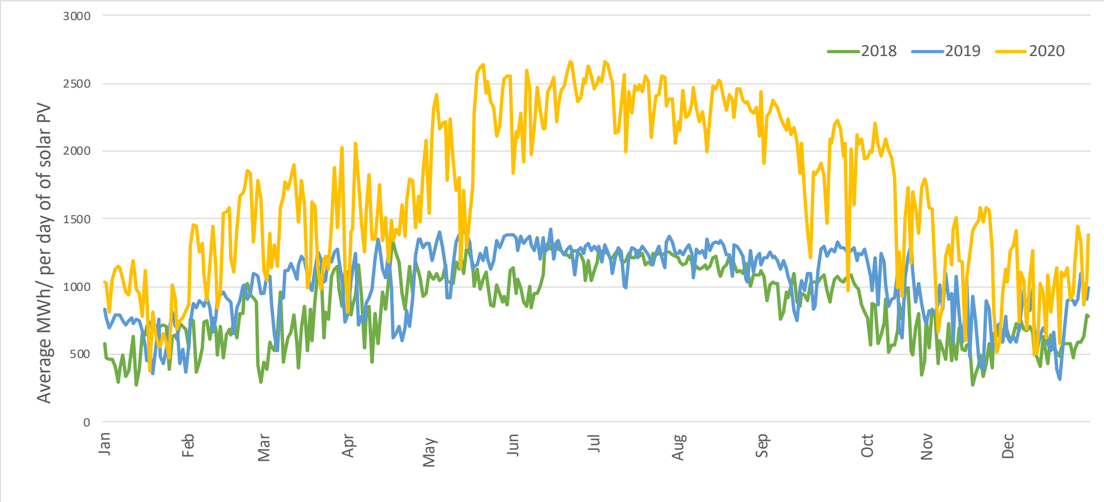


Figure 48. Average of MWh of solar PV generation in Spain in 2018,2019 and 2020. (Own elaboration, data: OMIE).

As we can see the increase of the install capacity has an important impact. Also, it can be seen the strong seasonality of this technology.
In order to study its effect in prices we are going to do it by months and also only taking into account the sun hours of each month.
In the following table the number of sun hours of each month is represented.

Table 9. Sunrise, sunset, and hours of sun by months in Spain

	Sunrise	**Sunset**	**Duration of the day (h)**
January	08:34	18:14	9.66
February	08:06	18:51	10.75
March	07:24	19:22	11.96
April	07:34	20:55	13.35
May	06:56	21:25	14.48
June	06:42	21:48	15.06
July	06:56	21:45	14.82
August	07:24	21:13	13.82
September	07:54	08:25	12.52
October	08:24	07:36	11.2
November	07:59	17:59	9.98
December	08:29	17:50	9.35

We can see than the sun hours for each month change a lot so taking this into account when evaluating the solar generation impact it is important.
We have evaluated the relation between hourly spot prices and solar PV generation.

If we plot both variables, we obtain the figure below where we can see that there is a change in the trend up to 3000 MWh of solar dispatch.

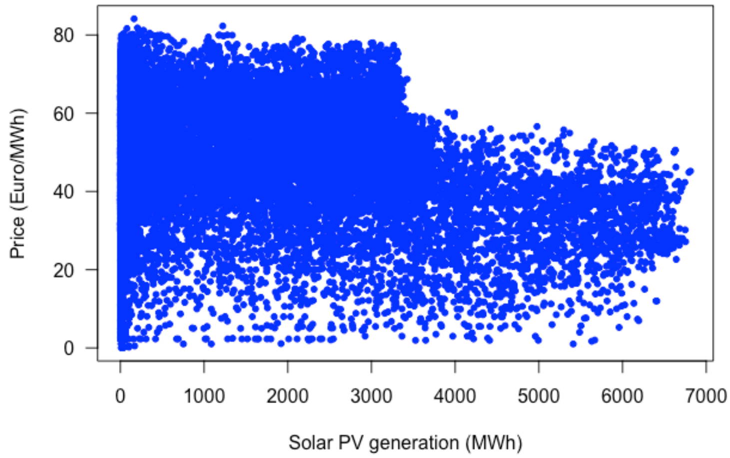


Figure 49. Spot price vs solar PV generation (MW) of years 2018, 2019 and 2020.

In order to know to which year, correspond this change we have differentiate with colors the different periods. In orange the years 2018 and 2019 and in blue the year 2020.

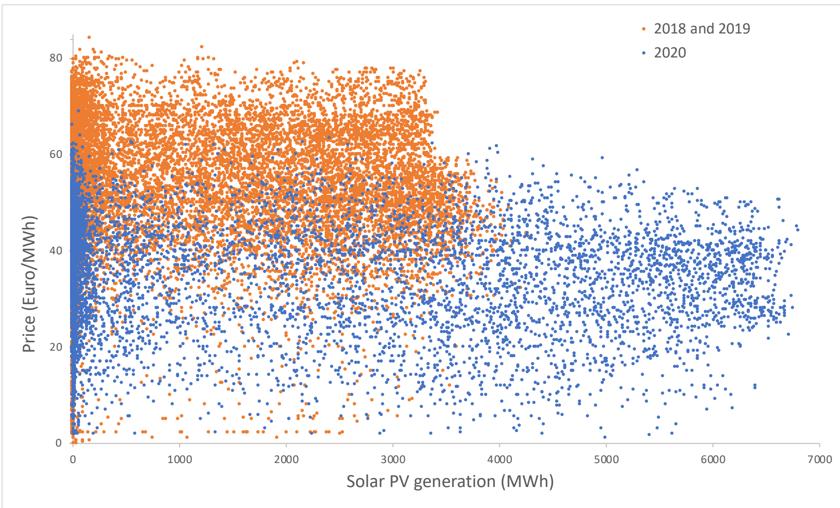


Figure 50. Spot prices (Euro/MWh) vs solar PV generation (MWh) of 2018, 2019 and 2020 with colors differentiation.

In the figure above can see that there is a clear change in the trend. This shift is due to the increased in the share of solar PV in the year 2020 as we have seen previously in Figure *39*. If we plot the years 2018, 2019 and 2020 separately we obtain the following:

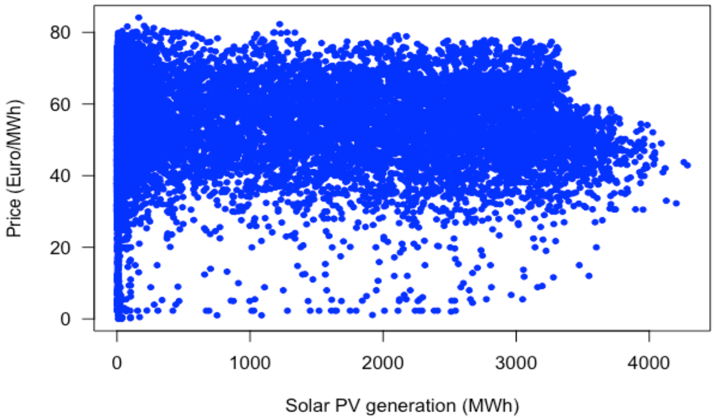


Figure 47. Spot price vs solar PV generation (MW) of years 2018 and 2019.

With a linear regression equation of:

$$Price\left({}^{Euro}/_{MWh}\right) = 51.19 + 6.36 \cdot 10^{-4} \cdot Sola\ PV\ generation\ (MWh)$$

For the year 2020 we have obtained the following graph:

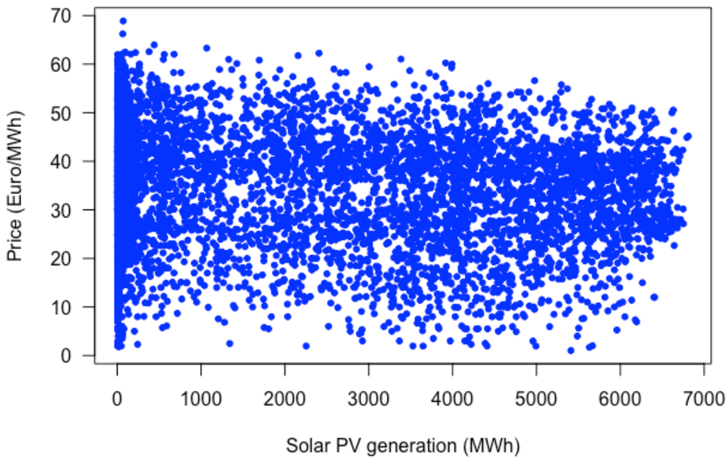


Figure 48. Spot price vs solar PV generation (MW) of year 2020.

In this case where the solar reach higher values of solar PV generation we can see that there is more relation on price reduction than in the case of the years 2018 and 2019. We can observe with the linear regression equation how the solar PV generation has a slightly effect on price reduction along the year 2020.

$$Price\left({}^{Euro}/_{MWh}\right) = 34.48 - 3.06 \cdot 10^{-4} \cdot Sola\ PV\ generation\ (MWh)$$

The correlation of prices and solar PV generation in 2018 and 2019 was 0.058 and in 2020 was -0.057 so again we can see this change due to the increase of the installed capacity of solar in the last year.
Taking into account the three years, all the months have negative correlation with price, but if we only chose the year 2018 and 2019 the results change as we have discussed previously.

With the result that have been found, it has decided to instead of using a time discrimination in order to capture the effect of solar photovoltaic, we are going to evaluate it by differencing between hours with higher solar share than 3000 MWh and hours with lower share than 3000 MWh, as we have found that the effect of price reduction it is only significant up to 3000 MWh. By doing that we aim to have a more accurate approximation to the real prices.

In the following study we have evaluate the correlation between price and solar PV generation each month of the year during sun hours.

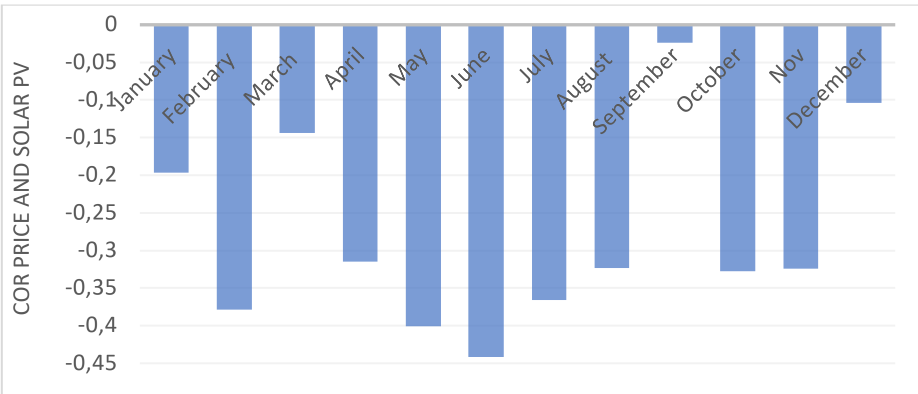


Figure 51. Correlation between prices and solar PV generation during sun hours in 2018, 2019 and 2020.

In order to capture the effect of solar PV during periods were the generation is not high enough to be able to have an effect in the sport price, will be interesting to study the effect as the amount of energy in MWh of the total demand that we have prevented from the use of technologies with higher variable cost such combine cycle.

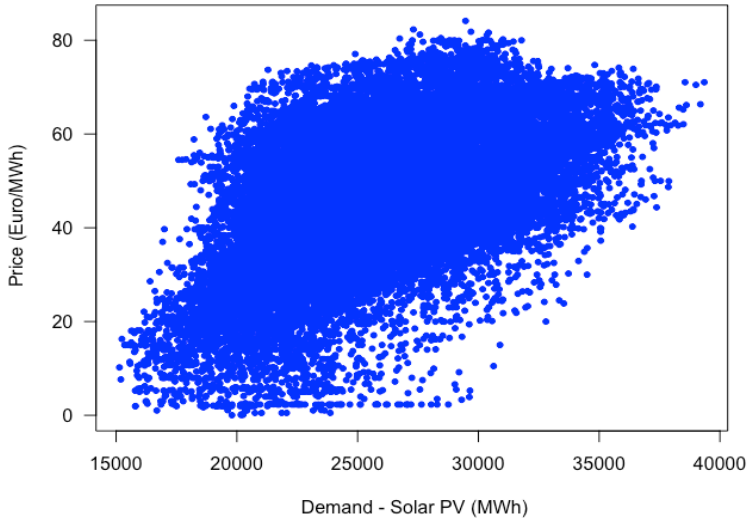


Figure 52. Price (Euro/MWh) and demand minus solar PV generation (MWh).

$$Price\left({}^{Euro}/_{MWh}\right) = -5.195 + Demand - Solar\ PV\ (MWh) \cdot 1.978 \cdot 10^{-3}$$

Correlation of 0.533, the lower the solar PV the higher the price.

- **Thermal solar**

There is also another solar technology which represents 2300 MW of capacity in the Spanish mix and it is also expected to grow the next years. As it is established in the National Plan for the energy transition, Spain will double the solar thermoelectric installed capacity by 2025 and in 2050 it is expected reach 7303 MW .[17]

Storage is a source that has been forgotten for so many years but in these days, its increasing importance in the system has attracted the interest. The goal of net zero emission electricity system will create a problem in order to supply energy in moments were renewables are not available. Storage systems will be a key in order to have the security of supply needed. This is probably the reason for the objective of increasing the installed capacity of thermal solar energy. Peak energy consumption take place after sunset and storage system can help to solve this problem.

As solar photovoltaic, solar thermoelectric has the potential to reduce prices as it has low variable cost. In this analysis this impact in price is going to be study during sun hours and also during the night because of the capability this technology has of storage energy during some hours.

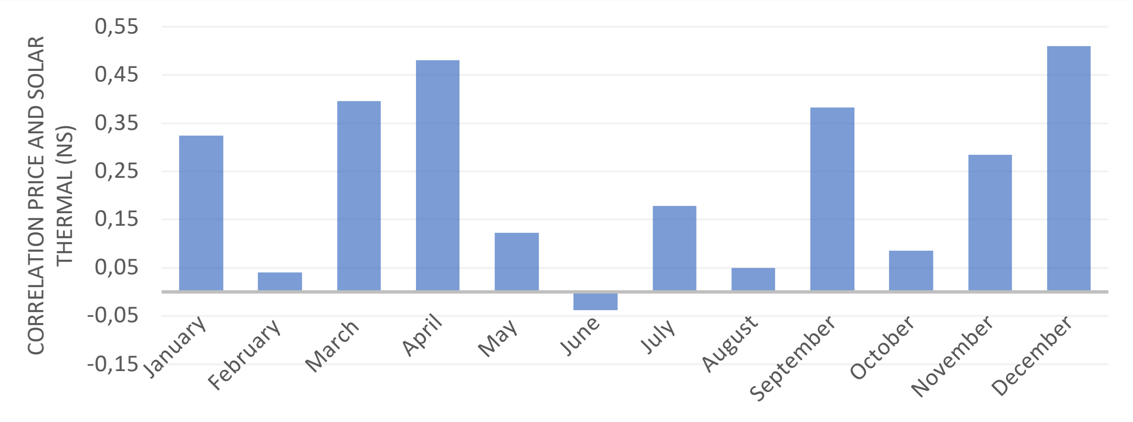


Figure 53. Correlation between prices and solar thermal generation during non -sun hours in 2018, 2019 and 2020.

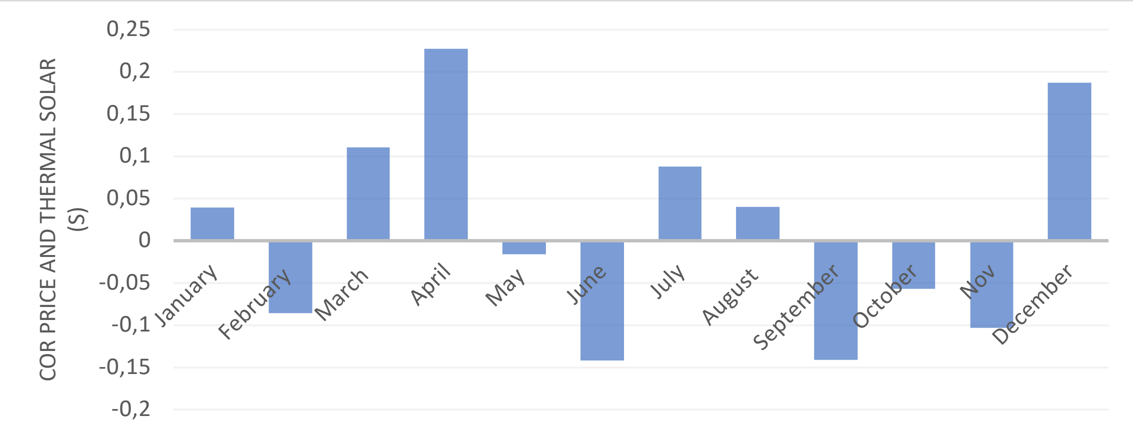


Figure 54. Correlation between prices and solar thermal generation during sun hours in 2018,2019 and 2020.

It can be observed in the figures above, both during the day and during the night this technology has not a clear impact on prices, it does not have a representative part of the mix therefore the effect by itself it cannot be seen by the correlation. As in the case of solar photovoltaic, it will be better to study the energy coming from more expensive technologies that have been avoided with this sun technology.
However, the expected increase of the install capacity in the coming years, probably will lead to higher impact in decreasing prices.

As overall, the effect of solar thermal generation it is not significant in the spot price formation. Nevertheless, as the solar forecast it is proportionate without distinction between both solar technologies, and as solar PV it has an effect of price reduction in some hours, solar forecast it is going to take into account in the price forecasting model.

5.7 Import and export effect on prices

In order to a more secure and efficient system, interconnection with border countries is crucial. Spain has interconnections with France, Portugal and Morocco. One of the first necessities to be consider in order to have a reliably system to be able to cope with the integration of renewables is the development of strongly interconnected power systems to allow energy exchanges between countries. A strong interconnected system with sufficient cross border capacity will provide the tool to achieve the decarbonization and the construction of the internal energy market in Europe.[20] Because of these Spain has several interconnection lines that are represented in the figure below.

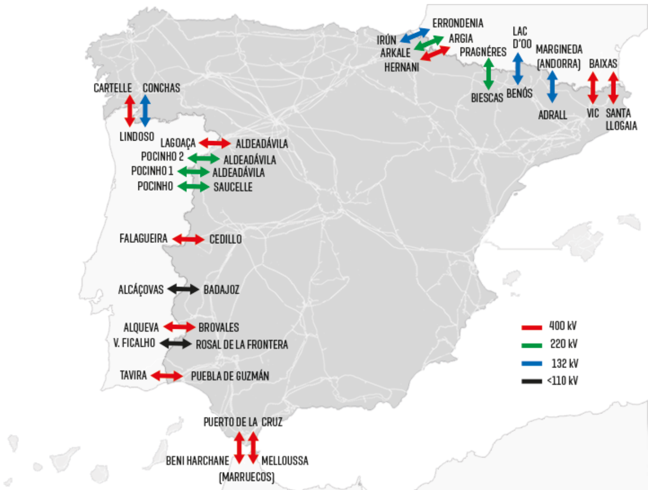


Figure 55. Interconnection lines in Spain in 2021 (Source: REE).

The European Union has established a minimum of interconnection ratio of 10% for 2020 and 15 % for the year 2030 and Spain, nowadays, only has a 2.8 %.[21] In order to achieve it, some projects have been proposed, like the project of the Bay of Biscay link, which has a capacity of 2 x1 GW and is going to provide an interconnection capacity of 4.8 GW which represents almost two times the existing one (2.8 GW). The interconnection ratio will reach 5% which is still not enough to fulfil the 2020 objective.

Spain was in the year 2020 95.9% of the hours with the same price as Portugal, on the other hand, France and Spain are only 39.3% coupled of the hours due to the not enough interconnection capacity.

[20] Roldan-Fernandez, J., Gómez-Quiles, C., Merre, A., Burgos-Payán, M. and Riquelme-Santos, J., 2018. Cross-Border Energy Exchange and Renewable Premiums: The Case of the Iberian System. *Energies*, 11(12), p.3277.

[21] Ree.es. 2021. *Strengthening interconnections | Red Eléctrica de España*. [online] Available at: https://www.ree.es/en/red21/strengthening-interconnections.

The direction of the exchange will depend on the prices of each isolated country, as we can see in *Figure 56* the country with lower price will export and the one with higher price will import. In the exporter country the price will increased as the demand is increasing, and in the importer country the price will be lower as there is more supply offered in the system. At the end the total welfare of both systems will be maximize. If there is infinite interconnection capacity the prices will be always equal but as we have said before in France are most of the time different.

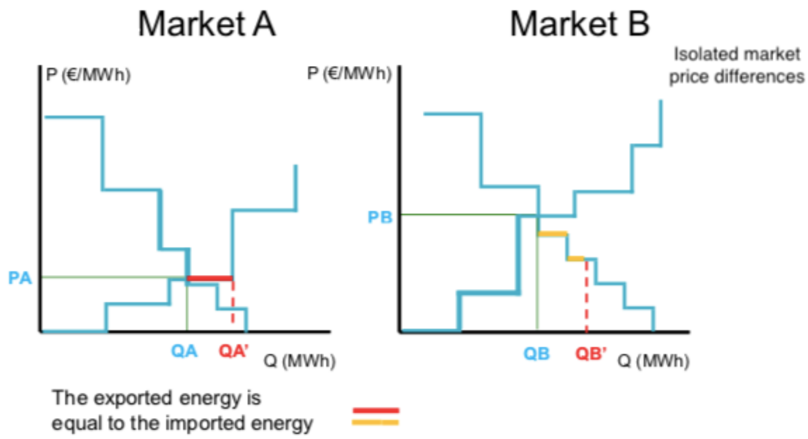


Figure 56. Graphical explanation of electricity imports and exports. (Source: OMIE).

The difference of prices it is what it is known as the spread:

$$Spread = Price\ Spain - Price\ France$$

To study the difference of prices and the imports and exports between countries, we haven't taken into account the hours were Spain and France have the same price.

In order to know if the imports and exports can give some information about the prices for the next day, we have studied the relation between the spread and the exports and imports of the day D-1 separately. The hours when there is price coupling have not been taken into account.

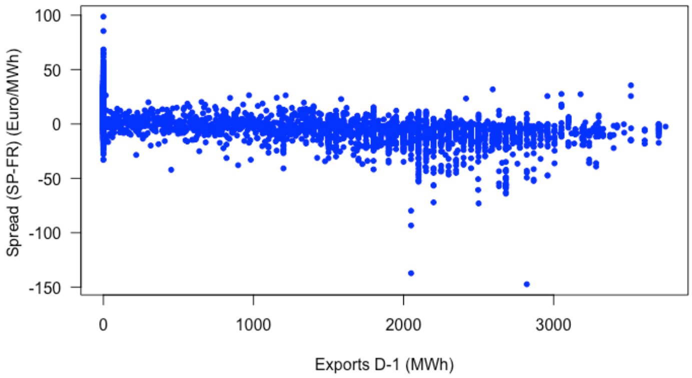


Figure 57. Spread (SP-FR) and exports of the previous day (MWh).

The correlation of these two variables is - 0.598 and the linear equation that explain this behavior is the following:

$$Spread = 10.874 - 0.0091 \cdot Exp_{D-1}(MWh)$$

The higher the exports of the day before tending to result in lower values of spread which means higher prices in France and lower in Spain. As the correlation is - 0.589 we can say that the interconnection data of the D-1 is explaining the spread correctly, lower prices in Spain when we are importing, of the day D in more than half of the hours of the year.

For the imports the same has methodology has been followed:

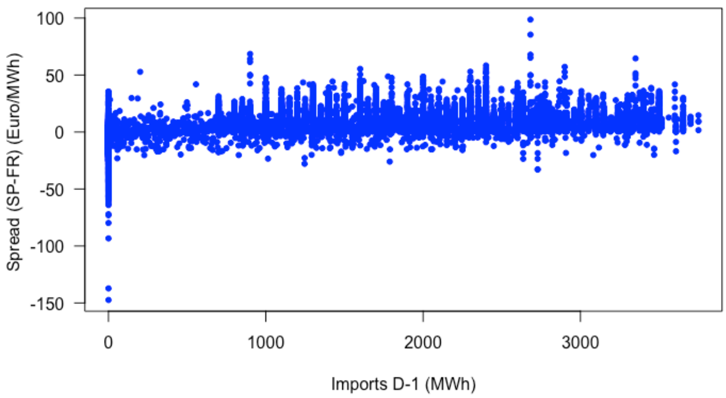


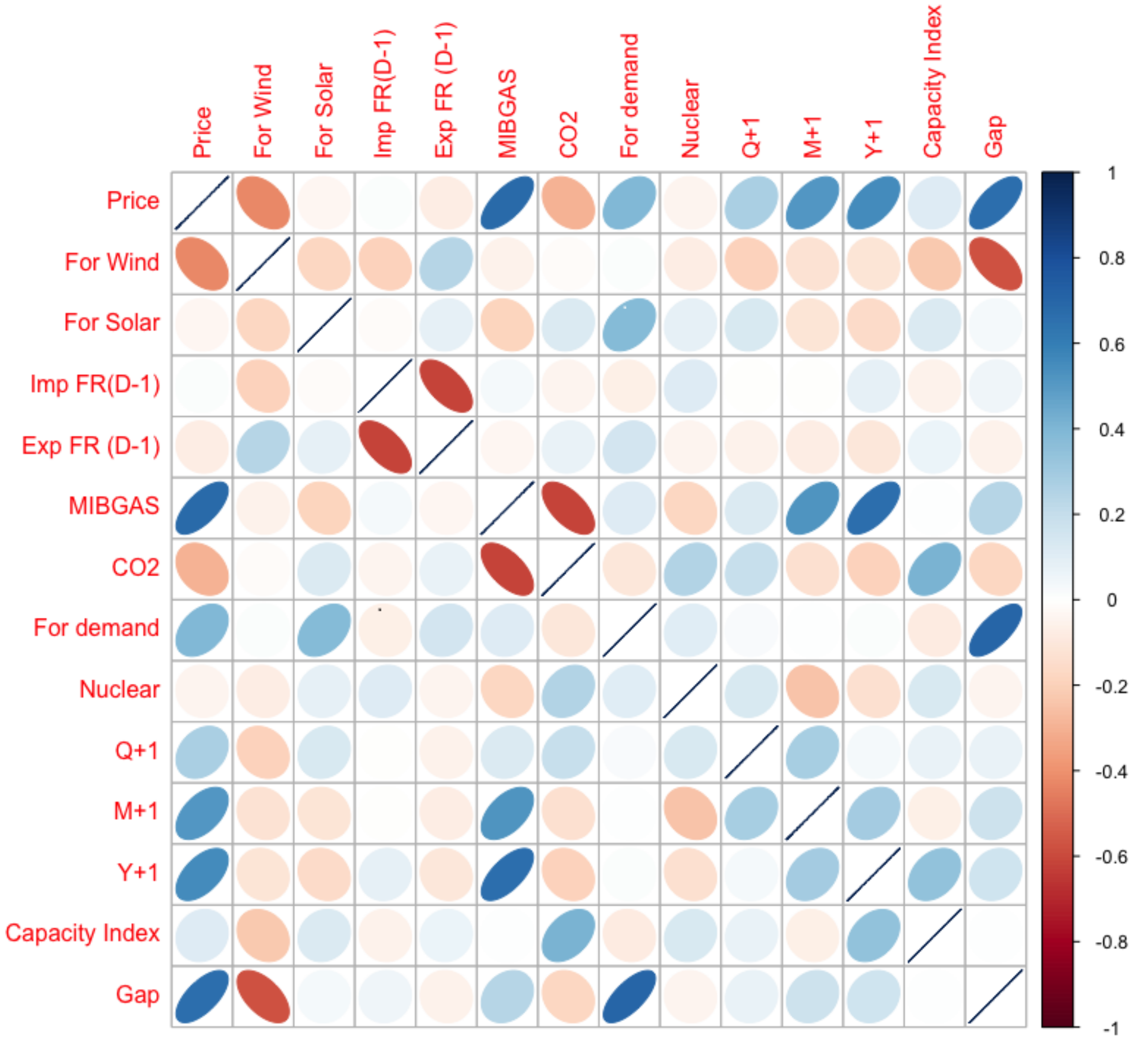
Figure 58. Spread (SP-FR) and imports of the previous day (MWh).

The correlation of these two variables is - 0.468 and the linear equation that explain this behavior is the following:

$$Spread = -1.5098 - 0.0056 \cdot IMP_{D-1}(MWh)$$

The higher the import of the day before tending to result in higher values of spread which means higher prices in Spain and lower in France. As in the same of the exporting hours it is explaining the behavior of the system correctly but in this case in less thana half of the hours.

5.8 Correlation summary



6. Forecast model and accuracy analysis

In order to evaluate if the results of the correlation study help to have a more accurate adjustment of the prices of the wholesale electricity market, we have carried out several multiple linear regression models (MLRM) based on the formula that it has been explained in section 4.

As we have seen there are considerable differences throughout the year, mostly in the case of hydro where we have seen different behavior depending on the period of the year because of the variation of the reservoir levels, the expected level for the next months and the future prices. As it has been obtained in section 5.6, solar generation lower than 3000 MWh does not have a significant impact in prices reduction and also, we have seen a clear change in the trend of this technology in the last year, based on the correlation study. In order to see if the distinction between the hours with high or low generation of solar (PV and thermal) has an effect of improvement in the linear approximation of the model, different equations for each case has been computed.

To joint everything together, 4 different MLRM has been calculated one for each quarter of the year, besides, a distinction between hour with larger and lower solar forecast than 3000 MWh.

For this analysis, it has been used hourly data for three year, 2018, 2019 and 2020. In the case of the quarters around 6000 hours.

With this study we want to demonstrate that by making a distinction between quarters, and taking into consideration variables like futures price, the capacity index that have been developed and the differentiation between periods with price reduction because of solar impact improve the accuracy of the electricity price forecasting.

Finally, an error analysis has been done to evaluate the goodness of the models.

6.1 Multiple linear regression equations

The following tables summarizes the equations obtained for each period:

Table 10. Multiple linear regression equations for the first quarter.

Multiple linear regression model for Q1		
	R^2	Equation
Q1	0.7713	Price = -22.23 - For Wind · $3.757{\cdot}10^{-3}$ - For Solar ·$3.211{\cdot}10^{-3}$ - Gap · 2.295 ·10^{-3}+ MIBGAS· 7.193 ·10^{-1} - Imp FR(D-1) · 9.527· 10^{-4} - Exp FR (D-1) · $9.403{\cdot}10^{-4}$ + M+1 · $3.352{\cdot}10^{-1}$ + Q+1 ·1.438 ·10^{-1} + Capacity Index ·21.21 + CO_2· $4.432{\cdot}10^{-1}$ + For Demand $3.571{\cdot}10^{-3}$
Q1 >Solar 1244h	0.8779	Price = -4.674 - For Wind· 1.471 10^{-3} - For Solar· $3.882{\cdot}10^{-4}$ - Gap· $2.580{\cdot}10^{-4}$+ MIBGAS· $7.646{\cdot}10^{-1}$ - Capacity Index · 14.79 + M+1 · 4.253 ·10^{-1} + For Demand· $1.512{\cdot}10^{-3}$ – CO_2 · $1.873{\cdot}10^{-1}$
Q1 < solar 2787 h	0.7883	Price = - 1.573 - For Wind· 1.735 10^{-3} - For Solar· $2.179{\cdot}10^{-3}$ - Gap· $4.861{\cdot}10^{-4}$ + MIBGAS· $8.204{\cdot}10^{-1}$ - Imp FR(D-1) · $1.291{\cdot}10^{-3}$ - Exp FR(D-1) · 2.219 ·10^{-3} - Capacity Index ·15.67 + M+1 ·2.726 ·10^{-1} + For Demand· $1.969{\cdot}10^{-3}$ + CO_2 · $6.769{\cdot}10^{-2}$

Table 11. Multiple linear regression equations for the second quarter.

Multiple linear regression model for Q2		
	R^2	Equation
Q2	0.8642	Price = -39.5 + For Solar · $1.379{\cdot}10^{-3}$ + For Wind· 6.227 ·10^{-4} + Gap· $2.3371{\cdot}10^{-3}$ + MIBGAS· 1.760 – Imp FR(D-1) · $4.903{\cdot}10^{-4}$ - Exp FR (D-1) $7.835{\cdot}10^{-4}$ - Capacity Index ·$2.144{\cdot}10^{-1}$ + M+1 ·$2.557{\cdot}10^{-3}$ + Y+1· $5.728{\cdot}10^{-1}$ - For Demand· $1.323{\cdot}10^{-3}$ + CO_2 · 1.105
Q2 >Solar 1599h	0.9087	Price = - 29.22 + For Solar · 1.140 ·10^{-3} + For Wind · $5.783{\cdot}10^{-4}$ + Gap ·2.090 ·10^{-3} + MIBGAS · 1.708 – Imp FR(D-1) · $6.707{\cdot}10^{-5}$ - Exp FR (D-1) $1.761{\cdot}10^{-4}$ - Capacity Index ·$2.712{\cdot}10^{-3}$ + M+1· $2.367{\cdot}10^{-3}$ + Y+1 · $3.777{\cdot}10^{-1}$ + CO_2 · $9.644{\cdot}10^{-1}$ - For demand · $1.023{\cdot}10^{-3}$
Q2 < solar 3491h	0.8403	Price = - 42.88 + For Solar · 1.575 ·10^{-3} + For Wind · $5.529{\cdot}10^{-4}$ + Gap ·2.431 ·10^{-3} + MIBGAS · 1.787 - Imp FR(D-1) · $6.901{\cdot}10^{-4}$ - Exp FR (D-1) $1.199{\cdot}10^{-3}$ - Capacity Index ·$2.893{\cdot}10^{-1}$ + M+1· $2.573{\cdot}10^{-3}$ + Y+1 · $6.185{\cdot}10^{-1}$ + CO_2 · 1.179 - For demand · 1.376 ·10^{-3}

Table 12. Multiple linear regression equations for the third period.

Multiple linear regression model for Q3		
	R^2	Equation
Q3	0.9273	Price = -15.58 – For Solar $1.477\cdot10^{-3}$ – For Wind $1.584\cdot10^{-3}$ – Gap $5.459\cdot10^{-4}$ + MIBGAS $\cdot1.367$ + For Demand $\cdot$ $1.581\cdot10^{-3}$ + M+1 $\cdot$ $1.305\cdot10^{-1}$ – Imp FR (D-1) $7.483\cdot10^{-4}$ - Exp FR (D-1) $\cdot1.690\cdot10^{-4}$ + CO_2 $\cdot4.326\cdot10^{-1}$
Q3 >Solar 2534 h	0.9463	Price = -13.01 – For Solar -1.716 $\cdot$ 10^{-3} – For Wind $1.903\cdot10^{-3}$ – Gap $8.034\cdot10^{-4}$ + MIBGAS $\cdot1.497$ + For Demand $\cdot1.702\cdot10^{-3}$ + M+1 $\cdot$ $6.945\cdot10^{-2}$ – Imp FR (D-1) $4.821\cdot10^{-4}$ - Exp FR (D-1) $\cdot3.060\cdot10^{-5}$ + CO_2 $\cdot4.635\cdot10^{-1}$
Q3 < solar 4092h	0.9165	Price = -15.76 – For Solar -1.228 $\cdot$ 10^{-3} – For Wind $1.227\cdot10^{-3}$ – Gap $2.052\cdot10^{-4}$ + MIBGAS $\cdot1.278$ + For Demand $\cdot1.323\cdot10^{-3}$ + M+1 $\cdot$ $1.718\cdot10^{-2}$ – Imp FR (D-1) $8.822\cdot10^{-4}$ - Exp FR (D-1) $\cdot3.862\cdot10^{-4}$ + CO_2 $\cdot4.237\cdot10^{-1}$

Table 13. Multiple linear regression equations for the fourth quarter.

Multiple linear regression model for Q4		
	R^2	Equation
Q4	0.8586	Price = -28.22 - For Solar 3.621 $\cdot$ 10^{-3} - For Wind $3.981\cdot10^{-3}$ - Gap $2.630\cdot10^{-3}$ + MIBGAS $\cdot$ 1.745 + For Demand $\cdot$ $4.031\cdot10^{-3}$ - M+1 $\cdot$ $4.278\cdot10^{-1}$ + Y+1 $\cdot$ $5.348\cdot10^{-1}$ + Q+1 $\cdot2.072\cdot10^{-1}$ - Imp FR (D-1) $1.065\cdot10^{-3}$ - Exp FR (D-1) $\cdot$ $4.591\cdot10^{-4}$ - CO_2 $\cdot$ $2.624\cdot10^{-1}$ - Capacity Index $\cdot$ $4.475\cdot10^{1}$
Q4 >Solar 1036h	0.8748	Price = 36.92 - For Solar $2.003\cdot10^{-3}$ - For Wind $2.385\cdot10^{-3}$ - Gap $1.122\cdot10^{-3}$ + MIBGAS $\cdot1.249$ + For Demand $\cdot2.512\cdot10^{-3}$ + M+1 $\cdot$ $2.323\cdot10^{-1}$ - Q+1 $\cdot$ $8.940\cdot10^{-1}$ + Y+1 $\cdot$ $5.670\cdot10^{-1}$ - Capacity Index $\cdot$ $6.693\cdot10^{1}$ + CO_2 $\cdot1.984\cdot10^{-1}$
Q4 < solar 5590h	0.8593	Price = -5.910 - For Solar $\cdot$ $3.394\cdot10^{-3}$ - For Solar $3.945\cdot10^{-3}$ - Gap $2.641\cdot10^{-3}$ + MIBGAS $\cdot$ 1.720 + For Demand $\cdot$ $4.091\cdot10^{-3}$ - M+1 $\cdot$ $4.549\cdot10^{-1}$ + Q+1 $\cdot3.063\cdot10^{-1}$ + Y+1 $\cdot5.085\cdot10^{-1}$ - Imp FR (D-1) $1.192\cdot10^{-3}$ - Exp FR (D-1) $\cdot1.476\cdot10^{-3}$ - Capacity Index $\cdot$ $4.242\cdot10^{1}$ - CO_2 $\cdot3.146\cdot10^{-1}$

By doing the regression of all the months together we obtain the following equations:

Table 14. Multiple linear regression equations for all the years 2018,2019 and 2020.

R^2	Equation
0.792	Price = - 9.480 - For Solar -1.200 · 10^{-3} - For Wind 1.729 ·10^{-3} –-Gap 1.975 ·10^{-4} + MIBGAS ·1.468 + For Demand ·1.402 ·10^{-3} + CO_2 · 3.329·10^{-1}
0.794	Price = -9.241 - For Solar -1.240 · 10^{-3} - For Wind 1.724 ·10^{-3} - Gap 2.2 ·10^{-4} + MIBGAS ·1.471 + For Demand ·1.455 ·10^{-3} - Imp FR (D-1) 2.391·10^{-4} - Exp FR (D-1) ·8.226· 10^{-4} + CO_2 ·2.950·10^{-1}
0.826	Price = -32.521 - For Solar · 1.122·10^{-3} - For Wind 1.567 ·10^{-3} - Gap 1.359·10^{-4} + MIBGAS · 7.677·10^{-1} + For Demand· 1.397·10^{-3} + M+1 · 2.502·10^{-1} + Q+1 · 2.140 ·10^{-1} + Y+1 · 4.291·10^{-1} - Capacity · 3.839 - CO_2 ·4.240·10^{-2} - Imp FR (D-1) 1.101·10^{-3}- Exp FR (D-1) · 1.415 10^{-3}

As we can see there are considerable differences of accuracy between the quarters, resulting the third one with the better R- square values. Moreover, only in the first quarter the adjustment of the equation to the price is worse than in the case with all the data of the years together.

6.2 Model simplification

As we have seen in the equations that describe each quarter, we are using a large number of variables and the results that we have obtain are good (R-square is close to 1). On the other hand, there are some variables with significant more weight than others. In this chapter we are going to analyze how the model works if the equations are simplified to the most representative variables. We have tried different combination of the variables with more weight until achieving the best R-square value with the least number of independent variables.

Table 15. Model simplification equations for each quarter.

First quarter	Price = 44.888 + MIBGAS ·1.014 + CO_2 ·1.176 - Capacity Index ·58.133 R^2 = 0.3787
Second Quarter	Price = -12.250 + CO_2 · 1.167 + MIBGAS ·2.113 R^2 = 0.6667
Third quarter	Price = 3.525 + MIBGAS ·1.855 + CO_2 ·0.662 R^2 = 0.8276
Fourth quarter	Price = 13.330 + MIBGAS · 1.830 + CO_2 · 0.531 - Capacity Index · 7.683 R^2 = 0.5876

The results that are obtained with the simplification are very acceptable except for the first case where the adjustment it is the worst of the four cases.

6.3 Accuracy analysis

In order to study the goodness of the model, the difference of the real and forecast price has been calculated and later on a graphic representation has been plotted. Moreover, a histogram of the difference of prices has been developed to illustrate the most relevant descriptive statistical values. Also, the results from the first equations and the simplified ones are going to be compared.
In this part, the data that has been used was the same as the one used for the development of the model in order to see if the equations are describing properly each period. In the next chapter some of the equations are going to be used with external data.

- **First quarter model errors**

The following table represented the statistical values obtain from the error of the first quarter model. As we can see the mean and the median are very close to zero so that means that our data has a clear tendency of going around this value. The quartiles are quite balance, but it has more weight the third one, hence the model is given lower values than the real ones more time than higher.

Table 16. Summary of statistical descriptive values of the first quarter.

Min.	-34.538
1st Qu.	-3.070
Median	0.188
Mean	0.001
3rd Qu.	3.682
Max.	30.145

The histogram representation shows more clearly the values and as it can be seen, the middle of the curve is around 0, which is a good sign. Also, it can be appreciated that the right part, which corresponds to the third quartile, reach higher values as we have seen in the statistical descriptive values.

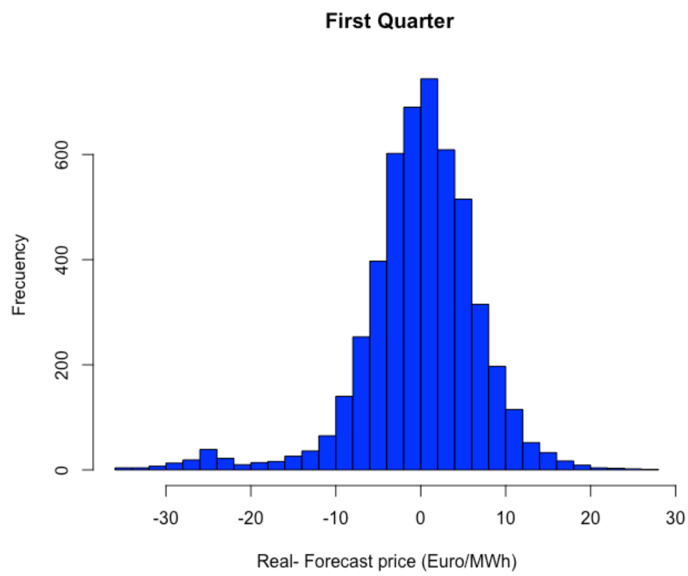


Figure 59. Histogram representation of the error of Q1 equation.

Furthermore, we can observe in the last part of the figure below a shift in the trend where the model starts to give higher values than the real ones very drastically. This alteration corresponds to the year 2020 where there was very low prices and lower demand due to Covid-19.

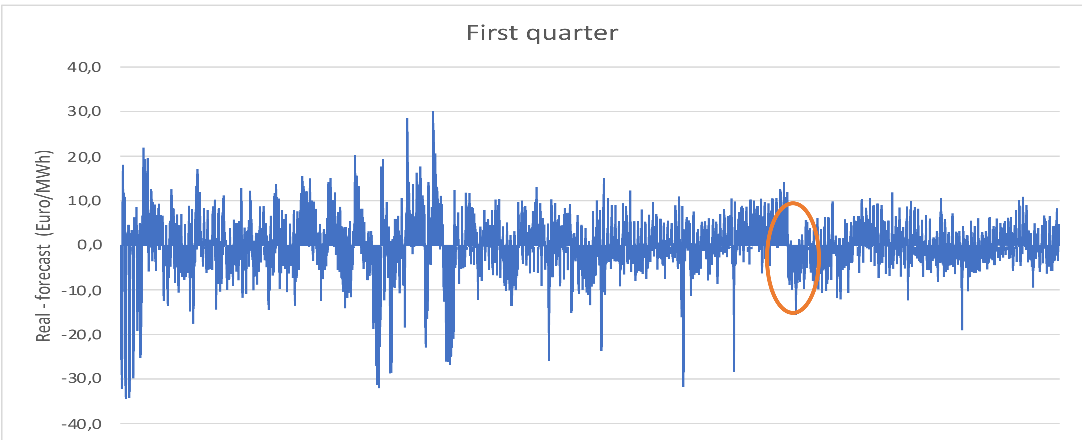


Figure 60. Difference of the real and the forecast hourly prices obtained with the model for Q1.

- **Second quarter model errors**

As in the first quarter, the mean is around zero and again the model tends to forecast lower values than the real ones as the third quartile has more weigh. This fact can be also seen in the histogram of the errors of this period.

Table 17. Second quarter errors statistical analysis.

Min.	-33.406
1st Qu.	-2.608
Median	0.724
Mean	0.433
3rd Qu.	4.033
Max.	17.195

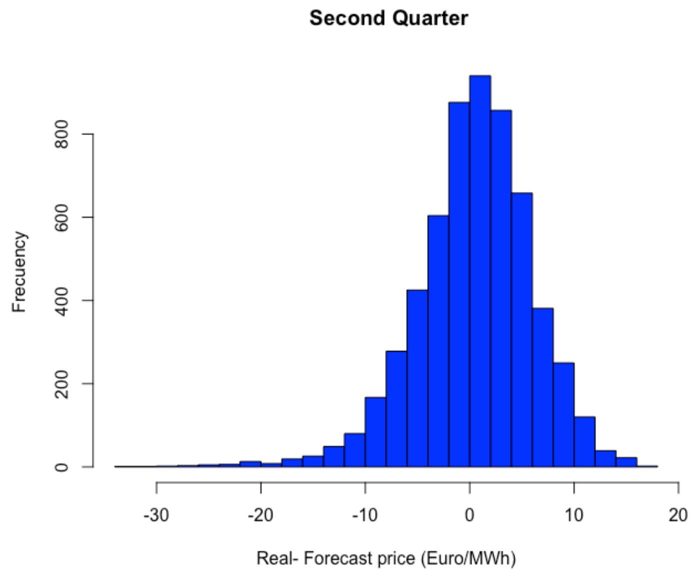


Figure 61. Histogram representation of the error of Q2 equation.

Figure *62,* like in the previous quarter, illustrate the change in the error due to the unexpected prices of Covid-19. However, the errors start to be smoother at the end of the graph which is a sign that the model it is stabilizing.

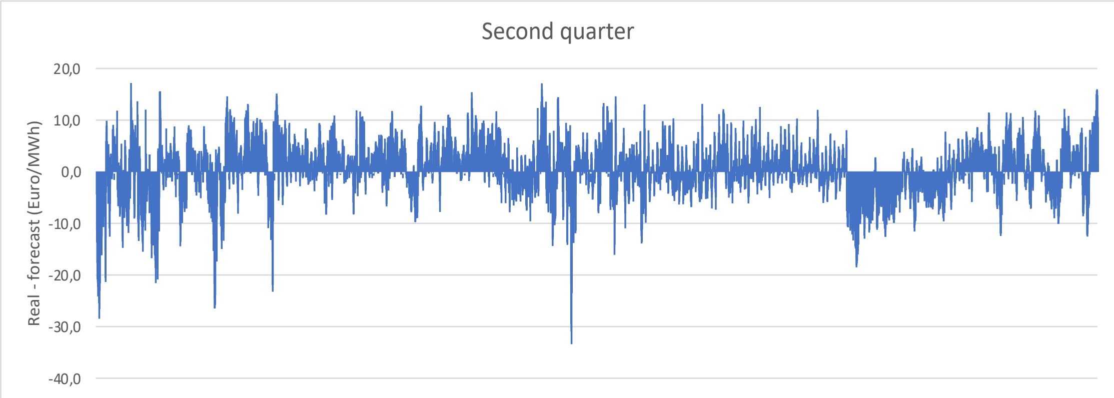


Figure 62. Difference of the real and the forecast hourly prices obtained with the model for Q2.

- **Third quarter model errors**

In the model of the third quarter, we can see an improvement of the results compared to the previous cases. The mean is even more near zero than in the other cases and the quartiles are balance. In this case the same number of times the model forecast prices lower or larger than the real.

Table 18. Third quarter errors statistical analysis.

Min.	-13.987
1st Qu.	-2.336
Median	-0.097
Mean	0.016
3rd Qu.	2.337
Max.	13.386

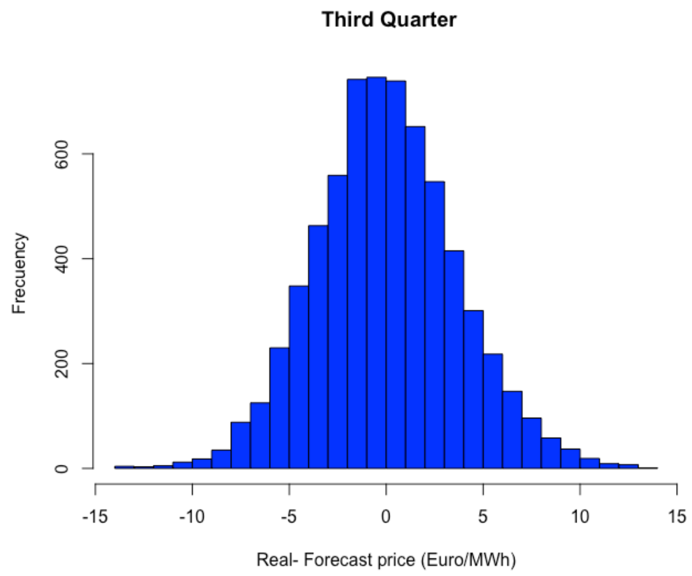


Figure 63. Histogram representation of the error of Q3 equation

In contrast with the previous cases, in this third period of the year we do not see the change in the trend of the error due to Covid-19. The error is smoother over the three years.

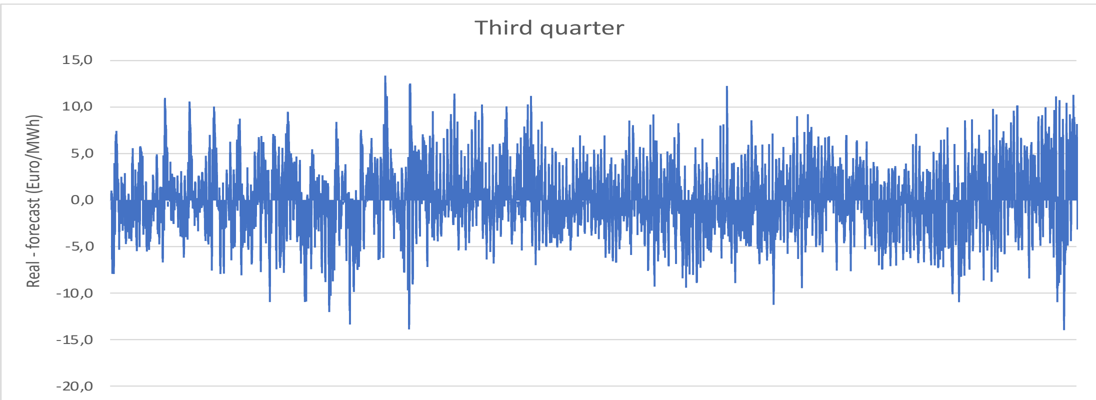


Figure 64. Difference of the real and the forecast hourly prices obtained with the model for Q3.

- **Fourth quarter model errors**

The model for the last quarter of the years 2018, 2019 and 2020 also it is also explaining with a good accuracy the prices throughout the period. The mean is higher than in the previous case, but it is still very close to zero. Furthermore, the quartiles are very balance which is a good characteristic of the model.

Table 19. Fourth quarter errors statistical analysis.

Min.	-28.618
1st Qu.	-3.215
Median	0.291
Mean	0.078
3rd Qu.	3.626
Max.	23.743

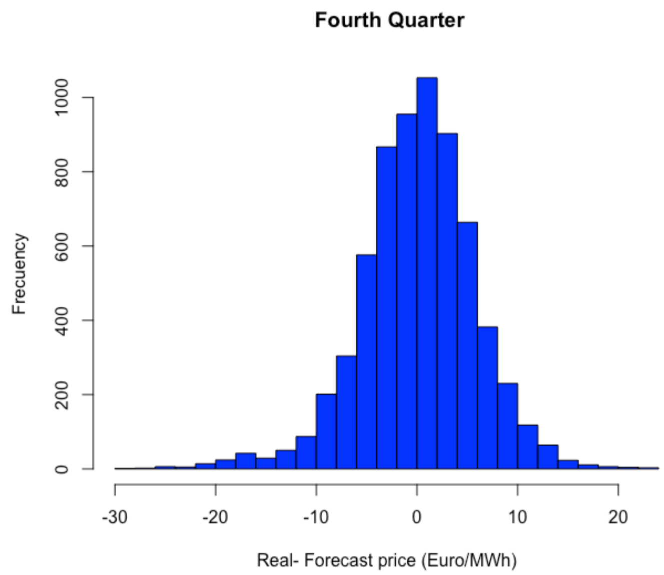


Figure 65. Histogram representation of the error of Q4 equation.

We can observe in the histogram of the error that there is a little more weight in the negative side therefore the model is given more times higher values than the real ones. However, it is very balance.

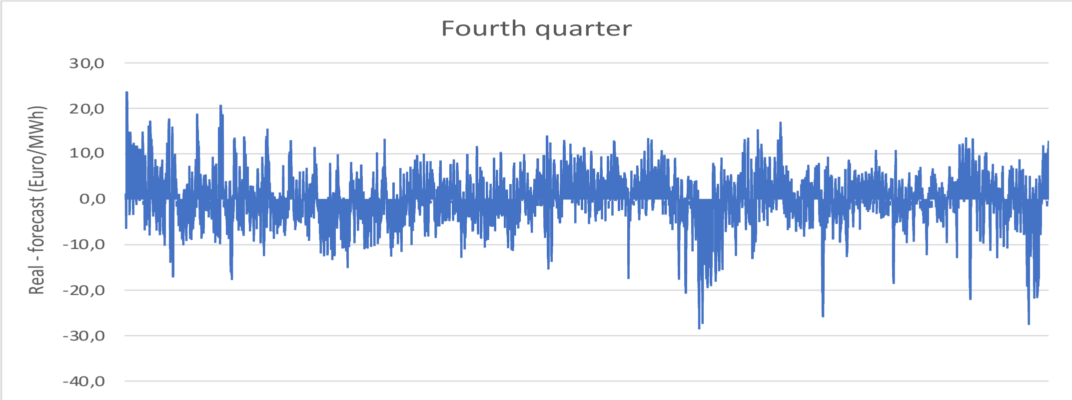


Figure 66. Difference of the real and the forecast hourly prices obtained with the model for Q4.

To sum up, in all the cases there is larger number of data that achieve values of error near zero as the highest frequency in the histogram representation is in that value.
Also, in the first, second and fourth quarter, the model tends to predict more times values lower than the real ones with the exception of the third quarter, which has the same value in both first and third quartile.
It can be concluded that the model that can describe in a more precise way the price during the period of study is the third quarter.

Finally, the errors of the simplified equations have been also computed:

- First quarter simplification errors

Table 20. First quarter simplified model errors statistical analysis.

Min.	-51.110
1st Qu.	-5.466
Median	1.938
Mean	-0.008
3rd Qu.	7.345
Max.	28.379

Based on the statistical values presented in *Table 20* the simplification of the first quarter equations it is not useful. The errors have much more weight in the positive sign therefore the model is given lower values than the real ones most of the time.

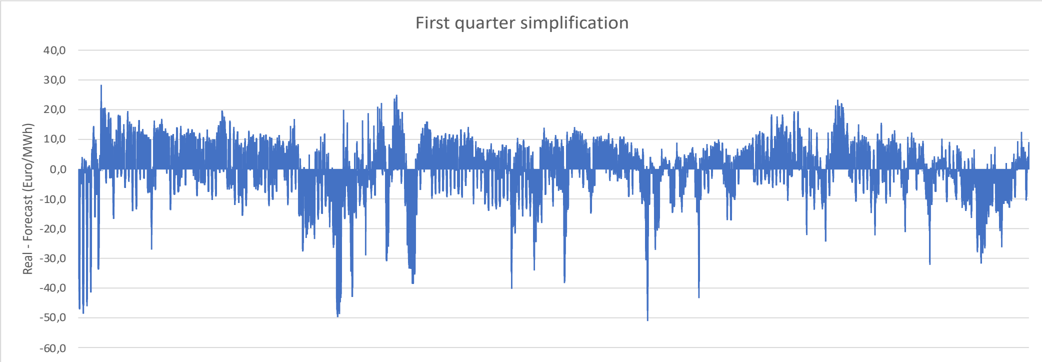


Figure 67. Difference of the real and the forecast hourly prices obtained with the simplified model for Q1.

We can observe in the figure above that the errors are more concentrated in the in the positive side (the model forecast lower prices than the real ones). On the other hand, the hours where the forecast it is making mistakes in the negative side (higher forecast that real) the error is much larger. It could be interesting to study why this is happening in further studies.

The histogram graph shows more clearly how all the data it is more concentrated in the right-hand side.

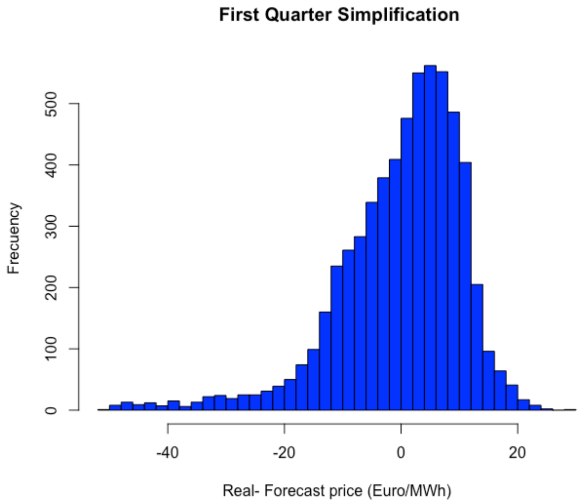


Figure 68. Histogram representation of the error of Q1 simplified model.

We can conclude than in the first quarter will be better to use the multiple linear equation model with all the variables.

- Second quarter simplification model errors

During this period the simplification results are better. In this case, both quartiles are more balance which means that the model tends to forecast the same number of times larger and lower prices than the real ones. Moreover, the mean error is very near zero.

Table 21. Second quarter simplified model errors statistical analysis.

Min.	-48.930
1st Qu.	-3.720
Median	1.730
Mean	0.000
3rd Qu.	5.813
Max.	19.660

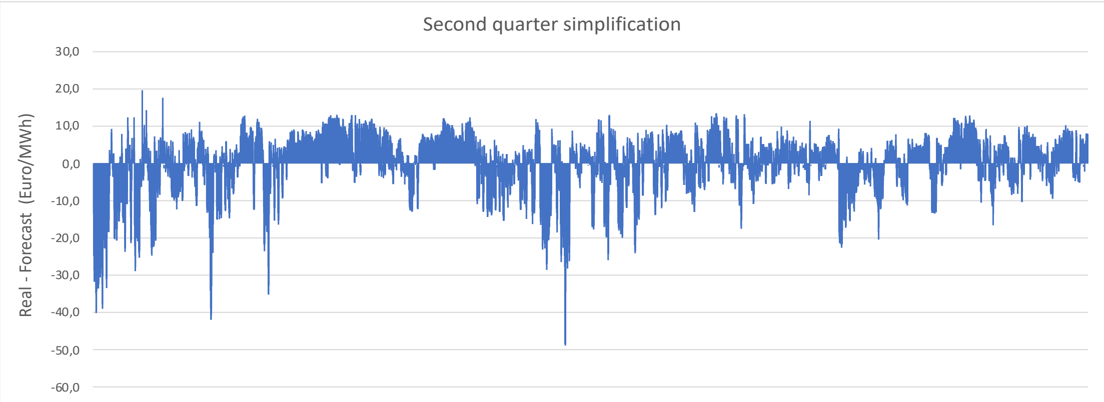


Figure 69. Difference of the real and the forecast hourly prices obtained with the simplified model for Q2.

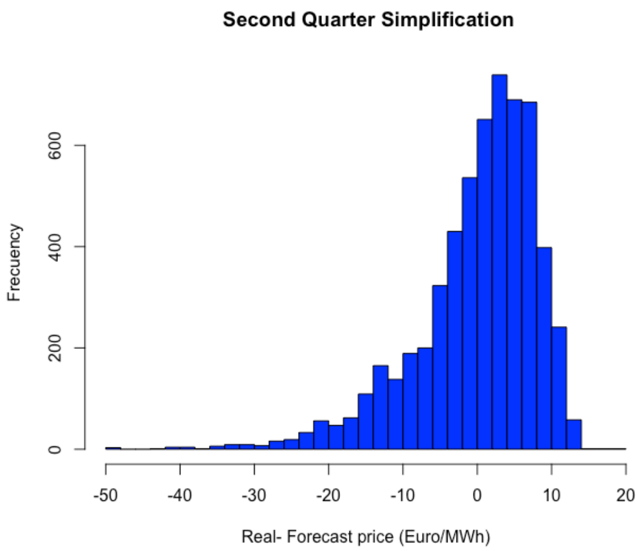


Figure 70. Histogram representation of the error of Q2 simplified model.

For the second quarter, could be interesting to take into account this simplification as the result is very similar to the ones obtained with the full equations and it is simpler and faster to calculate since the number of variables is much lower.

- Third quarter simplification model errors

The third quarter is showing the best results up to now, the mean it is close to zero and the quartiles are low and balance.

Table 22. Third quarter simplified model errors statistical analysis.

Min.	-27.900
1st Qu.	-3.600
Median	1.100
Mean	0.007
3rd Qu.	4.200
Max.	18.500

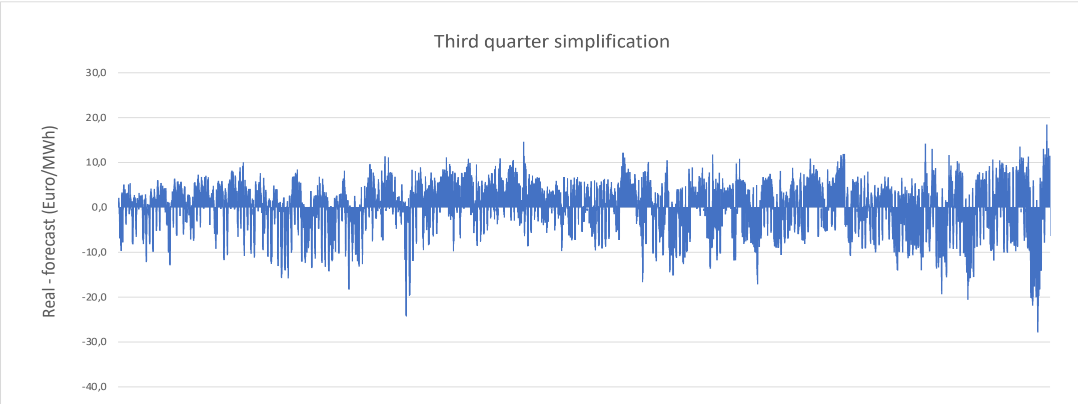


Figure 71. Difference of the real and the forecast hourly prices obtained with the simplified model for Q3.

In the figure above we can see how the error are distributed very well along the period of the study.

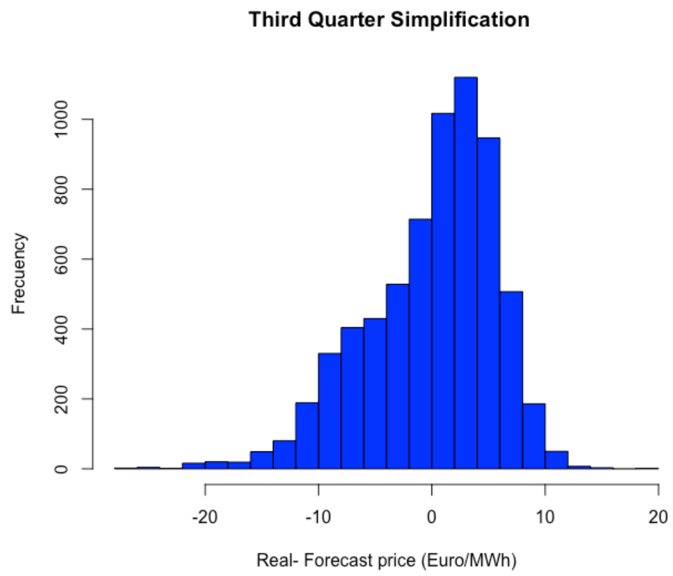


Figure 72. Histogram representation of the error of Q3 simplified model.

Because of the good results of this simplification, in this period will be good to use this formula than the complete one.

- Fourth quarter simplification model errors

Finally, in the last period the results of the adjustment of the simplified equation are good and will be consider in the prediction of this period.

Table 23. Fourth quarter simplified model errors statistical analysis.

Min.	-42.839
1st Qu.	-5.757
Median	1.828
Mean	0.000
3rd Qu.	6.615
Max.	25.514

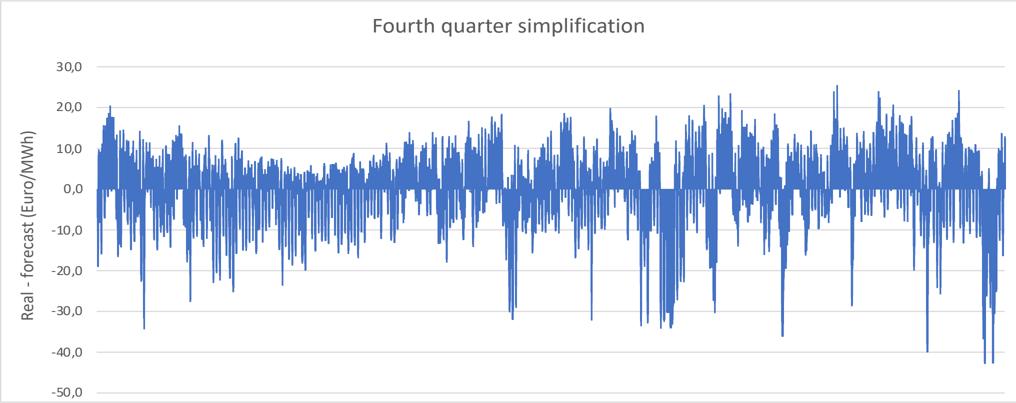


Figure 73. Difference of the real and the forecast hourly prices obtained with the simplified model for Q4.

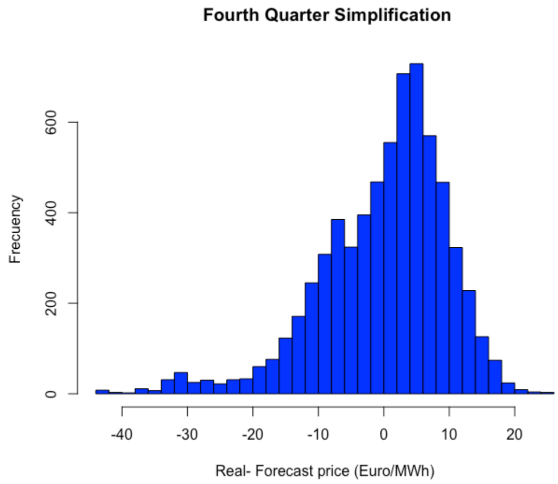


Figure 74. Histogram representation of the error of Q4 simplified model.

As we can see, in all the cases the maximum (negative and positive) errors have increased. There is a larger increased in the error in the positive part, which means that it is forecasting more times higher prices which is coherent as we are not considering any price reduction variables like wind or solar. The side of negative errors is very similar as in the case with all the variables.

Even if we have an increased in the maximum errors the values are good as in all the cases the highest frequency corresponds with the area near zero.

This specific simplifications for each period are very powerful. The number of variables it is been greatly reduced and the results are in all the cases except for the first one very similar to the models where we have used all the variable

7. Model verification with external data

To end up with the study, the model has been tried with data for the current year. We have used the January for the first quarter model and April for the second quarter model. Even if the year 2021 it is characterized with an unexpected behavior due to the very high prices, we have thought that could be interest to see how the equations can adjust this period. We have obtained the following results:

Firstly, using the equations for the first quarter in general and later on for the hours with larger solar generation than 3000 MWh.

The following table summarize the value of the errors (real – forecast price) when using the first quarter equation.

Table 24. January 2021 errors statistical analysis of the forecast.

Min.	-37.702
1st Qu.	-4.576
Median	2.799
Mean	2.254
3rd Qu.	10.828
Max.	43.096

As it can be seen, the higher number of values are around 2 as we can see in *Table 24*, which is not a very high error comparing with the results that we have obtain when calculating the error with the same data that we have used for the calculation of the equations.

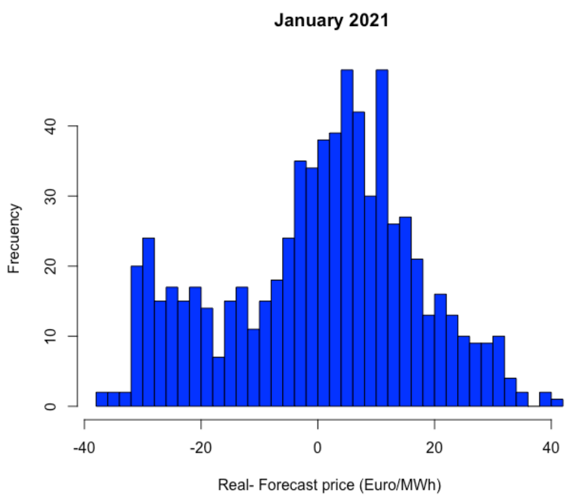


Figure 75. Histogram representation of the error from the price forecast of January 2021.

In the following figure both the real and forecast price has been represented graphically in order to help to see the accuracy of the results.

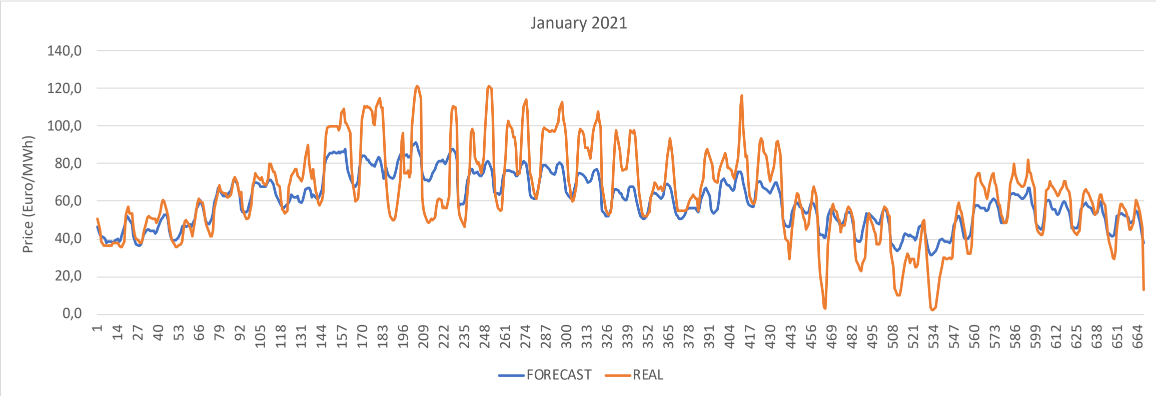


Figure 76. Real and forecast prices of January 2021 graphical representation.

As we can see the model it is making more mistakes in the peaks and the valleys, the reason can be that during this period of 2021 the Spanish market was reaching historical maximum and minimum prices. [22]

For the second quarter, we have done the same as in the first case. April 2021 has been chosen in order to the evaluate the accuracy of the formulas for this period.
Table 26 summarize the errors obtained in this forecast. As we can see this case presents higher errors than the first quarter model. The mistakes are more concentrated in the negative side. The model is forecasting higher values than the real ones

Table 25. April 2021 errors statistical analysis of the forecast.

Min.	-59.981
1st Qu.	-16.671
Median	-11.258
Mean	-11.865
3rd Qu.	-5.706
Max.	13.792

The histogram representation also illustrated that most of the hours the model is predicting prices higher than the real ones, as the histogram is moved to the left side that correspond with negative values.

[22] Elperiodicodelaenergia.com. 2021. *Un 'pool' de enero para la historia: la extrema volatilidad del mercado eléctrico hace ver precios récord máximos y mínimos*. [online] Available at: <https://elperiodicodelaenergia.com/un-pool-de-enero-para-la-historia-la-extrema-volatilidad-del-mercado-electrico-hace-ver-precios-record-maximos-y-minimos/>.

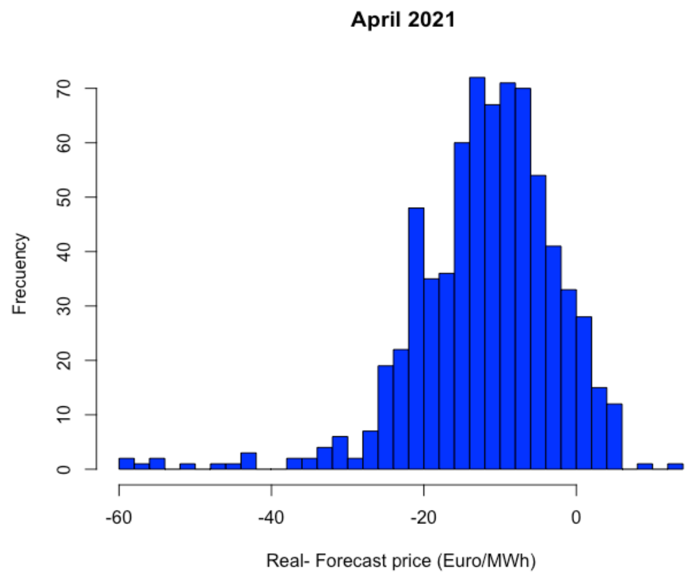


Figure 77. Histogram representation of the error from the price forecast of April 2021.

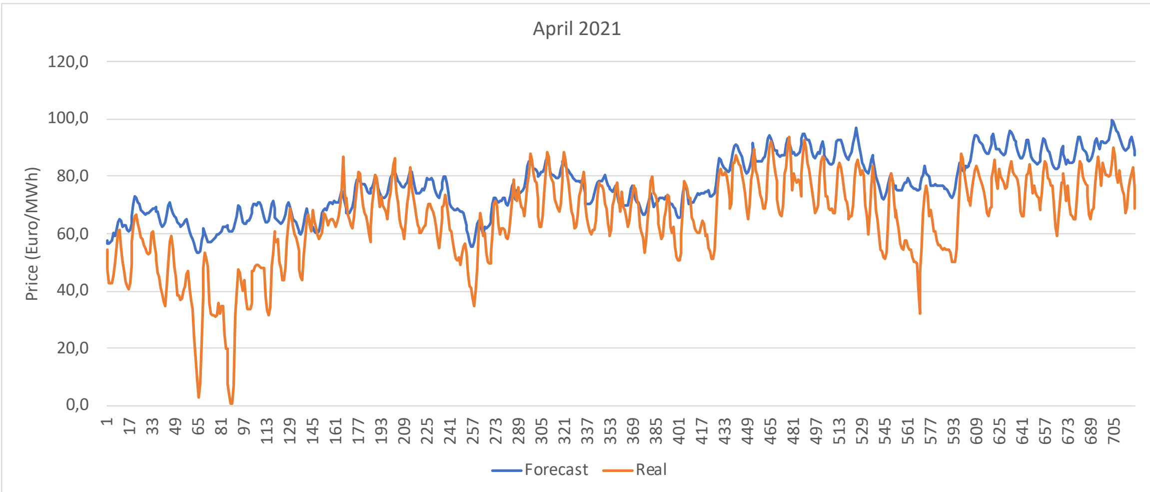


Figure 78. Real and forecast prices of April 2021 graphical representation.

An explanation of this result can be that the equation of this period was given a significant weight to MIBGAS and CO_2. Carbon prices in this period where considerable high, reaching values of 45 euros/ton.

In both cases the simplified formulas and the ones with the distinction between hours with higher and lower solar PV generation has been computed but with have not achieve good results. The historical characteristic of the year 2021, where we have reach prices that we cannot find in the years that we have used for the construction of the mode, can be the reason of the lack of adjustments of the models. Only in the case of January the results are good.

8. Results and discussion

In this section we are going to compare the results of the variables analysis and the linear regression analysis.
First of all, the variables that we have used in order to adjust the price dynamics, have been selected based on the significance criterion that has been explain in section 4. Due to this, the set of independent variables that describe each of the periods are the ones that are given useful information.

In order to take into account, the results that we have found in the hydroelectric generation analysis, we have distinguished between quarters to capture the effect of the behavior of the agents during the different seasons of the year.

In the case of the first quarter, M+1 and Q+1 futures price is significant in the model. This was expected based on the results of the correlation analysis, where these future prices were showing the highest correlation with hydro generation. Also, we have found that the capacity index it is significant in the forecast. It has an effect of price reduction as we have also seen in the correlation analysis. It is important to notice, that in this period we have not seen any effect of the difference between spot and future price in the hydro dispatch decisions. Therefore, the significance of the futures in the model can be due to the similar trend that futures and spot prices follow.

In the second quarter, yearly futures are significant. This is also in line with the correlation values where we have computed a correlation between Y+1 future price and the hydro schedule generation of -0.7362. This means that they are taking into account the opportunity cost of producing now or the year before. Monthly futures also result with a high significance level. Capacity index it is also relevant in this period with an effect of price reduction in contrast with what we have obtain the correlation analysis where we have not found a linear behavior between these variables.
During the third quarter it is interesting to see that capacity index it is not significant. As it is a period when the reservoirs are at minimum levels, the agents will produce only when they can have higher economic return, therefore, they may not take into consideration the water level in the reserves. In contrast, they will take into account the difference between spot price and future price. In this case, only monthly futures are significant. This make sense because in this period, there is not enough capacity to could think about producing in the next quarter or year.

In the last quarter which correspond with October, November and December, the capacity index starts to be important again due to fact that the reservoir levels start increasing so this will end up in a price reduction. Capacity index shows correlation with prices and it is significant in the equation. Furthermore, futures are also relevant again.

It is important to notice that in some cases future it is significant just because it follows a similar trend with spot prices. As we have seen in the hydro correlation analysis, only Y+1 future presents a more different trend with respect to the wholesale price, the others (M+1 and Q+1) follow a very similar trend. In other cases, future prices are relevant due to the opportunity cost that hydro agents see when they have to decide when to produce. This is the case of the third quarter, where we have seen in *Figure 33*, the strong relation between future and spot price variation percentage and hydro production in the system. The higher the future the lower the hydro production.
In the rest of the quarters this effect is not as clear as in the third one.

Based on the results obtained in multiple linear regression equations, we have seen that the addition of the futures prices, the capacity index and the imports and exports of the previous day have increased the accuracy of the model. We can say that these variables that we know the day before (D-1) of the day ahead market clearing can help to predict the price in the day D. We have improved the adjustment of the equation from 0.792 to 0.826, as it can be seen in *Table 14,* so the variable that we have add are helping to explain the price behavior during the period of study.

Secondly, we are going to analyze the results of the equations that have been obtained by doing the distinction between hours with higher than 3000 MWh of solar generation and hours with lower than 3000 MWh. As we have said before, we have decided to do it because we have found a substantial change in the trend of the price with respect to the solar generation up to 3000 MWh. We have found an enhancement of the results of the R-square values, however, when trying these models with external data the results are not good. This can be due to the low number of data that we have in this case as we have used only one month and also because the months that we have choose are have not a large share of solar. Further studies will be needed to determine if this distinction between hours with large or low solar generation, it is truly improving the results when trying to explain the electricity price.

In the model simplification section, we have tried to achieve a simpler equation with enough goodness. As we have seen all the errors are due to the lack of price decreasing variables. However, the results are quite good in all the cases except for the first quarter. The third quarter has the best results with an R-square of 0.8276, very close to the values when using all the variables. This can be explained because during this period, gas prices are very significant due to the fact that hydro it may be only dispatching in hours where it can maximize its profit to compensate the low reservoir capacity. These simplifications does not mean that the order variables are not given information, but having a simpler equation fast to compute, it is more powerful that having models with a large set of variables.

In the accuracy analysis we have seen that as overall, the characteristic of each of the price divers that we have found in the variable's analysis are given useful information to predict the price behavior, as the mean of the error (difference between real and forecast) it is around zero in all the cases.

Finally, in the last part of the project, where we have used external data to evaluate the goodness of the model, we have found that in the first quarter the adjustment of the equation is quite good. On the other hand, in the second period the accuracy of the model decreases sharply. Most of the times the model was predicting prices larger than the real ones. This change can be due to the fact the year 2021 the market is reaching historical minimum and maximum spot prices. Moreover, the remarkable increasing trend of carbon prices could influence the results of the model.
When using the specific formula for the hours with larger solar photovoltaic generation than 3000 MWh we do not see a relevant improvement in the results. The reason can be that in the months that we have used, January and April, solar generation it is not very relevant in the system.
As overall, the model was adjusting very well the period that we have used to compute the formulas, however, when we use it with external data, the forecast of January with the no simplified formula, is the only case that is given reasonable results
Further studied will be needed in order to try the equations in a period with more common scenarios of the electricity market. Also, can be consider the recalibration of the model in order to include in the computation of the multiple linear equations data from 2021.

9. Conclusions

All the variables that we have used and verify in the correlation analysis have a significant impact in the wholesale electricity price dynamics in Spain. We have obtained the following conclusions:

- MIBGAS prices have the highest correlation with spot prices with a value of 0.710. It is the highest correlation that we have found in the study

- Carbon prices produce an increase of the electricity prices as it is internalized in the agent's bids. However, it has been found that the correlation between the two prices mention before decrease in some periods with high renewable share in the system and low demand. The behavior that has been found can be a future trend in the electricity system.

- The capacity index, that has been computed as a value to reflect the capacity of the reservoirs, has an effect in the hydro dispatch more or less relevant depending on the period. Also, it is important to notice that in all the periods of the year, lower values of capacity restrict the schedule dispatch. On the other hand, large capacity values can reach either low or large hydroelectric dispatch. The capacity of the reservoirs does not limit the dispatch when the amount of water in the system is high.

- The variation between futures and spot prices it is also relevant as it represents the opportunity cost of hydro reservoirs agents. This variation that has been computed as the percentage difference between the spot price and the future price, has more or less effect in the dispatch depending on the moment of the year. During periods with low hydro availability, the correlation of this variation of prices just mention and the generation is higher.

- Pumped hydro it is mainly drive by the opportunity cost that the difference of prices during the day (peak and valley hours) creates. The correlation of the spread between peak and valley hours and the schedule generation of pumped hydro plants is 0.440. Moreover, it has been found than when the spread its lower or equal than 30 Euro/MWh they do not schedule more than 500 MWh. This fact means that the cost of producing more than 500 MWh may not be compensated with a spread of 30 Euro/MWh.

- Wind generation has in general, an effect of price reduction larger than solar photovoltaic. In the period that has been study (2018, 2019, 2020) it was determine that during the hours with a dispatch lower than 1700 MWh there is not an effect of price reduction. This is more relevant during summer, as there are more hours with lower wind generation. On the other hand, with the expected rise in the wind install capacity, these hours with higher generation will increased, therefore, the correlation between wind production in the system and low prices will be stronger.

- For the case of the two solar technologies, it has been found than only solar photovoltaic is able to have a significant effect in the electricity prices. We have clearly seen that the last years solar installed capacity increased, has made solar generation to reach higher values of generation share in the system. This change has made this technology to have more relation with hours with lower prices.

- The imports and export with France of the day before can explain the spread of prices between Spain and France half of the hours of the data that has been used. The correlation between the difference of prices between both countries and the electricity exchanges of the day before have a correlation of - 0.598 in the case of exports, and - 0.468 in the case of imports.

- In the multiple linear regression model, we have found that the most significant variables for the price adjustment of each quarter, are the same as the ones that have been found in the correlation analysis. The R-square values are very near one. This means that the variables that have been used can predict the price dynamics of the years 2018, 2019 and 2020 in a very accurate way.

- The results of the differentiation between hours with lower and larger solar photovoltaic generation than 3000 MWh are better than in the case of the model without this distinction. This happens when using the data that have been selected to compute the equation. However, when we use it with external data the errors that we have obtained are quite large. More studies will be needed in order to try these equations in months with higher solar generation and in a more stable situation of market prices.

- The results that we have obtained from the simplification of the equations where we have end up in a very reduce set of variables are in general very good, except for the first quarter. A deeper study will be needed to improve the adjustment of the first quarter equation. On the other hand, we have obtained the best result in the third quarter, where the model was simplified to MIBGAS and carbon prices. The reason can be that MIBGAS is one of the most relevant price setting technologies in this period of the year. Firstly, because hydro reservoir will produce based on the opportunity cost that the variation between future and spot price create. Also, the lower wind generation will make renewable to set the price less hours and finally, combine cycle power will be setting the price a significant number of hours.

- Finally, when trying the model with external data of the year 2021, we have found good results in the prediction of January. In the case of the April, the high carbon prices of the period and the relevant weight that this variable has in the formula results in large errors.

- This study shows the high sensitivity that electricity prices have. The significant changes in the markets, like the rise in the carbon emission allowances price, is an example of the fluctuation that the electricity system may undergo in the future. Deeper and permanent studies will be very necessary in order to be ready to face this volatile environment.

References

1. Omie.es. 2021. *Electricity market | OMIE*. [online] Available at: https://www.omie.es/en/mercado-de-electricidad.

2. IEA. 2021. *European Union 2020 – Analysis - IEA*. [online] Available at: https://www.iea.org/reports/european-union-2020.
3. Mosquera-López, S. and Nursimulu, A., 2019. Drivers of electricity price dynamics: Comparative analysis of spot and futures markets. *Energy Policy*, 126, pp.76-87.
4. Figueiredo, N. C. and Silva, P. P., 2018. "The price of wind power generation in Iberia and the merit-order effect", *International Journal of Sustainable Energy Planning and Management*, 15, pp. 21–30.

5. España, Ley Orgánica 21/2013, de 9 de diciembre, de evaluación ambiental. *Boletín Oficial del Estado*, 11 de enero de 2021, núm. 9, p. 2681 a 2741. (https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-421)
6. Weron, R. and Misiorek, A., 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24(4), pp.744-763.
7. Gianfreda, A. and Grossi, L., 2012. Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, 34(6), pp.2228-2239.
8. Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), pp.1030-1081.
9. Weher, E., 1977. Edwards, Allen, L.: An introduction to linear regression and correlation. (A series of books in psychology.) W. H. Freeman and Comp., San Francisco 1976. 213 S., Tafelanh., s 7.00. *Biometrical Journal*, 19(1), pp.83-84.
10. Ferreira, Â., Ramos, J. and Fernandes, P., 2019. A linear regression pattern for electricity price forecasting in the Iberian electricity market. *Revista Facultad de Ingeniería Universidad de Antioquia*, (93), pp.117-127.
11. Frondel, M., Schmidt, C. and Vance, C., 2012. Emissions trading: Impact on electricity prices and energy-intensive industries. *Intereconomics*, 47(2), pp.104-111.
12. Climate Action - European Commission. 2021. *EU Emissions Trading System (EU ETS) - Climate Action - European Commission*. [online] Available at: https://ec.europa.eu/clima/policies/ets_en.

13. Tan, X. and Wang, X., 2017. Dependence changes between the carbon price and its fundamentals: A quantile regression approach. *Applied Energy*, 190, pp.306-325.
14. Ree.es. 2021. *Demand for electricity in Spain falls by 0.4% in November | Red Eléctrica de España*. [online] Available at: https://www.ree.es/en/press-office/news/press-releases/2019/12/demand-electricity-spain-falls-04-november.
15. Elperiodicodelaenergia.com. 2021. *El precio del CO2 se dispara ya por encima de los 50 euros por tonelada: Ribera, preocupada, anuncia medidas para paliar este coste de transición*. [online] Available at: <https://elperiodicodelaenergia.com/el-precio-del-co2-se-dispara-ya-por-encima-de-los-50-euros-por-tonelada-ribera-preocupada-anuncia-medidas-para-paliar-este-coste-de-transicion/> [Accessed 15 July 2021].
16. Omie.es. 2021. [online] Available at: https://www.omie.es/sites/default/files/2021-01/informe_anual_2020_es.pdf.
17. Miteco.gob.es. 2021. *Plan Nacional Integrado de Energía y Clima (PNIEC) 2021-2030*. [online] Available at: https://www.miteco.gob.es/es/prensa/pniec.aspx
18. Morris, C., 2021. *German power prices negative over weekend | Energy Transition*. [online] Energy Transition. Available at: https://energytransition.org/2014/05/german-power-prices-negative-over-weekend/.
19. Lledó, L., Torralba, V., Soret, A., Ramon, J. and Doblas-Reyes, F., 2019. Seasonal forecasts of wind power generation. *Renewable Energy*, 143, pp.91-100.
20. Roldan-Fernandez, J., Gómez-Quiles, C., Merre, A., Burgos-Payán, M. and Riquelme-Santos, J., 2018. Cross-Border Energy Exchange and Renewable Premiums: The Case of the Iberian System. *Energies*, 11(12), p.3277.

21. Ree.es. 2021. *Strengthening interconnections | Red Eléctrica de España*. [online] Available at: https://www.ree.es/en/red21/strengthening-interconnections.

22. Elperiodicodelaenergia.com. 2021. *Un 'pool' de enero para la historia: la extrema volatilidad del mercado eléctrico hace ver precios récord máximos y mínimos*. [online] Available at: <https://elperiodicodelaenergia.com/un-pool-de-enero-para-la-historia-la-extrema-volatilidad-del-mercado-electrico-hace-ver-precios-record-maximos-y-minimos/>.

Annex

Capacity index values

Table 26. Capacity index calculation.

Month	Week	Average Capacity (GW)	Max (GW)	Min (GW)	Capacity 2019	Capacity 2020	INDEX 2019	INDEX 2020
January	1	10294,7647	14114	6661	10175	12485	0,7209154	0,88458268
	2	10578,9412	15142	7111	10022	12656	0,66186765	0,8358209
	3	10762,9412	15581	7248	9878	12576	0,63397728	0,8071369
	4	10957,7059	15367	7436	9671	12665	0,62933559	0,82416867
	5	11109,3529	15409	7556	9897	13129	0,64228698	0,85203453
February	6	11265	16033	7528	10582	13177	0,66001372	0,8218674
	7	11428,5882	16744	7490	10927	13108	0,65259197	0,78284759
	8	11671,3529	16706	7639	11094	12975	0,66407279	0,77666707
	9	11916,8235	16766	7653	11124	12840	0,66348563	0,76583562
March	10	12082,1765	17063	7726	11036	13388	0,64677958	0,7846217
	11	12266,7647	16799	7787	11351	13461	0,67569498	0,8012977
	12	12466,1765	16538	7865	11337	13814	0,68551215	0,83528843
	13	12679,2353	16318	7929	11289	13724	0,69181272	0,84103444
April	14	12953	16905	8169	11182	14053	0,66146111	0,83129252
	15	13144,2941	17167	8464	11012	14367	0,64146327	0,83689637
	16	13341,4706	17153	9169	11116	15046	0,6480499	0,87716434
	17	13552,6471	16941	9879	11339	15407	0,66932294	0,90945045
May	18	13677,6471	16896	9979	11849	15845	0,70129025	0,93779593
	19	13769,1176	16980	9863	12223	15865	0,71984688	0,93433451
	20	13860,0588	17004	10028	12332	16041	0,72524112	0,94336627
	21	13872	17122	10070	12424	16129	0,72561617	0,94200444
June	22	13861,2353	17123	9994	12435	16056	0,7262162	0,93768615
	23	13780,1176	17062	9873	12390	15852	0,72617513	0,92908217
	24	13739,8235	16945	9750	12300	15834	0,72587784	0,93443494
	25	13611,3529	16949	9543	12189	15639	0,71915747	0,9227093
July	26	13385,2353	16858	9207	12035	15364	0,71390438	0,91137739
	27	13092,7059	16715	8958	11807	15041	0,70637152	0,89985043
	28	12791,5294	16414	8718	11487	14753	0,69982941	0,8988059
	29	12441,2353	16061	8382	11228	14464	0,69908474	0,90056659
August	30	12065,7647	15773	8045	10848	14060	0,68775756	0,89139669
	31	11689,6471	15407	7789	10442	13492	0,67774388	0,87570585
	32	11299,7647	15059	7459	10117	12995	0,67182416	0,86293911
	33	10971,4118	14661	7372	9815	12674	0,6694632	0,86447036
September	34	10669,5882	14344	7077	9513	12175	0,66320413	0,84878695
	35	10332,8235	13918	6916	9258	11726	0,66518178	0,84250611
	36	10061,4118	13540	6666	8944	11362	0,6605613	0,83914328

	37	9790,41176	13139	6548	8646	10964	0,65804095	0,83446229
October	38	9584,47059	12851	6468	8388	10610	0,65271185	0,82561668
	39	9374,29412	12499	6215	8106	10355	0,64853188	0,82846628
	40	9248,94118	12188	6238	7923	10317	0,65006564	0,84648835
	41	9148,76471	12164	6142	7730	10248	0,63548175	0,84248602
November	42	9073,35294	12161	6193	7574	10039	0,62281062	0,82550777
	43	9077,76471	12139	6266	7631	10458	0,62863498	0,86152072
	44	9222,23529	12191	6206	7840	10558	0,64309737	0,86604872
	45	9343,82353	12259	6159	7905	10698	0,64483237	0,87266498
December	46	9491,05882	12677	6157	8316	10717	0,65599117	0,84538929
	47	9574,52941	12955	6103	8856	10643	0,68359707	0,82153609
	48	9728,29412	13251	5997	9285	10453	0,70070183	0,78884612
	49	9862,17647	13484	5890	10121	10383	0,7505933	0,77002373
	50	9979,76471	13794	5923	10220	10939	0,74090184	0,79302595
	51	9991,94118	13901	6267	10449	11383	0,75167254	0,81886195
	52	10120,1176	14080	6401	12314	11656	0,87457386	0,82784091

R Studio codes

- First quarter results

```
Call:
lm(formula = Price ~ ForSolar + ForWind + Hueco + M + Q + Capacity +
    Imp + Exp + Prevdemand + GAS + CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-34.541  -3.069   0.185   3.681  30.142

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.230e+01  2.241e+00  -9.953  < 2e-16 ***
ForSolar    -3.211e-03  2.294e-04 -13.994  < 2e-16 ***
ForWind     -3.757e-03  2.225e-04 -16.886  < 2e-16 ***
Hueco       -2.295e-03  2.257e-04 -10.169  < 2e-16 ***
M            3.352e-01  2.562e-02  13.086  < 2e-16 ***
Q            1.438e-01  4.588e-02   3.133  0.00174 **
Capacity    -2.121e+01  1.696e+00 -12.502  < 2e-16 ***
Imp         -9.527e-04  9.239e-05 -10.312  < 2e-16 ***
Exp         -9.403e-04  1.346e-04  -6.984 3.17e-12 ***
Prevdemand   3.571e-03  2.245e-04  15.910  < 2e-16 ***
GAS          6.620e-01  4.215e-02  15.706  < 2e-16 ***
CO2          4.432e-01  3.271e-02  13.550  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.452 on 6010 degrees of freedom
Multiple R-squared:  0.7717,    Adjusted R-squared:  0.7713
F-statistic:  1847 on 11 and 6010 DF,  p-value: < 2.2e-16
```

- First quarter results for solar PV larger than 3000 MWh

```
Call:
lm(formula = Price ~ ForSolar + ForWind + M + Hueco + Capacity +
    GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max
-28.3878  -2.2072   0.1032   2.5932  21.1490

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.674e+00  2.937e+00  -1.591  0.11177
ForSolar    -3.910e-04  1.489e-04  -2.626  0.00875 **
ForWind     -1.471e-03  3.888e-05 -37.833  < 2e-16 ***
M            4.253e-01  4.219e-02  10.079  < 2e-16 ***
Hueco       -2.580e-04  2.695e-05  -9.572  < 2e-16 ***
Capacity    -1.479e+01  2.751e+00  -5.377 9.03e-08 ***
GAS          7.646e-01  6.357e-02  12.027  < 2e-16 ***
Prevdemand   1.512e-03  5.201e-05  29.070  < 2e-16 ***
Imp          1.243e-04  1.562e-04   0.796  0.42645
Exp          1.215e-05  2.300e-04   0.053  0.95786
CO2         -1.873e-01  6.956e-02  -2.693  0.00718 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.826 on 1232 degrees of freedom
Multiple R-squared:  0.8786,    Adjusted R-squared:  0.8776
F-statistic: 891.4 on 10 and 1232 DF,  p-value: < 2.2e-16
```

- First quarter results for solar PV lower than 3000 MWh**:**

```
Call:
lm(formula = Price ~ ForSolar + ForWind + M + Hueco + M + Capacity +
    GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-33.603  -3.168   0.116   3.710  29.628

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.573e+00  1.497e+00  -1.051    0.293
ForSolar    -9.221e-04  1.233e-04  -7.481 8.62e-14 ***
ForWind     -1.573e-03  2.638e-05 -59.625  < 2e-16 ***
M            3.387e-01  1.903e-02  17.795  < 2e-16 ***
Hueco       -3.603e-04  2.085e-05 -17.279  < 2e-16 ***
Capacity    -1.555e+01  1.838e+00  -8.460  < 2e-16 ***
GAS          8.028e-01  2.986e-02  26.887  < 2e-16 ***
Prevdemand   1.649e-03  2.981e-05  55.321  < 2e-16 ***
Imp         -8.280e-04  9.511e-05  -8.705  < 2e-16 ***
Exp         -1.118e-03  1.284e-04  -8.705  < 2e-16 ***
CO2          3.196e-02  4.160e-02   0.768    0.442
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.478 on 5199 degrees of freedom
Multiple R-squared:  0.7887,    Adjusted R-squared:  0.7883
F-statistic:  1941 on 10 and 5199 DF,  p-value: < 2.2e-16
```

- Second quarter results

```
Call:
lm(formula = Price ~ ForSolar + ForWind + M + Hueco + M + Y +
    Capacity + GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-33.916  -2.843   0.071   3.148  16.492

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.950e+01  2.531e+00 -15.607  < 2e-16 ***
ForSolar     1.379e-03  1.125e-04  12.249  < 2e-16 ***
ForWind      6.227e-04  1.137e-04   5.478 4.49e-08 ***
M            2.557e-03  1.191e-04  21.471  < 2e-16 ***
Hueco        2.371e-03  1.085e-04  21.844  < 2e-16 ***
Y            5.728e-01  4.815e-02  11.896  < 2e-16 ***
Capacity     2.144e-01  4.519e-02   4.744 2.15e-06 ***
GAS          1.760e+00  2.234e-02  78.786  < 2e-16 ***
Prevdemand  -1.323e-03  1.086e-04 -12.180  < 2e-16 ***
Imp         -4.903e-04  8.260e-05  -5.936 3.09e-09 ***
Exp         -7.835e-04  1.007e-04  -7.781 8.48e-15 ***
CO2          1.105e+00  2.914e-02  37.917  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.227 on 5820 degrees of freedom
Multiple R-squared:  0.8645,    Adjusted R-squared:  0.8642
F-statistic:  3376 on 11 and 5820 DF,  p-value: < 2.2e-16
```

- Second quarter results for solar PV larger than 3000 MWh

```
Call:
lm(formula = Price ~ ForSolar + ForWind + M + Hueco + M + Y +
    Capacity + GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max
-31.0174  -2.0735   0.0954   2.3587  13.2043

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.922e+01  3.982e+00  -7.339 3.18e-13 ***
ForSolar     1.140e-03  1.854e-04   6.152 9.30e-10 ***
ForWind      5.783e-04  1.717e-04   3.367 0.000774 ***
M            2.367e-03  1.766e-04  13.408  < 2e-16 ***
Hueco        2.090e-03  1.673e-04  12.487  < 2e-16 ***
Y            3.777e-01  6.924e-02   5.456 5.51e-08 ***
Capacity    -2.712e-03  6.370e-02  -0.043 0.966042
GAS          1.708e+00  3.924e-02  43.524  < 2e-16 ***
Prevdemand  -1.023e-03  1.663e-04  -6.154 9.21e-10 ***
Imp         -6.707e-05  1.480e-04  -0.453 0.650449
Exp         -1.761e-04  1.492e-04  -1.180 0.238160
CO2          9.644e-01  4.271e-02  22.582  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.534 on 1902 degrees of freedom
Multiple R-squared:  0.9092,    Adjusted R-squared:  0.9087
F-statistic:  1732 on 11 and 1902 DF,  p-value: < 2.2e-16
```

- Second quarter results for PV lower than 3000 MWh

```
Call:
lm(formula = Price ~ ForSolar + ForWind + M + Hueco + M + Y +
    Capacity + GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-33.345  -3.184   0.085   3.439  17.089

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.288e+01  3.321e+00 -12.909  < 2e-16 ***
ForSolar     1.575e-03  1.858e-04   8.480  < 2e-16 ***
ForWind      5.529e-04  1.465e-04   3.774 0.000163 ***
M            2.573e-03  1.583e-04  16.252  < 2e-16 ***
Hueco        2.431e-03  1.386e-04  17.541  < 2e-16 ***
Y            6.185e-01  6.418e-02   9.637  < 2e-16 ***
Capacity     2.893e-01  6.210e-02   4.658 3.30e-06 ***
GAS          1.787e+00  2.834e-02  63.054  < 2e-16 ***
Prevdemand  -1.376e-03  1.392e-04  -9.883  < 2e-16 ***
Imp         -6.901e-04  1.014e-04  -6.805 1.16e-11 ***
Exp         -1.199e-03  1.397e-04  -8.581  < 2e-16 ***
CO2          1.179e+00  3.872e-02  30.449  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.455 on 3906 degrees of freedom
Multiple R-squared:  0.8408,    Adjusted R-squared:  0.8403
F-statistic:  1875 on 11 and 3906 DF,  p-value: < 2.2e-16
```

- Third quarter results

```
Call:
lm(formula = Price ~ ForSolar + ForWind + Hueco + GAS + Prevdemand +
    M + Imp + Exp + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max
-13.9252  -2.4098  -0.0693   2.3661  13.3690

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.558e+01  1.177e+00 -13.239  < 2e-16 ***
ForSolar    -1.477e-03  1.837e-04  -8.040 1.05e-15 ***
ForWind     -1.584e-03  1.832e-04  -8.647  < 2e-16 ***
Hueco       -5.459e-04  1.826e-04  -2.990  0.00280 **
GAS          1.367e+00  3.511e-02  38.947  < 2e-16 ***
Prevdemand   1.581e-03  1.831e-04   8.636  < 2e-16 ***
M            1.305e-01  2.028e-02   6.435 1.32e-10 ***
Imp         -7.483e-04  5.733e-05 -13.051  < 2e-16 ***
Exp         -1.690e-04  6.483e-05  -2.607  0.00916 **
CO2          4.326e-01  2.370e-02  18.251  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.711 on 6614 degrees of freedom
Multiple R-squared:  0.9274,    Adjusted R-squared:  0.9273
F-statistic:  9389 on 9 and 6614 DF,  p-value: < 2.2e-16
```

- Third quarter results for PV larger than 3000 MWh

```
Call:
lm(formula = Price ~ ForSolar + ForWind + Hueco + GAS + Prevdemand +
    M + Imp + Exp + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max
-11.4183  -1.9689  -0.0022   1.9442  11.7232

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.301e+01  1.916e+00  -6.791 1.38e-11 ***
ForSolar    -1.716e-03  2.655e-04  -6.464 1.22e-10 ***
ForWind     -1.903e-03  2.587e-04  -7.353 2.60e-13 ***
Hueco       -8.034e-04  2.590e-04  -3.102  0.00195 **
GAS          1.497e+00  5.313e-02  28.177  < 2e-16 ***
Prevdemand   1.702e-03  2.605e-04   6.532 7.81e-11 ***
M            6.945e-02  3.192e-02   2.176  0.02967 *
Imp         -4.821e-04  9.058e-05  -5.322 1.12e-07 ***
Exp          3.060e-05  9.537e-05   0.321  0.74837
CO2          4.635e-01  3.427e-02  13.527  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.212 on 2523 degrees of freedom
Multiple R-squared:  0.9464,	Adjusted R-squared:  0.9463
F-statistic:  4954 on 9 and 2523 DF,  p-value: < 2.2e-16
```

- Third quarter results for PV lower than 3000 MWh

```
Call:
lm(formula = Price ~ ForSolar + ForWind + Hueco + GAS + Prevdemand +
    M + Imp + Exp + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max
-13.2623  -2.7128  -0.0791   2.6051  13.6767

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.576e+01  1.568e+00 -10.051  < 2e-16 ***
ForSolar    -1.228e-03  2.629e-04  -4.671 3.09e-06 ***
ForWind     -1.227e-03  2.499e-04  -4.910 9.47e-07 ***
Hueco       -2.052e-04  2.486e-04  -0.826    0.409
GAS          1.278e+00  4.561e-02  28.031  < 2e-16 ***
Prevdemand   1.323e-03  2.484e-04   5.326 1.06e-07 ***
M            1.718e-01  2.644e-02   6.501 8.96e-11 ***
Imp         -8.822e-04  7.509e-05 -11.749  < 2e-16 ***
Exp         -3.862e-04  9.560e-05  -4.040 5.43e-05 ***
CO2          4.237e-01  3.202e-02  13.231  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.959 on 4081 degrees of freedom
Multiple R-squared:  0.9167,	Adjusted R-squared:  0.9165
F-statistic:  4990 on 9 and 4081 DF,  p-value: < 2.2e-16
```

- Fourth quarter results:

```
Residuals:
     Min       1Q   Median       3Q      Max 
-28.6949  -3.2935   0.2138   3.5462  23.6655 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.822e+00  3.794e+00  -0.744    0.457    
ForWind     -3.981e-03  1.167e-04 -34.128  < 2e-16 ***
Y            5.348e-01  3.054e-02  17.509  < 2e-16 ***
M           -4.278e-01  2.882e-02 -14.845  < 2e-16 ***
Q            2.072e-01  4.257e-02   4.869 1.15e-06 ***
Capacity    -4.475e+01  2.113e+00 -21.172  < 2e-16 ***
Imp         -1.065e-03  7.874e-05 -13.529  < 2e-16 ***
Exp         -4.591e-04  9.185e-05  -4.999 5.92e-07 ***
ForSolar    -3.621e-03  1.302e-04 -27.803  < 2e-16 ***
Hueco       -2.630e-03  1.186e-04 -22.164  < 2e-16 ***
GAS          1.745e+00  4.125e-02  42.307  < 2e-16 ***
Prevdemand   4.031e-03  1.211e-04  33.276  < 2e-16 ***
CO2         -2.624e-01  3.593e-02  -7.302 3.18e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.714 on 6611 degrees of freedom
  (23 observations deleted due to missingness)
Multiple R-squared:  0.8589,	Adjusted R-squared:  0.8586 
F-statistic:  3353 on 12 and 6611 DF,  p-value: < 2.2e-16
```

- Fourth quarter results for PV larger than 3000 MWh:

```
Call:
lm(formula = Price ~ ForWind + Imp + Exp + Y + Q + M + Capacity + 
    ForSolar + Hueco + GAS + Prevdemand + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max 
-23.9221  -2.6939   0.2696   2.7528  19.1577 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.692e+01  9.452e+00   3.906 9.98e-05 ***
ForWind     -2.385e-03  3.198e-04  -7.458 1.88e-13 ***
Imp         -2.751e-04  2.020e-04  -1.362 0.173528    
Exp         -4.194e-04  2.227e-04  -1.884 0.059895 .  
Y            5.670e-01  8.722e-02   6.501 1.25e-10 ***
Q           -8.940e-01  1.012e-01  -8.833  < 2e-16 ***
M            2.323e-01  6.293e-02   3.691 0.000235 ***
Capacity    -6.693e+01  4.763e+00 -14.051  < 2e-16 ***
ForSolar    -2.003e-03  3.727e-04  -5.375 9.49e-08 ***
Hueco       -1.122e-03  3.317e-04  -3.384 0.000742 ***
GAS          1.249e+00  8.839e-02  14.136  < 2e-16 ***
Prevdemand   2.512e-03  3.356e-04   7.485 1.54e-13 ***
CO2          1.984e-01  1.036e-01   1.915 0.055717 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.909 on 1022 degrees of freedom
Multiple R-squared:  0.8762,	Adjusted R-squared:  0.8748 
F-statistic:   603 on 12 and 1022 DF,  p-value: < 2.2e-16
```

- Fourth quarter results for PV lower than 3000 MWh:

```
Call:
lm(formula = Price ~ ForWind + ForSolar + M + Hueco + M + Y +
    Q + Capacity + GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
     Min       1Q   Median       3Q      Max
-27.9256  -3.4224   0.1843   3.5332  24.8627

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.910e+00  4.135e+00  -1.429    0.153
ForWind     -3.945e-03  1.258e-04 -31.357  < 2e-16 ***
ForSolar    -3.394e-03  1.663e-04 -20.412  < 2e-16 ***
M           -4.549e-01  3.159e-02 -14.401  < 2e-16 ***
Hueco       -2.641e-03  1.278e-04 -20.668  < 2e-16 ***
Y            5.085e-01  3.366e-02  15.109  < 2e-16 ***
Q            3.063e-01  4.624e-02   6.624 3.83e-11 ***
Capacity    -4.242e+01  2.325e+00 -18.244  < 2e-16 ***
GAS          1.720e+00  4.479e-02  38.402  < 2e-16 ***
Prevdemand   4.091e-03  1.303e-04  31.391  < 2e-16 ***
Imp         -1.192e-03  8.951e-05 -13.318  < 2e-16 ***
Exp         -1.476e-03  1.001e-04 -14.735  < 2e-16 ***
CO2         -3.146e-01  3.799e-02  -8.281  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.762 on 5576 degrees of freedom
Multiple R-squared:  0.8596,    Adjusted R-squared:  0.8593
F-statistic:  2844 on 12 and 5576 DF,  p-value: < 2.2e-16
```

- Result for the three years and months at the same time

```
Call:
lm(formula = Price ~ ForWind + ForSolar + Hueco + GAS + Prevdemand +
    Imp + Exp + CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-35.783  -3.306   0.189   3.818  24.196

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) -9.241e+00  3.912e-01  -23.621  < 2e-16 ***
ForWind     -1.724e-03  1.309e-05 -131.691  < 2e-16 ***
ForSolar    -1.240e-03  2.337e-05  -53.073  < 2e-16 ***
Hueco       -2.212e-04  1.004e-05  -22.040  < 2e-16 ***
GAS          1.471e+00  7.829e-03  187.850  < 2e-16 ***
Prevdemand   1.455e-03  1.504e-05   96.716  < 2e-16 ***
Imp         -2.391e-04  4.076e-05   -5.866 4.53e-09 ***
Exp         -8.226e-04  5.874e-05  -14.004  < 2e-16 ***
CO2          2.950e-01  1.145e-02   25.774  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.319 on 24780 degrees of freedom
Multiple R-squared:  0.7936,    Adjusted R-squared:  0.7935
F-statistic: 1.191e+04 on 8 and 24780 DF,  p-value: < 2.2e-16
```

- Result for the three years and months at the same time with futures and capacity included

```
Call:
lm(formula = Price ~ ForWind + Y + M + Q + Capacity + ForSolar +
    Hueco + GAS + Prevdemand + Imp + Exp + CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-33.542  -3.116   0.198   3.486  25.328

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) -3.712e+01  5.857e-01  -63.382  < 2e-16 ***
ForWind     -1.563e-03  1.306e-05 -119.663  < 2e-16 ***
Y            4.226e-01  8.339e-03   50.670  < 2e-16 ***
M            2.559e-01  6.194e-03   41.314  < 2e-16 ***
Q            2.783e-01  9.357e-03   29.742  < 2e-16 ***
Capacity     2.580e-01  8.033e-03   32.115  < 2e-16 ***
ForSolar    -1.121e-03  2.185e-05  -51.292  < 2e-16 ***
Hueco       -1.168e-04  9.732e-06  -12.006  < 2e-16 ***
GAS          7.305e-01  1.314e-02   55.594  < 2e-16 ***
Prevdemand   1.391e-03  1.401e-05   99.293  < 2e-16 ***
Imp         -1.083e-03  4.028e-05  -26.875  < 2e-16 ***
Exp         -1.396e-03  5.528e-05  -25.247  < 2e-16 ***
CO2         -7.983e-02  1.220e-02   -6.542  6.2e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.801 on 24776 degrees of freedom
Multiple R-squared:  0.8261,    Adjusted R-squared:  0.826
F-statistic:  9808 on 12 and 24776 DF,  p-value: < 2.2e-16
```

- Result for the three years and months at the same time without capacity, futures and imports and export of the day before.

```
Call:
lm(formula = Price ~ ForWind + ForSolar + Hueco + GAS + Prevdemand +
    CO2, data = Prediccion1)

Residuals:
    Min      1Q  Median      3Q     Max
-35.435  -3.329   0.239   3.856  25.014

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.480e+00  3.897e-01  -24.33   <2e-16 ***
ForWind     -1.729e-03  1.303e-05 -132.68   <2e-16 ***
ForSolar    -1.200e-03  2.327e-05  -51.57   <2e-16 ***
Hueco       -1.975e-04  9.921e-06  -19.91   <2e-16 ***
GAS          1.468e+00  7.750e-03  189.43   <2e-16 ***
Prevdemand   1.402e-03  1.462e-05   95.93   <2e-16 ***
CO2          3.329e-01  1.113e-02   29.91   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.344 on 24782 degrees of freedom
Multiple R-squared:  0.792,     Adjusted R-squared:  0.7919
F-statistic: 1.573e+04 on 6 and 24782 DF,  p-value: < 2.2e-16
```