POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Prediction and Interpretation of bids in the Italian Electricity Market for Ancillary Services

LAUREA MAGISTRALE IN INGEGNERIA MATEMATICA - MATHEMATICAL ENGINEERING

Author: **ANDREA PASOTTI | 10493242**
Advisor: **PROF. PIERCESARE SECCHI**
Co-advisor: **ENG. FILIPPO BOVERA**
Academic year: **2020-2021**

---

1. Introduction

This Executive Summary introduces the results of a joint collaboration between **Politecnico di Milano** and the research institute **RSE - *Ricerca sul Sistema Energetico***, based on an extensive statistical analysis of the **Italian Electricity Markets**. These marketplaces have experienced drastic changes in recent years, mainly driven by the general desire to transition towards a more **sustainable economy**, supported by the diffusion of power plants fueled by non-programmable Renewable Energy Sources. Indeed, these units are able to produce much cheaper and cleaner energy than traditional thermoelectric power plants but are causing a sharp **increase** in the measures required for the **regulation of the Power System**, with balancing resources that have to be procured by the Italian Transmission System Operator on the **Ancillary Services Market (MSD)**. MSD is part of the so-called Spot market for electricity (the one managing short-term energy tradings), it takes place after the Day-Ahead and Intra-Day markets, and covers *ancillary services*, which is a term used to denote special services and functions adopted by the TSO to meet the demand for electrical energy in real time. Indeed, the market is used by the TSO to anticipate every possible imbalance that may happen to the Power System, and thus it has a very specific nature, being **restricted** to only a **selected subgroup** of Production Units, that have at least 10 MW in installed capacity and meet other specific safety prerequisites. In contrast to the previous markets, **MSD is not based on a merit order** and therefore it is generally characterized by higher prices than the Day-Ahead Market (MGP), where, however, the greatest portion of electrical energy is exchanged. Furthermore, the **participation** to the market **is compulsory**, meaning that all the units that are eligible for it have to bid at every hour. Eventually, the **unpredictability** behind the need for ancillary services generates a setting that is both quite complicated and intriguing to study, from a statistical standpoint.

1.1. Overview of the conducted work and prior literature

The scope of my work is to present an **initial overview** of the **context** of electricity markets and the **Power System**, subsequently moving to the **collection** of all the **relevant data** to analyse the Production Units that operate in

electricity markets, with a special attention to the aforementioned Market for Ancillary Services. I then move on to the **selection** of a set of **117 Production Units** operating on MSD, studying their *macro-behaviour* across the electricity markets MGP, MI and MSD.
Eventually, I decided to restrict the domain of the analyses to the three-year period 2017 to 2019, **focusing on Northern Italy**, where the greatest portion of energy exchanges happen. The principal target variable of my work will be the specific service denoted as *"Gradino 1"* (**GR1**) on MSD ex-ante, which translates to "First Step" and it is associated to a request from the Production Unit to increase (OFF type) or decrease (BID type) the amount of energy to be produced.
The Italian market for ancillary services was previously analysed in [1], [2] and [3], each time with a specific focus on one single topic, such as the event denoted *opposite call* in [1] or the prediction of the acceptance of bids by the TSO Terna in [3]. This work has the purpose to built on the existing literature, providing the **first *wide-ranging* analysis** of the Italian electricity markets to our knowledge, posing a particular attention to the offer profiles submitted on MSD by the PUs operating in Northern Italy.

2. Data collection process

The first achievement of this work was the construction of a set of massive datasets, accounting for all the relevant features to our analyses. This was done with the active support of RSE and eng. Filippo Bovera, who helped me defining five macro-classes of variables. All the data are publicly available, and were mainly retrieved from [5] and [6].
The main areas of interest are addressed below:

- **Baseline Data**:
  A set of calendar variables and other useful features, such as the day and hour of each bid, the day of the week and a Boolean variable to flag each festivity.
- **Markets' Data**:
  Comprise all the bids submitted on the electricity markets MGP, MI and MSD in the years 2017 to 2019, across all Italian zones.
- **Production Units' Data**:
  A dataset containing information on 117 selected PUs, comprehending their technology, operator, installed capacity and voltage level of the connection to the Power Grid. The units were chosen from all the available ones, by filtering for those active in the bidding zone `NORD` during the three years under consideration.
- **Structural Data**:
  Data describing characteristics of the structure of the energy system, including the forecasted load and generation, the cumulative water reservoir in basins used by hydroelectric plants and the scheduled commercial exchanges, both with foreign countries and between Italian zones.
- **Exogenous Data**:
  I have grouped here data extraneous to the electricity markets, but possibly correlated to it. This group comprehends weather data (including average temperature in the day, wind speed and pressure), together with the spot price of natural gas on the respective Day-Ahead market.

3. Clustering the Production Units in Northern Italy

As anticipated in the introduction, this work has the primary scope to provide a **wide-ranging analysis** of the **Ancillary Services Market**, with the final aim being that of **interpreting** and **predicting** the bids on MSD GR1. To this extent, an initial **visual analysis** of the offer curves presented by the 117 PUs in our dataset highlighted major differences between the various power plants. Indeed, some units presented **extremely variable bids**, trying to adapt their offers to the price of electricity on the Day-Ahead Market, while other units submitted **constant bids for extended periods of time**. This quite evident difference in behaviour suggests that not all the units eligible to take part in MSD are equally interested to place competitive bids in it, at least at every hour. As a consequence, I decided to perform a **clustering analysis** of the units in the dataset, segmenting the power plants on the basis of their ***macro-behaviour*** across all the electricity markets.
To do that, I tracked the following variables:

- **Equivalent Operating Hours of Production**, in the years 2017 to 2019;
- **Total Net Income** on the three-year period, across the markets MGP, MI and

MSD. Data were weighted by dividing for the Installed Capacity of the PU, thus measured in €/MW;

- **Earnings from Day-Ahead Market (MGP)**, considering only OFF bids and dividing this value by the Installed Capacity of the PU [€/MW];
- **Intra-Day Market Total Expenses from BID bids**, divided by the Installed Capacity of the PU [€/MW];
- **Intra-Day Market Total Earnings from OFF bids**, divided by the Installed Capacity of the PU [€/MW];
- **Expenses from Ancillary Services Market - BID bids**, divided by the Installed Capacity of the PU [€/MW];
- **Earnings from Ancillary Services Market - OFF bids**, divided by the Installed Capacity of the PU [€/MW].

To perform the analysis, I decided to use the **algorithm K-Means**[1], which is one of the most established tools in unsupervised machine learning. This said, since the underlying setting is quite complex, it was necessary to **refine** and further interpret the results of the algorithm, by looking at the **offer profiles on MSD** of the units in each identified cluster, as well as their technology and operator.

After the refinement of the K-Means results, I managed to identify **eight** different **well-defined clusters** of Production Units, each characterized by a unique behaviour on the electricity markets and having units placing similar bids on MSD. The identified clusters are briefly presented below, posing particular attention to their strategies on the market MSD[2], which is the main purpose of this work.

3.1. Cluster of thermoelectric PUs from Enipower

It is a set of **nine CCGT power plants** from the operator **Enipower**, which have extremely **high values** for the **Equivalent Operating Hours** of Production, and report **great revenues** from MSD OFF bids. Analysing the offer curves from this cluster, I found that these PUs present bid profiles on MSD that are extremely close one another. Furthermore, the bids from this cluster have a really **high variability** and are updated almost on a hourly basis, **following** the trend of the zonal price of electricity on the **Day-Ahead Market** (MGP).

This makes this cluster one of the most interesting to be analysed in greater details, and I have fit with incredibly accurate results a Random Forest model to predict its MSD GR1 bids, of both upward and downward type.

3.2. Cluster of dispatchable hydroelectric PUs from ENEL

This group comprehends a set of **17 hydroelectric power plants** of *dispatchable* type, a production technology that indicates units operating with water basins and dams, excluding all the Pumped-Storage hydroelectric ones. These units belong to the operator ENEL, report **almost no revenues** from the **Ancillary Services Market** (MSD) and present quite low values of Equivalent Operating Hours of production. However, they appear to have a lot of accepted bids on the Intra-Day market (MI).

Looking more in detail at their offers on MSD, we notice that these units seem to be **not-interested** in taking part to this market, since they all submit constant BID offers at the value 0 €/MWh, for almost the entirety of our time span. Regarding OFF offers, they present more variability than BID ones, but still remain constant for many consecutive weeks.

Indeed, my analysis identified that the **operator ENEL** adopts the **same pricing strategy** across most of its units, with a common baseline value from which the units deviate only on the basis of some **auto-regressive variables** that are unit-specific, such as the programmed production after the market MI or the amount of water stored in their basin.

3.3. Cluster of "ENEL to Alperia" hydroelectric PUs

This cluster comprehends a set of 4 units that **switched management** from the operator **ENEL** to the operator **Alperia**, during the period of our analysis, precisely in March 2019.

---

[1] This was not the first considered approach, since initially we tried to perform the clustering directly on the offer curves of the Production Units, in a functional way. However, the results of this procedure were quite unsatisfactory, due to the difficulty to define a powerful notion of distance in the space of MSD GR1 bids.

[2] I refer to the relative part of the thesis to their complete analysis, as well as a graphical visualisation of their bids on MSD, necessary to fully appreciate the presented topics but omitted in this summary for space constraints.

All the units in this group are geographically located in *Val d'Ultimo*, in South Tyrol. Looking at the bids of these units on MSD, we can see a **profound difference** between those submitted under the management of ENEL, which are quite close to the ones presented for the previous cluster, and those under the management of **Alperia**. In particular, under Alperia **BID bids** are **more variable** and are on average at a higher price than those from ENEL, effectively implying **better price competitiveness**.

3.4. Cluster of thermoelectric PUs from ENEL

This cluster comprehends a group of 11 PUs that belong to a total of **4 power plants** controlled by ENEL (*La Casella*, *Fusina*, *Porto Corsini* and *La Spezia*), including all the thermoelectric PUs by ENEL in our dataset. All the selected units have in common a very **high activity** in the **Intra-Day Market**, especially for bids of BID type, in an analogue way to what was already observed for hydroelectric units of this operator. Regarding MSD bids, those of **BID type** present **less variability** than OFF ones, and are generally constant for many months. On the opposite, OFF bids seem to adapt to the seasonal oscillations of the MGP price, even if they are not updated on a hourly basis.

Furthermore, by looking more in depth at the offers from this cluster, I discovered how the operator ENEL places **MSD GR1 bids** that are usually **constant on a daily basis**, from midnight to 11 P.M. of almost every day, **regardless** the **hourly oscillation** of the zonal price on MGP or the forecasted load and generation. This behaviour is quite a surprising one and differs significantly from that of ENEL's competitors. It also **appears** to be somehow **"sub-optimal"**, due to the fact that at certain hours the upward bids on MSD result being quite close to the price of energy on the Day-Ahead Market. Furthermore, focusing only on **OFF bids**, I found that about 80% of them have **values that are multiple of 5**. This indicates with great likelihood a **human involvement in the process of bid definition**, with offer prices that are often rounded to the closest multiple of 5.

3.5. Cluster of thermoelectric PUs active on MSD-BID

This cluster comprehends six **thermoelectric plants of CCGT production technology**, belonging to various operators and including all the units that were most active in **MSD** for **BID types of offers**, according to the algorithm K-Means. The units selected by the clustering are those of *Leini*, *Livorno Ferraris*, *Torviscosa*, *Vado Ligure*, *Voghera* and *Tavazzano*, and reported the highest amount of money exchanged with the TSO Terna for MSD bids of BID type, in the period of our analysis. Geographically, the units are located in different parts of Northern Italy, but still present a **similar behaviour** on the electricity markets.

I decided to pose particular attention to this cluster, due to the peculiarity of MSD offers of BID type. Indeed, those offers are associated to a request made by the PU to "buy back" energy from Terna, and this is generally associated to the definition of **more complicated tactics** also on **prior markets**, that should take into account the possibility to reduce the working range on MSD. To address this issue, I fit a Random Forest model to predict and interpret downward bids on MSD, from the units in this cluster.

3.6. Cluster of thermoelectric PUs active on MSD-OFF

This cluster comprehends a set of six different Production Units (*Azotati*, *Cassano*, *Piacenza*, *Sermide*, *Turano Lodigiano* and *Turbigo*) that earned the greatest amount of money on MSD through bids of OFF type, according to the K-Means algorithm. All the units in this group are of **thermoelectric technology** and belong to various operators, with A2A being the most represented in the cluster.

In my analyses, I focused on the units from the **operator A2A** and conducted an in depth study of their offers on MSD. What I found is that these units present significant differences between OFF and BID types of bids, with the latter ones that are **mostly constant** throughout the whole time span of this analysis, fixed at a price that is about 33 €/MWh. This value is quite interesting for two reasons: on one side it implies a lack of interest from A2A to adjust its downward offers according to the MGP price. On the other side, though, the offered price is quite high (especially if compared to the value 0 €/MWh of hydroelectric units from ENEL), therefore these offers end up being very competitive in certain periods of the year, when the zonal price on MGP is low.

Regarding **upward MSD OFF bids** from A2A's thermoelectric units, these seem to present a **much greater variability than BID ones**. The variability, however, is given by sudden "jumps" to extremely high values, which are usually set to 500€/MWh, and thus indicate a general will not to be accepted on MSD. This said, looking more closely at the OFF bids, we can notice that they are generally characterized by a **constant baseline during the majority of the days**, to which the bids deviate only for the "extreme" offers at very high prices. Moreover, similarly to the operator ENEL, this baseline is usually updated at midnight.

3.7. Cluster of pumped-storage hydro PUs from ENEL

This cluster concerns a group of 8 **pumped-storage hydroelectric power plants from ENEL**, that present similar behaviour for MSD bids, especially those of BID type.

It was originally identified by the K-Means algorithm, which grouped together a set of 5 units with the same technology and operator, to which I then decided to add also the units of *Bargi Centrale*, *Entracque Rovina* and *Fadalto*, in order to include all the pumped-storage hydroelectric units from the operator ENEL in a single group. This decision was taken after a manual analysis of the offers on MSD from these PUs, which identified how both BID and OFF bids presented common traits, such as the same variability and similar price values. Furthermore, all the units present a very **peculiar behaviour** around the month of **October 2018**, when they all change the **minimal value** of their BID offers from zero to the value 35 €/MWh, once again confirming that the operator ENEL shares his knowledge across all his units of the same technology.

3.8. Cluster of hydroelectric PUs from CVA

The operator **CVA - *Compagnia Valdostana delle Acque*** is a minor company that produces **renewable energy** through hydroelectric power stations, in the localized region of *Valle d'Aosta*. This operator does not have a particular economic relevance and only owns 4 power plants eligible on MSD, however, I believe that it is quite interesting to conclude this analysis by studying the **behaviour of a smaller company**. A graphical visualisation of MSD bids, identified that CVA adopts very *naive* strategies, with offers that are slightly different among the PUs, but often **constant in both directions** for many weeks, or even months. This said, there seems to be a **seasonal pattern** in the offer curves on MSD from CVA's units, related to the presence of "hollows" in the bids of BID type. These happens in correspondence to the **summer months** and we could possibly interpret this fact by considering that all the units in the cluster are of hydroelectric type and located in the Alps. This fact implies that these four PUs will have **lots of primary resources** available when the summer heat melts the surrounding glaciers, increasing considerably the amount of water in the rivers of *Valle d'Aosta*, thus explaining the identified tendency.

4. Random Forest analysis of MSD GR1 bids

The last part of my work was dedicated to the **creation of prediction models** for some of the identified clusters. In particular, I chose to focus on the units from **Enipower**, as well the cluster of **thermoelectric PUs active with MSD bids of BID type**, since they are the most interesting ones to be analysed for the specific GR1 service on the Ancillary Services Market. The algorithm adopted for this analysis is called **Random Forests**, and was chosen because it combines great predictive performances and a high degree of interpretability. The algorithm was tuned mainly with the help of [4] and managed to achieve incredible accuracy results in all the models fitted.

The **target variable** is the price of the MSD GR1 offer for a specific hour in the time span 2017-2019. Regarding the predictors, they are chosen among the variables collected at the beginning of the analyses, and are all related to the same hour as the target variable. Every predictor is publicly **available** to all the operators of electricity markets **before** the actual observation of the **MSD GR1 prices**, to ensure meaningful results. Some variables are referred to the specific Production Unit that is placing the bid on MSD, such as the programmed production profile after MI (`MI_PROFILE`) and in that case they are always normalized by the installed capacity of the PU, in order to obtain comparable values among different units in the same cluster.

4.1. Cluster of Enipower, prediction of MSD bids of BID type

I will begin with the study of **BID offers on MSD GR1** from the cluster of thermoelectric units of Enipower. To analyse them, I fit a Random Forest Random Input model, with a total of 50 trees. For this model, the Out Of Bag cross-validation error is steadily decreasing after the first few iterations, eventually reaching the extremely low value of 0.51 $(€/\text{MWh})^2$. Furthermore, the **percentage of variability** explained by the model is more than acceptable, being equal to 99.31, meaning that the model is performing extremely well on the available data.

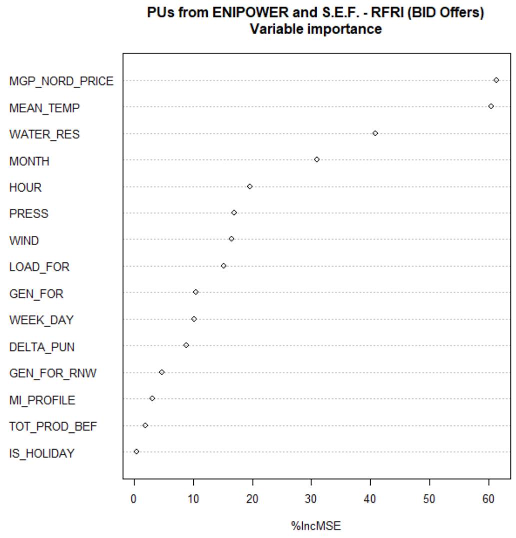


Figure 1: Cluster of Thermoelectric PUs from Enipower, Random Forest algorithm for MSD GR1 **BID Offers**: Variable Importance.

Moving on to the interpretation of the results of the algorithm, we can observe the **ranking of the importance** of the variables composing the model from Figure 1. This picture shows how the units within this cluster place their downward bids on MSD considering mainly the **price of electricity** on the **Day-Ahead Market**, adapting their bids by looking at many indicators of seasonality. Among those, we find the **Mean Temperature**, the variable indicating the **Water Reservoir** in artificial basins used by hydroelectric power plants, and the two variables for the **hour and month** of the considered bid. Moreover, it is also quite interesting to notice that the units in this cluster **do not seem to adapt their BID offers** according to their **percentage of work after MI**, since the variable `MI_PROFILE` has low importance. This is for sure a peculiarity of this cluster of units, and it is a direct consequence of the decision from the operator Enipower to place **almost identical bids** for all his units. This said, the units in the cluster are characterized by the **highest values for Equivalent Operating Hours**, meaning that they will almost always be active as a result of the markets MGP and MI.
Furthermore, it is interesting to notice how the two variables **Load Forecast** and **Generation Forecast** do not appear to be much useful in the definition of the final price of MSD GR1 bids of BID type.

Focusing on the two most important variables in the model, I have reported a visualization of their **marginal effects**, via Partial Dependence Plots, in Figure 2. From the graph on the left side of the picture, we can notice that the the **price of the bids tend to increase** with the **zonal price of electricity on MGP**, meaning that the operator Enipower tends to increase the prices of its BID offers, when the zonal price of electricity on the Day-Ahead Market grows. This is effectively a direct consequence of the fact that the operator tends to keep a **somewhat constant price delta** between the price of electricity on MGP and its downward bids, effectively encouraging the TSO Terna to accept its BID bids when the price of electricity is high. Moreover, from the plot reported on the rightmost part of Figure 2, we can see that the price of bids presented by Enipower seems to **decrease with the mean temperature**. This fact is particularly true if we consider the colder days in our time span, when we know that the price of electricity is generally higher, due to an expected increase in demand. Quite interestingly, though, we do not observe a similar pattern for the hottest days of the observed three-years period.

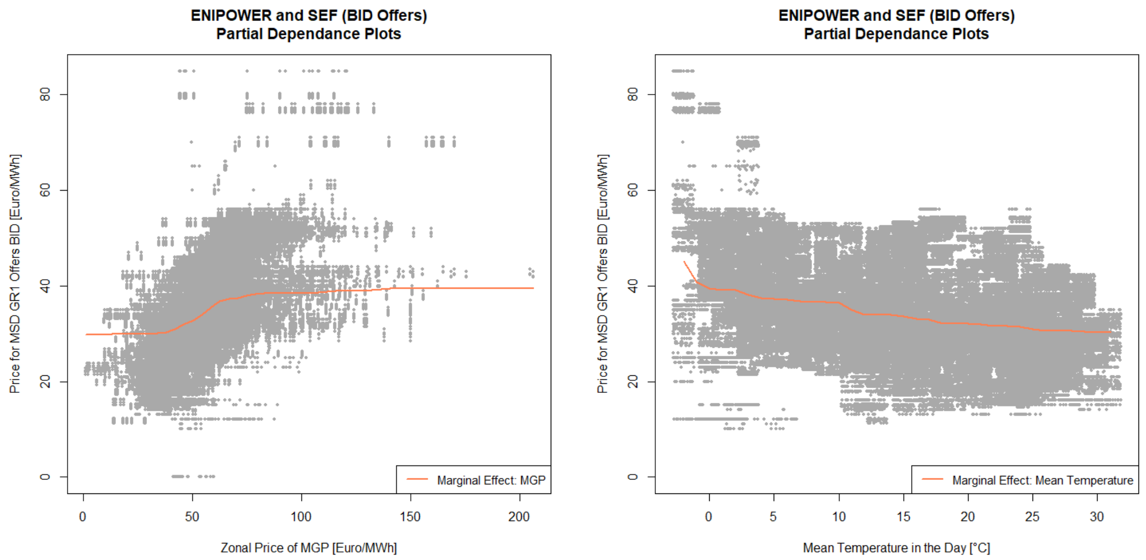


Figure 2: Cluster of Thermoelectric PUs from Enipower, prediction of MSD **BID bids**: visualization of marginal effects via Partial Dependence Plots for the variables **Zonal Price on MGP** (left) and **Mean Temperature** (right).

4.2. Cluster of Enipower, prediction of MSD bids of OFF type

Keeping the attention on the cluster of thermoelectric units of Enipower, I found **analogue results** for MSD bids of OFF type. Even in this case, I have decided to fit a Random Forest Random Input model, with a total of 50 trees. The OOB Mean Squared Error is decreasing with the number of trees, eventually levelling after the first 35 iterations. The final model presents a value for the **OOB cross-validation error** of 14.73 $(€/\text{MWh})^2$, together with a **percentage of variability explained by the model** equal to 97.61%. These values are slightly worse than those of the model for BID offers, but this is mostly due to the fact that **OFF bids have a wider range than BID ones**, and thus we can consider the model to be highly accurate even in this second case.

Looking at variable importance for the model with OFF offers as a target, reported in Figure 3, we can observe that the highest ranked variables are the same as in the previous case, meaning that the operator is somewhat consistent for both types of bids.

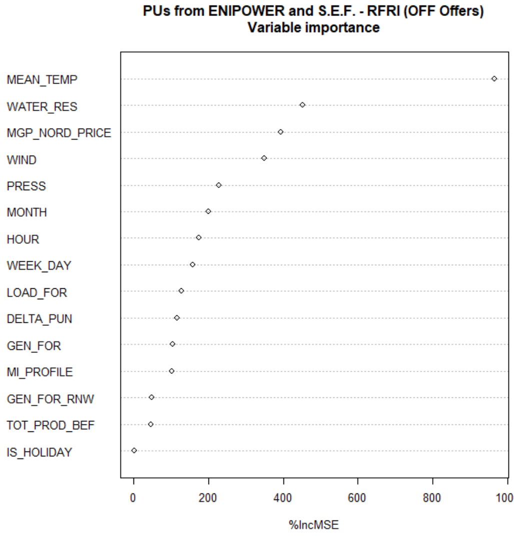


Figure 3: Cluster of Thermoelectric PUs from Enipower, Random Forest algorithm for MSD GR1 **OFF Offers**: Variable Importance.

To conclude this section, in Figure 4 I have reported the visualization of the **marginal effects for the two most important variables** in the model. From it, we can observe how the **Mean Temperature** influences the bids in quite a similar way than the previous case, since it is appears to be particularly relevant in the coldest days of the time span. Moreover, the price of upward MSD GR1 bids tends to increase with the **MGP zonal price**, in an analogue way to the BID case, meaning that the operator tends to keep his offers within a defined range from the price of energy on the Day-Ahead Market.

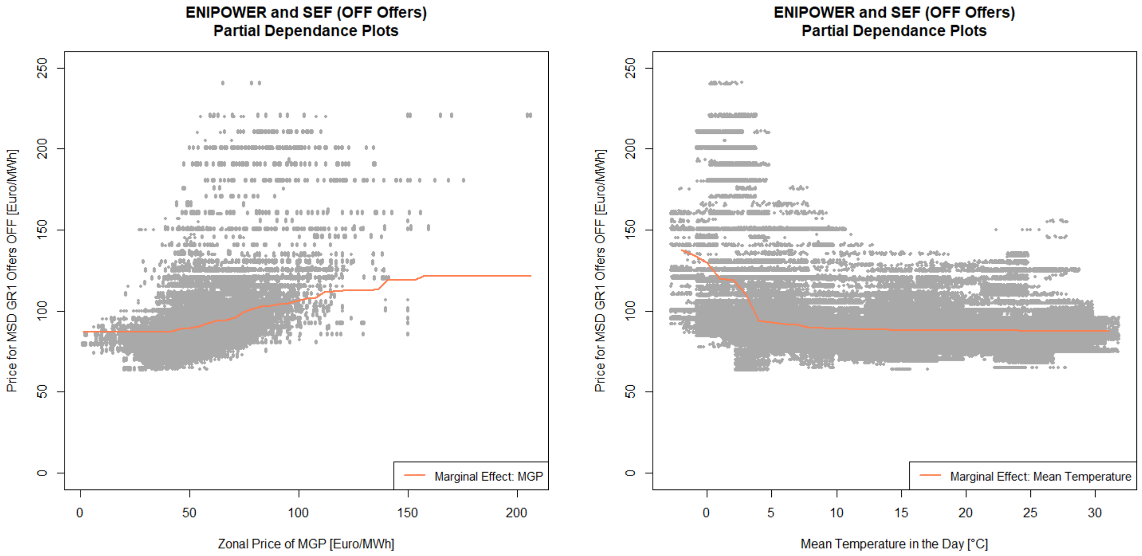


Figure 4: Cluster of Thermoelectric PUs from Enipower, prediction of MSD **OFF bids**: visualization of marginal effects via Partial Dependence Plots for the variables **Zonal Price on MGP** (left) and **Mean Temperature** (right).

4.3. Cluster of "Thermoelectric PUs active on MSD - BID", prediction of MSD bids of BID type

In the conclusion of my work, I predict and interpret the bids from the cluster of **Thermoelectric Production Units active with BID offers on MSD GR1**. Eventually, I have decided to only fit the model for the **prediction of BID offers** for this cluster, focusing on the five Production Units of *Leini*, *Livorno Ferraris*, *Torviscosa*, *Voghera* and *Tavazzano*.

I proceeded to fit the Random Forest Random Input algorithm for this specific case, setting the number of trees in the forest to 70. In Figure 5 I have reported the **Out Of Bag error** as a function of the number of trees in the forest. From the graph, we can observe that the OOB error is decreasing in a smooth way, eventually reaching the value of 16.80 (€/MWh)^2. Regarding the **percentage of variance explained by the model**, we find that it is equal to 90.5%, a value that is lower than previous cases, but still incredibly good, considering that we are dealing with units from three different operators.

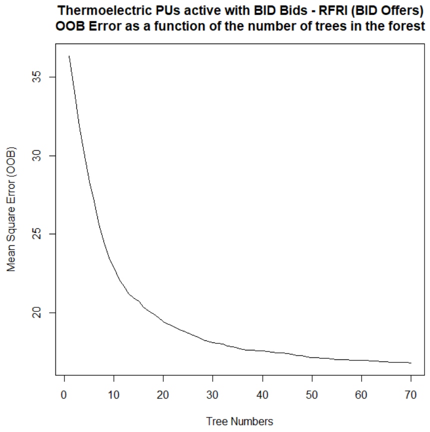


Figure 5: Cluster of Thermoelectric PUs active with BID bids, learning curve of the Random Forest algorithm for MSD GR1 **BID Offers** (OOB error as a function of trees in the forest).

Moving on to the **study of Variable Importance** for the model under consideration, we find that the **highest ranked variables are similar to those of the cluster of Enipower**, with the addition of the one indicating the **Equivalent Hours of Production on the previous day** (`TOT_PROD_BEF`). Indeed, the Equivalent Hours of Production end up being the most important variable in the cluster, causing a worsening of the error of about 100%, when it is permuted within the observations in the OOB sample. Soon after we find the two variables that dominated the previous models, namely the **Zonal Price of electricity** on MGP and the **Mean Temperature**.

Moreover, in this case it is quite important also the variable `MI_PROFILE`, related to the production percentage of the units as a result of the

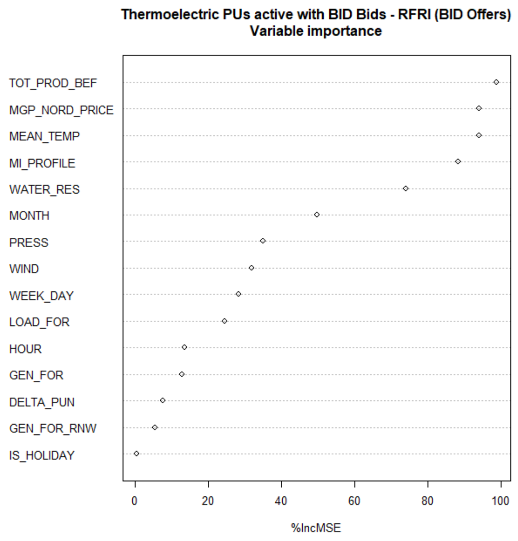


Figure 6: Cluster of Thermoelectric PUs active with BID bids, Random Forest algorithm for MSD GR1 **BID Offers**: Variable Importance.

Intra-Day Market (MI).

Regarding Partial Dependence Plots, they show how the bid price tend to increase with the **zonal price on MGP**, exactly as for the operator Enipower. Furthermore, even in this case, we can observe a prominent effect of the variable **Mean Temperature**, with offer prices on MSD GR1 - BID that decrease, on average, as the climate gets hotter, even though in this case there are no particular peaks in colder days.

5. Future Analyses

The conducted work lays the basis for a set of additional interesting analyses. Indeed, many straightforward extensions are possible, starting from the consideration of also other Italian bidding zones in the clustering.

Moreover, it would be quite interesting to **expand the temporal horizon of these analyses**, both to the **Covid-19 pandemic** months (which sharply decreased the demand of energy in Italy) and to the month of October 2021, when the price of electricity abruptly increased.

Another possible extension of my work on MSD, would be the **definition of more complex models, accounting for other ancillary services** rather than only the GR1. In fact, a more refined model would comprehend also the service denoted as "AS" - *Accensione/Spegnimento* in our dataset, which stands for Switch On/Off[3].

Finally, one last interesting topic to study would be the **acceptance of bids on MSD by the TSO Terna**, since during my research work I have found two interesting facts in this direction and since I believe it has **crucial importance for future analyses**. This subject was already partly studied in [3], and I decided to dedicate a separate section to this topic in my work. The first issue is the apparent existence of a **merit-order** procedure for the acceptance of bids on MSD in each specific **Grid Supply Point**. Indeed, if this hypothesis was confirmed, it would imply the existence of **segments of units** that are effectively in **close competition** to produce on the **highly remunerative market MSD**.

The second discovery is related to a **sudden change** in the acceptance mechanics by Terna for **hydroelectric units**, which **stopped being accepted on MSD** after February 13th, 2019, even if they left unchanged their bids from previous months. Unfortunately, our analysis only consider the years 2017 to 2019 and the identified tendency should be studied more in depth in recent times, assessing if it is still in place nowadays. Nevertheless, this appears to be an evident change in behaviour by Terna.

6. Conclusions

The liberalization of electricity markets gave birth to extremely competitive environments, where **operators** can **bid** in **strategic ways**, in order to maximise their revenues from *all* the units they control.

Indeed, during my work I found out that the great majority of Production Units powered by **Renewable Energy Sources**, especially those with **limited** installed capacity, tend to submit non-competitive bids on MSD. In fact, they generally offer at constant prices for many weeks, or even months and seem effectively **uninterested in taking part in MSD**. On the opposite, **thermoelectric power plants are more "ac-**

---

[3] Indeed, the units from the operator Enipower, which seem to place the best possible GR1 bids, have almost no accepted offers of the AS type. In contrast, the units from ENEL and A2A seem to begin their production only on MSD quite often, as a result of accepted upward AS bids. If confirmed, this hypothesis may indicate a profound difference in the approach of the units to the market MSD, possibly driven by the specific technological characteristics of each power plant.

tive" on the market, and consider MSD as a **potential** way to cope with their limited revenues from the markets that are based on a merit order, where the price of electrical energy is lower. In general, though, the **units eligible for MSD** place their bids only **considering few variables**, that are generally limited to the price of electrical energy on MGP, seasonal variables, and some **unit-specific conditions**, such as the production profile after the market MI and, for hydroelectric Production Units, the amount of water in their basin.
Focusing on **thermoelectric units**, I found major differences among the analysed operators. For instance, the **company Enipower** places the **same bids** across **all the units** it controls, with a high degree of **variability** and bids that change on a hourly basis, but without taking too much care of the working range after the Intra-Day Market of its Production Units. In contrast, the **operators A2A and ENEL** seem to differentiate more their bids across the units they control, but they update them less often than the cluster from Enipower. Specifically, the operator **ENEL places its bids on a daily basis** (i.e. constant prices for 24 hours straight), thus disregarding completely the forecasted load/generation and the consequent fluctuations of the MGP and MI prices. For two of the identified clusters, I was able to **predict with incredible accuracy** the submitted bids, thanks to the algorithm **Random Forest**.

To conclude, most of the offer curves for the **service GR1 on MSD** appear to be **somewhat perfectible**, with many plants that do not care to bid at competitive prices at every hour. This is probably due to the **difficulty** of the operators to **predict** when there will be an **effective need** for their service on MSD. Thereby, I suggest that future research should focus on the prediction of the need for each ancillary service at every hour, also **explaining the acceptation dynamics** on MSD from the Transmission System Operator Terna.

References

[1] Pietro Innocenti - Politecnico di Milano [AY 2019-2020]
Master's thesis: *Statistical Models and Algorithms for the analysis of the Electricity Market and the Economic Dispatch Service*

[2] Virginie Marchionni - Politecnico di Milano [AY 2019-2020]
Master's thesis: *TAJIVE: a new tool for integrative analysis of multidimensional data. Application to the Italian Energy market.*

[3] F. Bovera, A. Blaco, G. Rancilio, M. Delfanti [2019]
Assessing the Accuracy of Different Machine Learning Classification Algorithms in Forecasting Results of Italian Ancillary Services Market. - In: 2019 16th International Conference on the European Energy Market (EEM), pp. 1–5.

[4] Robin Genuer, Jean-Michel Poggi
Random Forest with R - Springer, "Use R!" series

[5] GME - Gestore dei Mercati Energetici SpA
Mercatoelettrico.org - Download Section

[6] ENTSOe - European Network of Transmission System Operators
Electricity Market Transparency Platform

POLITECNICO
MILANO 1863

SCHOOL OF INDUSTRIAL AND INFORMATION ENGINEERING
Master of Science in Mathematical Engineering: Applied Statistics

Prediction and Interpretation of bids in the Italian Electricity Market for Ancillary Services

Andrea Pasotti
10493242

Advisor:
Prof. Piercesare Secchi

Co-Advisor:
Eng. Filippo Bovera

Master Thesis developed in collaboration with:
R.S.E. - Ricerca sul Sistema Energetico

Academic Year 2020-2021

Vita

Ho cercato
E non ho trovato
Ho pensato
E non ho risolto
Ho agito
E non ho raggiunto!
...Ma allora?
Ho vissuto
È quanto basta

Vin,
14 Febbraio 2016

Abstract

Recent rises in the price of electrical energy have drawn incredible attention to Italian Electricity Markets. These marketplaces have experienced drastic changes in past years, mainly driven by the desire to transition to a more sustainable and green economy, which entailed the growing diffusion of power plants fueled by non-programmable Renewable Energy Sources. Indeed, these units are able to produce much cheaper and cleaner energy than traditional CCGT power plants but are causing a sharp increase in the measures required for the regulation of the Power System, with balancing resources that have to be procured by Terna, the Italian Transmission System Operator (TSO), on the Ancillary Services Market (MSD).

The following Master's thesis has the aim to provide a wide-ranging analysis of MSD, thanks to a thorough study of the offer profiles submitted by the Production Units operating in Northern Italy. In particular, the work will be focused on the specific service denoted as *"GR1"* on MSD ex-ante, considering bids of both upward and downward types. These offers will be interpreted on the basis of many predictors, including both endogenous and exogenous data to the electricity markets.

Our findings confirmed the intricacy of the market MSD and showed how different operators tend to manage their power plants in various ways, with disparate degrees of both variability and *effectiveness* of the submitted bids. Furthermore, the presented results combine the technical knowledge of the departments of Energy and Mathematics of Politecnico di Milano and the practical experience on electricity markets of the Research Institute RSE, which supported the initial portion of the work.

Sommario

I recenti aumenti del prezzo dell'energia elettrica hanno attirato un'incredibile attenzione sui mercati elettrici italiani. Questi hanno inoltre subito drastici cambiamenti negli ultimi anni, spinti principalmente dal desiderio di transizione verso un'economia più sostenibile e *green*, che ha comportato la crescente diffusione di centrali elettriche alimentate da fonti energetiche Rinnovabili di tipo non programmabile. Le unità di questo tipo sono infatti in grado di produrre energia molto più economica e *pulita* rispetto alle tradizionali centrali termoelettriche, ma stanno determinando un forte aumento delle misure necessarie per la regolazione del Sistema Energetico, con risorse di bilanciamento che devono essere procurate da Terna, il gestore di trasmissione energetica italiano (denominato, in inglese, *TSO*), sul Mercato dei Servizi di Dispacciamento (MSD).

La seguente tesi di Laurea Magistrale si pone come scopo quello di fornire un'analisi ad ampio raggio di MSD, grazie ad uno studio approfondito dei profili di offerta presentati dalle Unità Produttive operanti nel Nord Italia. In particolare, il lavoro si focalizza sullo specifico servizio denominato *"gradino 1"* (GR1) su MSD ex-ante, considerando offerte sia al rialzo che al ribasso. In aggiunta, le offerte saranno interpretate sulla base di molti predittori, incluse variabili sia endogene che esogene ai mercati dell'energia elettrica.

Le nostre analisi confermano sostanzialmente la grande complessità del mercato MSD e mostrano come vari operatori tendano a gestire le proprie centrali elettriche in maniera differente, con diversi gradi sia di variabilità che di *efficacia* delle offerte presentate.

Infine, i risultati esposti uniscono le conoscenze tecniche dei dipartimenti di Energia e Matematica del Politecnico di Milano, con l'esperienza pratica sui mercati energetici dell'Istituto di Ricerca RSE, il quale ha supportato tutta la parte iniziale del lavoro di tesi.

Contents

List of Figures

List of Tables

List of Acronyms

CARTs	Classification And Regression Trees	101
CCGT	Combined Cycle Gas Turbine	57
EOH	Equivalent Operating Hours	53
GME	Gestore dei Mercati Energetici S.p.A.	11
MGP	Mercato del Giorno Prima	18
MI	Mercato Infragiornaliero	20
MSD	Mercato dei Servizi di Dispacciamento	21
PUN	Prezzo Unico Nazionale	20
PU	Production Unit	22
RSE	Ricerca sul Sistema Energetico	13
RTN	Rete di Trasmissione Nazionale	13
SEN	Sistema Elettrico Nazionale	13
TSO	Transmission System Operator	11

Introduction 1

This work of thesis presents the results of a joint collaboration between the departments of Energy and Mathematics of **Politecnico di Milano**, together with the research institute **RSE - *Ricerca sul Sistema Energetico***, and it has the purpose to perform an extensive statistical analysis of the Italian Electricity Markets, with a particular attention to the **Ancillary Services Market (MSD)**.
The market for ancillary services is the one in charge of real time energy tradings, and it is used by the Italian Transmission System Operator Terna to guarantee the security of the Power System, in terms of balancing energy and voltage profiles.

Indeed, the **growing diffusion of power plants fueled by non-programmable Renewable Energy Sources** is causing an increase in the measures required for the regulation of the Power System, with balancing resources that have to be procured by the TSO on the Ancillary Services Market. Moreover, the unpredictability behind the need for these ancillary services generates a market that is both **extremely complicated and intriguing to study** under a statistical point of view.

The Italian market for ancillary services was already analysed in the master's thesis work of Pietro Innocenti [2], with a specific focus on the event denoted as *opposite call*, that happens when a power reserve created by Terna in MSD ex-ante is not used in real time. Virginie Marchionni studied the variability in the offer curves by two operators active in Northern Italy, within her master's thesis [3], thanks to the use of a novel tool for the integrative analysis of multidimensional data. Furthermore, F. Bovera et. al. [4] provided a comparison of different classification algorithms in forecasting the results of "Step 1" bids of upward type, on MSD ex-ante.
This work has the aim to built on the existing literature, providing a **wide-ranging analysis of MSD**, thanks to a thorough study of the offer profiles submitted by the Production Units operating in Northern Italy.

In particular, I will mainly focus on the specific service denoted as "GR1" on MSD ex-ante, considering bids of both upward and downward types. While doing so, I will highlight the relationship between the price offered by the Production Units eligible to take part in MSD and many predictors, including both **endogenous and exogenous data** to the electricity markets. I will also pose particular attention to the **technology and managing operator** of each power plant, identifying and interpreting any strategic behaviours that will emerge from the analyses.

As previously introduced, the work was primarily developed during eight months of **internship at the research institute RSE**, which I did between December 2020 and July 2021. For the initial months, I have worked in close collaboration with Sara Martucci, another intern from *Politecnico di Milano*, who helped me collecting the datasets for the statistical analyses and collaborated with me on the preliminary studies of the Italian electricity markets. Due to this, Chapters 2, 3, 4 and 5 were written as a **four-handed project** together with Sara Martucci.

My original desire was to produce a master's thesis that was **accessible** even to the least experienced in the topic of energy markets. For this reason, I begin in Chapter 2 with a quite general discussion on the topic of electricity and the power system, aimed at explaining the underlining complexity of energy production, transmission and distribution. In Chapter 3, then, I explain how electrical energy is traded on public sessions, introducing the three Italian electricity markets named MGP, MI and MSD. I then move on to a better specification of the research question, presented in Chapter 4, before moving on to the data collection process for the subsequent statistical analyses, reported in Chapter 5. Chapter 6 marks the beginning of the statistical analyses, and comprehends a first overview of the units operating in Northern Italy, that were clustered on the basis of their macro-behaviour on electricity markets. Eventually, in Chapter 7 I dive deeper on the topic of the Ancillary Services Market (MSD), analysing and predicting the bids from four of the previously identified clusters, thanks to the help of the algorithm **Random Forest**.

Eventually, I report an overview of future analyses that may be done on the topic, in Chapter 8, focusing on the importance of further understanding the process of acceptation of bids by the Transmission System Operator Terna.

Electricity and the power system 2

Electricity is fascinating.

> The energy used by a light bulb to light up a room is being produced **in almost the exact same time** at possibly hundreds of kilometers of distance, and we have no way to perfectly know from which power plant it comes from.

Most of us are lucky enough to experience and use electricity on a daily basis, in fact we may say that our lives rely heavily on the **continuous availability** of this source of power. This said, we generally know very little about the complexity behind its production and management.
This chapter is dedicated to the introduction of this incredibly interesting, although intricate, topic and has the goal to lay the necessary foundations to fully understand some of the decisions that were made in our statistical analyses.

We will therefore start with few notable **historical notions** around electricity in Section 2.1, to then proceed with the description of how it is massively **produced and transferred** to end users, thanks to current electrical grid systems, in Section 2.2. Finally, we will also introduce a practical example of the importance of creating a reliable and connected power grid, by explaining the **power crisis** that happened in Texas in early 2021.

2.1 Historical Notions on Electricity

Many great personalities in the history of mankind have worked on this subject, contributing to our evolution and shaping our current society.
It is staggering to think that the first concrete studies on electricity date back to Ancient Greece and **Thales of Miletus**, who observed and described static electricity with rods of amber[1].

---

[1] It is exactly from this phenomenon that stands the etymology of the word "electric", indeed *elecktron* is the Greek word for amber.

This said, the topic would remain little more than an intellectual curiosity for millennia[2], until major breakthroughs were achieved in the 18th century, mainly due to **Benjamin Franklin**, who proved the electrical nature of lightnings.

However, we can consider the first big step forward in electrical science as made by the Italian physicist **Alessandro Volta**, who created the first battery (an electrochemical cell, made of alternating layers of zinc and copper) in the year 1800. The 19th century also marked the birth of **electromagnetism**, a new branch of physics centered on the discovery of the unity of electric and magnetic phenomena. **Hans Christian Ørsted** observed that a compass needle was deflected from magnetic north by a nearby electric current; this idea captured the attention of the French physicist **André-Marie Ampère**, who begun developing a theory to understand the relationship between electricity and magnetism, stating the principle that later on became known as Ampère's law.

All the experiences and experiments presented so far have still very little in common with our modern idea of electricity, yet we may consider as a first *tangible link* the idea that the alternating current technology (the one we experience and use nowadays) is rooted in **Michael Faraday**'s discovery in 1831 that a changing magnetic field can induce an electric current in a circuit. This idea is at the basis of the technology that is currently used to produce the vast majority of electrical energy, that is by rotating magnets within closed loops of copper coils or other conductive materials, and we will cover this topic in much greater details in Section 2.2.1.

In the last 200 years we have learnt how to massively produce and store electricity: we live in cities entirely powered by electric current and we have domestic appliances that we are able to plug in and use at any time of the day. Furthermore, electricity plays a big role in our **safety and health**. As we may imagine, it is not trivial to ensure that the correct amount of energy is produced and transferred where it is needed, at any time. The next section is devoted to delve more into the technicalities of this topic, since we will present the infrastructure behind electric energy generation and the processes of transmission and distribution.

2.2 The Power System

In the introduction of this chapter we affirmed how the electricity we use in our daily lives has a very particular aspect, that is it must be consumed at the exact same time as it is produced, no matter how far the producer and consumer might be. In this section we will talk about the infrastructure

---

[2] To obtain a broader overview on the topic we refer to the following web page, which inspired many of the following concepts: **Electricity - History** from Wikipedia, the free encyclopedia

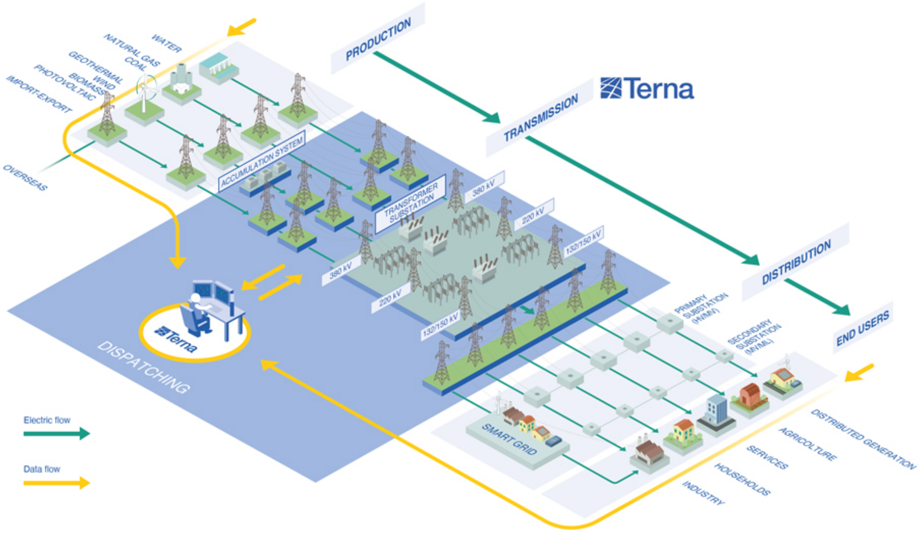


FIGURE 2.1: Structure of the Italian Power Grid, from [9]

that enables this process and allows it at an incredibly large scale: possibly being *one of humanity's most important engineering achievements*[3].

All notable notions to understand how this process works are summarized in the concept of the **Power System**[4], which deals with the generation, transmission and distribution of electricity. A first schematic visualization of the Italian Power System is given by Figure 2.1.
In the following sections we will mainly keep the focus on the first two aspects of the Power System, explaining some fine details that will be key to understand the production strategies of plants from different technologies.

Before doing so, let us just talk a little more about a particular aspect of the power system, related to the fact that all the energy produced in the grid needs to be consumed at exactly the same time. This might appear counter-intuitive at first glance, since we experience every day that electrical energy can be stored in the lithium batteries of our smartphones. This technology, though, mainly works at a low-scale and trying to replicate it for the electricity used to power a manufacturing plant or a house is very hard and expensive[5].

---

[3] We refer to an interesting series of videos on this topic, from where we took great inspiration in the structural organization of this section, as well as in some of the concepts presented: *Grady Hillhouse [Practical Engineering]* - ***The Power Grid***

[4] We use the term *Power System* to indicate a network of interconnected electrical infrastructures with the aim to supply, transfer and use electricity. At the same time, the term *Power Grid* (or electrical grid) can be used to denote only the phases of transmission and distribution.

[5] To be even more precise, there now exist also large-scale electrochemical systems,

This said, there exists actually a version of "battery" that operates at a large scale, but it has a completely different aspect than the usual concept of electrochemical energy generators. What we are talking about is often referred to as **pumped-storage hydroelectricity** and deals with a specific type of power plants that is composed of two different water reservoirs located at different altitudes, an electric turbine close to the lower reservoir and a pump which lifts water from the lower reservoir to the upper one.
This setup stores **gravitational energy** in the water contained in the higher storage point, which is filled with water pumped from the lower reservoir. The pumping process requires energy (electricity) to happen, this is usually done in periods of the day when electricity has a lower cost, such as at night. Conversely, at the times of the day when energy is more demanded, the gravitational energy stored in the upper source is converted via a turbine into electricity: the upper source is gradually emptied and the lower one is filled, producing electrical energy as a result. Obviously, most of the water is retained in this process, mimicking a process similar to that of a "common" battery.

2.2.1 Power System: Generation

The process of generating electricity is very complex and many types of **power plants** exist, each of them with different characteristics, advantages and disadvantages. They may be powered by both renewable and non renewable resources; nevertheless, all the plants have in common the concept of converting a particular kind of energy, into a different one: electrical energy.

At the end of the previous section we have talked about conversion from gravitational to electrical energy, however, the most widely used method to generate electrical energy comes from the conversion of heat, or **thermal energy**. This accounts for all the different ways by which water is heated to pressurized steam, which is then used to put a turbine into rotation.

Mechanical energy is then transformed into electric energy (at alternating voltage and current) by linking the turbine with a *rotor* (made of magnets) and a *stator* (generally composed of three coils of copper wire), which together form an electric generator. This process happens at an incredibly large scale, with turbines that can reach more than 1500 MW of power. Examples of this kind of energy transformation are all the plants fueled by oil, coal or natural gas but also renewable geothermal plants.

The remaining methods deal with different types of energy conversion, generally from renewable resources. Examples of this are the conversion of potential energy within **hydroelectric stations**, the conversion from

---

which work as "MW-scale" batteries. These are still quite new resources, but their adoption could provide to the power system useful functionalities in the setting of Ancillary Services.

rotational energy within **wind-powered stations** or the direct conversion made by solar photovoltaic cells: from light, into electricity.

Solar energy is basically cost-free with respect to any fuel-powered source of energy; it is also incredibly more environmentally-friendly than the common thermoelectric plants, since it does not produce carbon dioxide as a side effect. However, it also has two major disadvantages, the first one comprehensible to everybody: it is not always available, since it is affected by bad weather and stops after daylight hours. The second, though, is more subtle to understand and is represented by the absence of ***mechanical inertia*** of this power source.

As we stressed in the previous paragraphs, electrical energy needs to be generated at the same time as it is requested for consumption, since loads cannot normally store it; therefore it is crucial to ensure that supply always meets demand exactly. This concept of being able to generate the right amount of power even during localized faults or disturbances is sometimes called *regulating capacity* of the system, and plays a key part in keeping our energy supply stable and secure. This capacity includes both power variations due to inertial response from large rotating masses connected to the grid and power variations due to the intervention of controllers, in particular the ones in charge of regulating the grid frequency.

2.2.2 Power System: Transmission

Large power plants are generally located far from urban areas, usually for a matter of cost reduction or, as for the case of some renewable-energy power plants (like wind plants), for an easy access of the primary resources used in the production process. This is the reason why electrical energy generally needs to be transmitted for **long distances**: from power plants, to the most densely populated or industrial zones, where it is consumed.

Transmission can be carried out in AC or DC, usually at a **high voltage**: the current is boosted in voltage by transformers at the power plants before entering the transmission system, to ensure that the energy dispersed, so wasted, by the Joule effect is minimized in the transportation. This happens at a cost of a greater safety concern, since current at high voltages is able to flow across materials where it usually can not[6].

Another important aspect of the energy transmission process involves how the various **transmission networks are connected**. This issue is so crucial that in the last section of this chapter we will analyze more in depth an example underlining its importance.

All in all, the continuous rise of renewable sources of energy, the diffusion of new electric loads, such as electric vehicles and domestic heating, and the related process of decentralization of energy production, will most

---

[6] Such as, for instance, across air, creating an extremely dangerous phenomenon called *electric arc*.

likely contribute in the following years to a change in the structure of the transmission infrastructure. This said, the topic remains of crucial importance and we will encounter it again while building the dataset for our analyses.

2.2.3 Power System: Distribution and Consumption

The last component of the Power System is distribution. We have talked in the previous sections about how energy is transmitted towards urban areas, where it is mostly consumed, at a high voltage. Very close to big urban areas are then located some facilities known as **substations**, where transformers reduce the voltage of the alternating current arriving from the power plants, both for safety reasons and also to make it more easily disposable. Another important feature of substations is represented by the so called **breakers**, protection devices that serve the goal of isolating parts of the grid in case of faults.

We know that electric energy flows almost in real time from the power sources to the consumers, thus if perturbations occur in the grid (e.g. imbalances due to stumbling generation units), these may propagate instantly to every connection in the system, creating a cascade of problems and posing a serious threat to the stability of the whole infrastructure.

To summarize, **substations** exist to mitigate any problem that may occur to the grid, and they do so by interposing a layer between energy generation, transmission and its distribution to the public and to industries. Starting from substations, all the finer processes of distribution of electric current to any industry and building that may need it are also managed.

2.3 Texas, US: 2021 Power Crisis

It is crucial that every part of the previously presented system works as it is supposed to, with particular emphasis on the equilibrium between the energy produced and consumed in real time.

To further stress the importance of this topic, we may just consider the devastating **power crisis** occurred in Texas in February 2021. From February 10^{th}, in fact, the US State was hit by a series of severe winter storms, which caused great damages to some of the power plants in Texas, freezing part of the wind turbines and leaving all the power equipment which was not correctly winterized vulnerable to the extraordinarily low temperatures.

This, together with the increased demand of energy due to the cold temperatures, created a big imbalance between the production and supply of electricity.

To cope with the dangerous situation, the ERCOT[7] decided to begin

---

[7] **ERCOT**: Electric Reliability Council of Texas, in charge of operating Texas' power grid and managing part of the related (deregulated) market.

rotating outages[8], a procedure that consists of removing power from area to area, so that no single neighborhood is down for a long time, and has the aim to lessen the stress on the power system by reducing its total load.
This managed to avoid severe damages to the grid for just a matter of minutes, leaving though more than 4 million people without power and causing losses to real estate for approximately 195 million of US dollars.

This said, the extreme meteorological conditions were not the only cause of the power crisis in Texas[9], since we need also to take into account the very **peculiar characteristics of the grid** in this State. In fact, Texas is the only US State to be almost autonomous for what concerns the production and distribution of energy, and has an isolated power system from all other States in the USA. This decision was taken with the (mainly political) aim of obtaining greater autonomy in the regulation and commercialization of electric energy, and implied that Texas was unable to import enough energy from other States in a moment of desperate need and was finally unable to cope with the internal shortages.

This example highlights how critical it is to operate the grid in a safe and controlled way, but also stresses the importance of the prevention of possible shortages via the construction of **powerful connections** with neighbouring States, to be used in case of need. Indeed, this is what happens in Italy and Europe, where things work in a very different way.
The focus of the following chapter will be to explain how the Italian electrical grid is organized, focusing on its division into six different *zones*, which are interconnected and are also able to exchange electricity with all the neighbouring States.

---

[8] **NowThisEarth**: *Millions Without Power in Texas During Coldest Temps in Years*

[9] **Alvin Chang - The Guardian**: *Why the cold weather caused huge Texas blackouts*

The Italian Electricity Market 3

In the previous chapter we talked about electricity, focusing on its production and transmission via the **Power System**. The infrastructure behind this process works essentially in the same way in every part of the globe, with the main difference being represented by the technologies of the power plants connected to the grid, which can vary widely from nation to nation.

We will now describe how electricity is **commercialized**, with great emphasis on how the production process is controlled by the Transmission System Operator (TSO) and regulated via the market. To do so, we will take into account only the **Italian case**, since it represents the setting of our research analysis. We recall that the TSO in Italy is Terna, which is the institution in charge of the management and operation of the Italian Power Grid. In this regards we will present in Section 3.2 the division of the Italian electricity market into *bidding zones*, together with the structure of the Italian Power System.

In Section 3.3 then, we will introduce all the specific **market regulations** implemented by the market operator, called Gestore dei Mercati Energetici S.p.A. (GME). While doing so, we will go into the details of the Italian spot market *"Mercato a Pronti"*, introducing the intricate problem of **ancillary services**, which constitute the core of this master's thesis.

Finally, in Section 3.4, which concludes this chapter, we analyse again all the previous notions from the **standpoint of a Production Unit**. We made this decision since we found, during our research work, that there was still a big gap between the theory of the markets' functioning and the actual way in which each operator presented its bids.

For this reason, we thought that it would have been interesting to sum up everything from the point of view of a Production Unit that operates in the Italian market and has to decide how to make offers based on the surrounding context.

3.1 History of the Italian Electricity Market

We should start by specifying that in Italy the process of producing and commercializing big amounts of energy undergoes precise regulations, with both electrical energy and natural gas that are traded in two different and complex markets. All these exchange platforms, with their procedures, are called **IPEX**, which stands for Italian Power Exchange market and, for the electricity portion, undergo the directives of the so called *Borsa Elettrica Italiana.*

The Electricity Market was born with the liberalisation of the Italian electric sector, that happened in 1999 with legislative decree *D. Lgs. n. 79/1999*, best known as *"Decreto Bersani"*. This decree marks a structural change in the electricity sector, driven by the goal [11] to promote competition in the activities of production and wholesale of electricity. This was accompanied with the more specific desire to guarantee the maximum transparency and efficiency of the ancillary services, a crucial topic that we shall tackle in the following sections and will constitute the backbone of this thesis.

Among those changes, the aforementioned decree set the birth of three different entities, that (even though with some organizational changes in the following years) are still active to this day and play a crucial role in the Italian energy sector:

- **Terna S.p.A.**:
  Terna is the company entitled to manage the energy **transmission process** in Italy, as well as to ensure the safety and stability of the National Power System. It does that by monitoring the well-functioning of the electric infrastructure, studying the outcomes of the Electricity Markets and deciding the new strategic investments, according to the long-term needs of the Power Grid.
  Terna is effectively the most important player in the electricity market, having the role of **Transmission System Operator**, henceforth **TSO**.

- **GME**:
  The company that, as reported in [11], is responsible for the **organization and management** of the Electricity and Natural Gas Markets. In particular, it is in charge of the whole **trading infrastructure** behind the so called *Spot Electricity Market*[1], which consists of:

  1. Day-Ahead Market - MGP
  2. Intra-Day Market - MI
  3. Daily Products Market – MPEG
  4. Ancillary Services Market - MSD

---

[1] See also **GME's website**: www.mercatoelettrico.org/en/mercati/mercatoelettrico

In practice, the GME collects all the offers in these markets, it evaluates them (with the exception of MSD, which is managed in a slightly different way) and then communicates the related results.
Regarding the **Ancillary Services Market (MSD)**, we have to specify that it is managed by the GME on behalf of Terna. This happens because MSD is quite a peculiar market and it has the major objective to ensure the stability and secure operation of the Power System, as we will see in further details in Section 3.3.3. For this reason, it is the TSO that establishes the requirements for this market and it communicates them to the GME.
Most importantly, the GME performs its duties by undergoing principles of neutrality, transparency and objectivity: promoting fair competition in the market.

- **Ricerca sul Sistema Energetico (RSE)**:
  A research institute, whose goal is to perform public interest programs of Research and Development addressing national energy, also tackling environmental and economic goals.
  RSE performs studies of public importance on subjects related to the *Electric System Research*, that is everything concerning the electricity system innovation, its technological development and topics of general interest such as efficiency, economics and materials, process and device experimentation.

It is also worth noting [7] that the aforementioned structure is of relatively new establishment, with Italy that was one of the last Western countries to institute a TSO. This holds true also for the Italian Power Exchange Market, including the *Borsa Elettrica* (Electricity Market), since it only started its trading operations on the 1^{st} of April, 2004.

3.2 The Italian Power System

As we mentioned in Chapter 2, the Power System is composed of three different processes: production, transmission and distribution, where we have considered the last to also take into account the consumption of electric energy by industries and end users.
We will now focus on the Italian Power System, called in Italian *Sistema Elettrico Nazionale (SEN)*, defining its main features and peculiarities.

The Italian Power System is built in such a way as to optimize its management, in terms of both security and efficiency. In particular, the transmission stage relies on the National Transmission Network, in Italian *Rete di Trasmissione Nazionale (RTN)*, which accounts for substation facilities (electrical nodes) and transmission lines, and that is depicted in Figure 3.1[2].

---

[2] An interactive version of the same map is available on **ENTSO-E website**, at the following link: https://www.entsoe.eu/data/map/

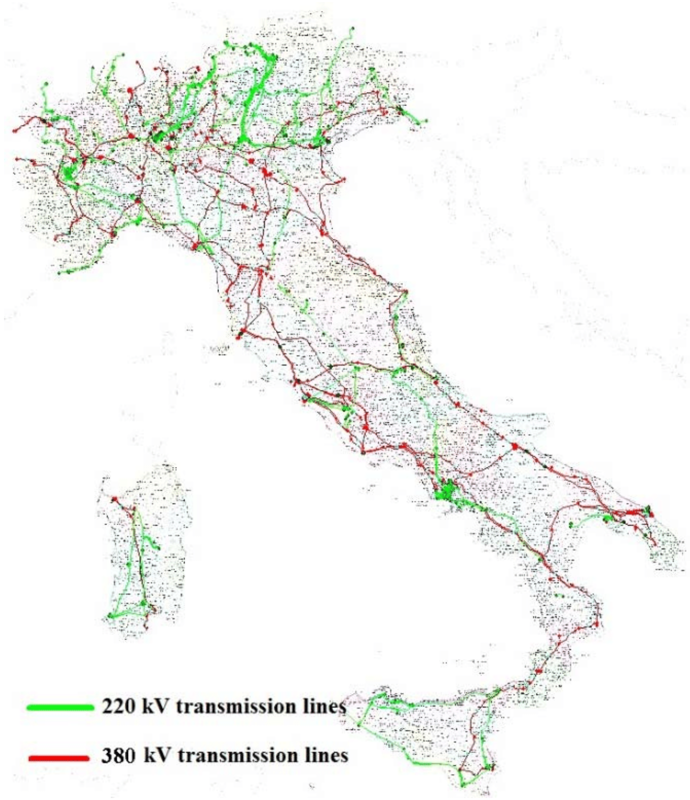
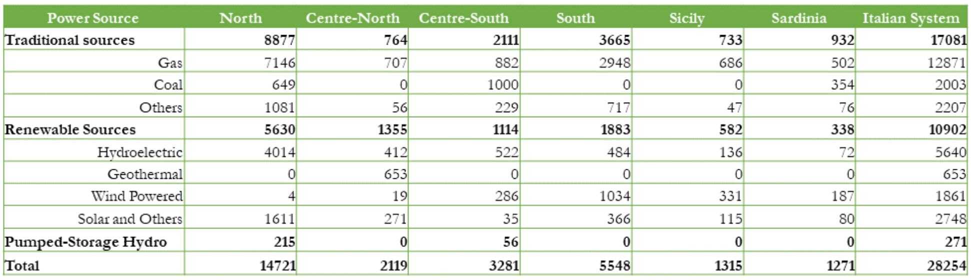


FIGURE 3.1: Power lines in Italy, at high and ultra high voltage [9]



Power Source	North	Centre-North	Centre-South	South	Sicily	Sardinia	Italian System
Traditional sources	**8877**	**764**	**2111**	**3665**	**733**	**932**	**17081**
Gas	7146	707	882	2948	686	502	12871
Coal	649	0	1000	0	0	354	2003
Others	1081	56	229	717	47	76	2207
Renewable Sources	**5630**	**1355**	**1114**	**1883**	**582**	**338**	**10902**
Hydroelectric	4014	412	522	484	136	72	5640
Geothermal	0	653	0	0	0	0	653
Wind Powered	4	19	286	1034	331	187	1861
Solar and Others	1611	271	35	366	115	80	2748
Pumped-Storage Hydro	**215**	**0**	**56**	**0**	**0**	**0**	**271**
Total	**14721**	**2119**	**3281**	**5548**	**1315**	**1271**	**28254**

FIGURE 3.2: Zonal Production by Technology, average hourly values [**MWh**] for the year 2018 [12]

Most importantly, in order to ease the management and the control of the Grid taking into account the limited transmission capacity of some transmission lines (i.e. critical cross-sections), Italy has been divided into **six macro areas**, called *geographical zones*: North, Centre-North, Centre-South, South, Sardinia and Sicily.

These *geographical zones* were defined by Terna, studying where it was most likely to observe power grid congestions, and were defined in such a way

Zone	C-North	C-South	North	Sardinia	Sicily	South
Load [10^6 MW]	94.95	136.42	494.83	25.90	52.79	77.91
Percentage	0.11	0.15	0.56	0.03	0.06	0.09

Table 3.1: Total Load for the Italian *Geographical Zones*, in the three-year period from 2017 to 2019

that some technical parameters are uniquely valid for every single zone. We have reported them in Figure 3.3, where we can observe how they trace the geography of the country, with two of them being the regions composing Insular Italy. Each zone is characterized by very different volumes of energy (both as to consumption and as to generation), with the **North zone being by far the biggest one** in this regard, accounting for 56% of the total energy demand in the three-year period of our analysis: 2017-2019[3].

One great difference among the zones is the diversification in the **production technology**, that logically reflects the geographical characteristics of the zone's environment. In particular, we have that wind-powered plants constitute a great share of the total energy generation in the most southern regions, while the same technology is less adopted in the North. For hydroelectric energy the exact opposite occurs: with the biggest concentration of power-generating stations located in the Northern Alps. We have reported a representation of this topic in Figure 3.2, where we can appreciate the average production quantity of each zone, by technology, for the year 2018. It is interesting to notice how Italy is effectively the seventh country worldwide by **geothermal** installed capacity[4], but all its production plants are located in Center-North, inside the region Tuscany.

Some of the six *geographical zones* are directly interconnected with each other (as depicted in Figure 3.3) and there exist some technical limits to the amount of energy that can be transferred from one zone to a connected one. These limits are determined by the TSO, who makes data about them available to market operators; it also publishes such data on its website ex post.

Most importantly, border zones communicate with **foreign power systems** through 8 *virtual zones* (France, Switzerland, Austria, Slovenia, BSP, Corsica, Corsica AC, Greece). This organization is promoted at a European level and is done to contain any problem of similar nature to that introduced in Section 2.3.

---

[3] More information are available in Table 3.1, and were retrieved from our database, which will be presented in Chapter 5.
We refer to Figure 3.2 to gain an overview of the zonal production. From that picture we can observe that in the year 2018, the North zone accounted for about 52% of the total energy production in Italy, a datum quite close to the total load from that zone.

[4] Alexander Richter - Top 10 Geothermal Countries [28 September 2018]: Article from **ThinkGeoEnergy** Website

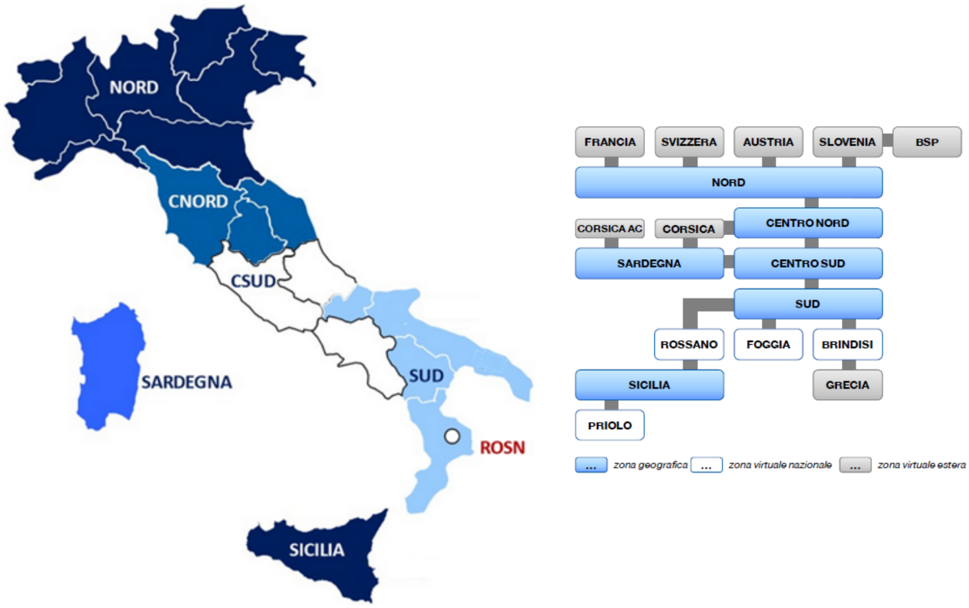


Figure 3.3: Division of Italy into *geographical zones* (left) and connections among them and with foreign zones [11] (right); situation valid prior than 2021 and including the period of our analysis

Moreover, this connection between different countries is of crucial importance, since it allows TSOs to make purchase and sell contracts of energy bundles among their closest countries. Just to make an example, it is commonly known that in France electrical energy usually costs quite less than in Italy, and for this reason Italy buys great amounts of energy from the transalpine country.
This difference is mainly due to the presence of **Nuclear Power Plants in France**, which existence is forbidden in Italy for safety concerns, but still are able to provide large amounts of energy at a relatively low price. One big exception to this trend happened in early 2017, when major problems occurred to some French electrical facilities and Italy was very actively exporting electricity towards France in order to cope with their shortages. In the following chapters we will see how this highly affected electricity prices, especially in Northern Italy.

To conclude this section, we just have to mention that the Italian Power System undergoes **continuous development**, for instance to generation plants and the transmission infrastructure. Besides, from time to time there are organizational changes, such as the one that happened in 2021 (therefore after the period of our analysis) to the sub-division into different bidding zones, which (among other measures) included some changes related to the management of the poles of limited production.
All these modifications contribute to the high variability of the markets and contribute to the definition of an incredibly interesting, yet challenging, data analysis setting.

3.3 The Italian *Borsa Elettrica*

The Italian Electricity Market, which is organized and managed by **GME** [11], is a telematic marketplace for the negotiation and wholesale of electric energy, where all the programs of intake and off-take of energy from the Power System are defined.
Offers in the market are rewarded according to the merit criteria, i.e. an offer is accepted if it has the "best" price among the feasible ones. Moreover, it is useful to specify that the *Borsa Elettrica* is not a compulsory market, since operators can make private deals outside the marketplace, through the so called *bilateral contracts*, that are usually identified by special tags in the offers.

From a structural standpoint, the exchange of electricity is organized into two different parts: a spot market called *Mercato a Pronti* and a forward market, the *Mercato a Termine.* In our work we only focus on the first and more important one that deals with short-term energy tradings and we will now describe the three sub-markets into which it is divided: MGP, MI and MSD.
MSD, in turn, is divided into two stages: a procurement stage and a subsequent and final balancing market (denoted by **MB**), that is used by the **TSO** to balance demand and supply in real time. This one, though, does not change much the amounts defined in the prior markets and it is quite unpredictable, since it is based on the real time difference between the forecasted load and the actual observed load, difference which can also be heavily affected by non-programmable renewable generation. For all these reasons, we have decided not to consider MB in our analysis.

The electricity markets are sub-divided for every Italian bidding zone, but work in essentially the same way in each of them. In fact, all the zones belong to the same market but, in case of congestion, transfer capacity limits between zones are taken into account while solving the market. Looking at this subject in detail, we find that each *geographical* or *virtual* zone described in Section 3.2 is the union of **offer points** [11], that are the minimal units with respect to which the hourly programs of inputs and off-takes have to be defined.
Input programs should be unique to every single production unit, so that the **TSO** is able to identify in a unique way how to supply resources for the ancillary services from all the eligible units. For what concerns the off-takes, instead, the offer points that withdraw energy can correspond to both a single intake unit or to an aggregation of such intake points. With that being said, in our analysis we will not consider such *consumption points* and only focus on the production units.
To every offer point is then associated a "dispatching user". This is responsible for the execution of the intake and off-take programs; moreover, it is also required to execute every dispatching order that might be required by Terna in order to ensure the security of the system.

This concludes the preliminary description of the organization of the Italian Electricity markets. Starting from the following section, we will describe each of the mentioned markets in greater details, with a particular emphasis on the Ancillary Services Market, that will be the final target for our research analysis.

3.3.1 MGP - Day-Ahead Market

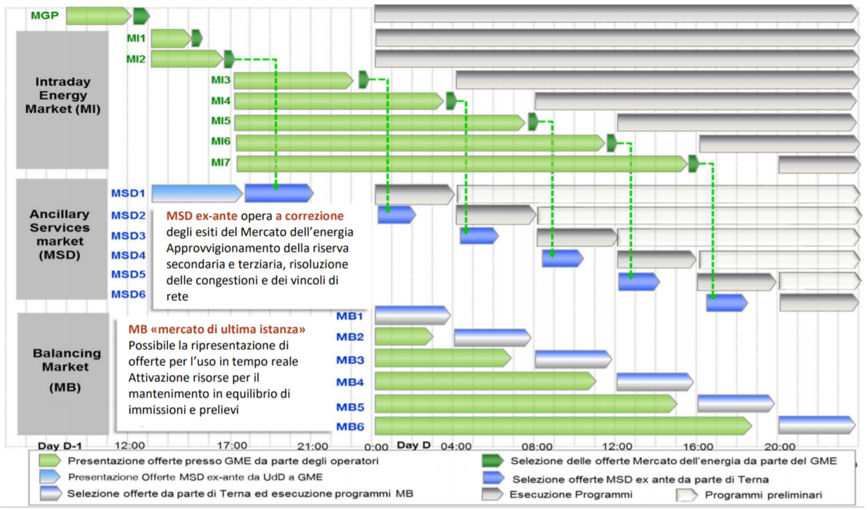


Figure 3.4: Structure of the Italian *Mercato a Pronti* [7]

In the previous section we introduced the Italian *Mercato a Pronti*, where all the major producers of electrical energy negotiate their production programs for every specific hour of the day. In order to guarantee a greater flexibility to the operators, the offers are presented within three different sub-markets, and we refer to Figure 3.4 to better understand their sequentiality.

The **Day-Ahead Market** (in Italian *Mercato del Giorno Prima*, henceforth MGP) is time-wise the first and it is also characterized by the highest amount of energy exchanged. In this market, bids for a specific hour of day d are collected up to day $d-1$ at most, and, if accepted, will represent the baseline for the production or consumption profile of the unit under consideration.

The price limit for selling bids is now equal to 3000 €/MWh (even though initially a 500 €/MWh price cap existed, then removed) and purchase bids can be without indication of price, meaning that the unit is willing to buy electricity at any cost [7]. Furthermore, all bilateral bids that may be defined with private contracts are added in the market as couples of selling bids with zero price and purchase bids without price indication, meaning that there is no actual way for us to understand the real price behind such bids.

The *Mercato del Giorno Prima (MGP)* is a very peculiar market, since the final energy price coincides with the **market clearing price**, which is the equilibrium price obtained as the intersection between the quantities (i.e. the amounts of electric energy) demanded by consumers and those offered by the production units that take part in the market. In Figure 3.5

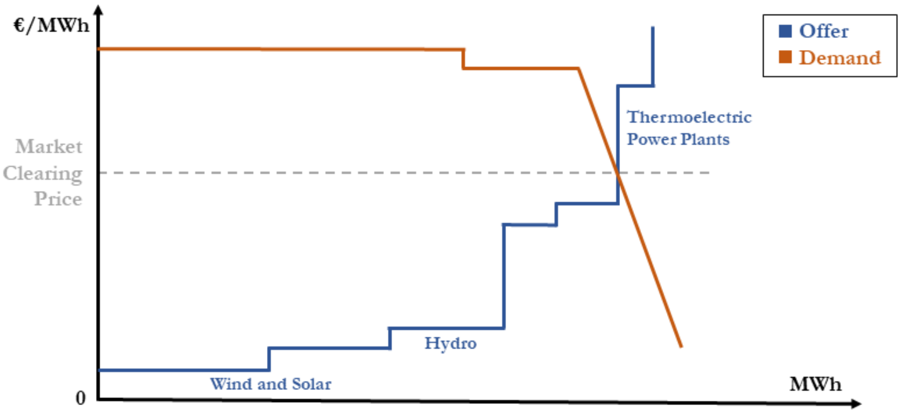


FIGURE 3.5: Demand and supply curves, plus their intersection (adapted from [7])

we reported an example of such procedure, where we can see that selling bids are ordered to form a growing blue line according to their price, and, at the same time, purchase bids are depicted with a dark orange monotone non-increasing curve. The intersection of these curves defines the final price for electric energy and this procedure is carried out for each hour of day *d*.

As we were saying, the outcome of MGP is evaluated at a national level, accounting for the purchase and sell curves of the units from all the six bidding zones. However, while doing that, the GME also takes into account the **transmission constraints** of the power grid. In fact, it may not be true that the energy required by each single *geographical zone* is equal to the quantity to be produced within the same zone[5]. On the contrary, it is usually the case that these two quantities differ, and a portion of the total energy might be transferred across the zones, in order to meet the demand and supply at the market clearing price. Sometimes though, it could not be possible to transfer all the requested energy, due to the transmission constraints across the zones, and we have a situation known as **congestion** of the grid.

When this is the case, the market is split into all the geographical zones, with exporting zones that will be characterized by lower prices for electric energy, while the opposite will happen in the importing zones. This happens because it is not feasible to physically transfer all the energy from one zone

---

[5] Just to give a fictitious example, in a particularly windy and sunny day we may observe massive quantities of energy accepted to the renewable power plants located in Southern Italy, that bid at a very low price. It is possible that this energy is partially requested in the zone North, but it is not feasible to transfer it all for that great distance. This creates a grid congestion that have to be resolved by the GME.

to the other, and in one of the two zones will be necessary to activate more expensive production units, in order to cope with the demand.

The market splitting procedure is performed automatically by the market operator (i.e. by GME), on the basis of information supplied by the TSO, thanks to an algorithm called **EUFEMIA** [7], and it also implies that selling bids are remunerated at the *system marginal price* of their *geographical zone*, that we recall might not be the same in every part of the peninsula. In contrast, purchasing bids will follow the so called ***Prezzo Unico Nazionale (PUN)*** that is the **same price** for every zone, evaluated as an average of the marginal prices in each zone, weighted for the demanded quantities of the zones. We also notice that, in the presence of congestions, the TSO collects money from the market, since it effectively buys electricity in places where it costs less, to sell it to other zones where it costs more: this revenue is a part of what is known as *congestion rent*.

The decision to opt for a unique *market clearing price* might seem counter-intuitive at first, since in practice every bid is rewarded as the highest price accepted to satisfy the demand; however, in the presence of a strategic behaviour from the operators in the market, this has been proven in many papers (we report [6] as an example) to be the **optimal solution for the public customers**, since it lowers the final price of electrical energy.

3.3.2 MI - Intra-day Market

The Intra-day Market, *Mercato Infragiornaliero (MI)* in Italian, is the second market and it serves the purpose to "correct" the production programs defined in MGP [7], since at the time when MI takes place, all the units know if their bids were accepted or not in the first market.

Indeed, what may happen after MGP is that the profile of a production unit may be unfeasible, due to ramp constraints not satisfied or start-up/shutdown problems (e.g. time requirements not met).

The market is divided into 2 sessions that are run after MGP but still on day $d-1$ (called *MI1* and *MI2*), plus 5 more sessions (*MI3* to *MI7*) that take place on day d, as reported in Figure 3.4.

In each session a unit may place two different types of bids, to increase their energy production (analogously, to reduce their energy consumption) or to decrease their energy production (analogously, to increase their energy consumption). It is also worth noting that MI is still remunerated with a system marginal price, this time only evaluated at a zonal level, without the presence of a unique price such as the **PUN** for MGP.

As a result of MI every unit knows its updated programs for energy production or consumption, with respect to the MGP programs.

3.3.3 MSD - Ancillary Services Market

The last market we take into account is the Ancillary Services Market, which in Italian is called *Mercato per il Servizio di Dispacciamento* and it is often

abbreviated to just MSD. This market is used by Terna [7] to guarantee the security of the Power System, in terms of balancing energy and voltage profiles.
In the specific case of **MSD ex-ante**, the stage we are considering, Terna has the goal to supply the necessary resources for secondary power reserve, tertiary power reserve and resources for congestion management.

As we stressed in the previous chapter, in a Power System the quantity of energy produced has to be exactly equal to the consumed one, at every time. When this is not the case we may observe a variation in the voltages or frequency in the Grid, which may lead to system instabilities and, eventually, blackouts.
Among the possible reasons for such variations in demand or production we may identify power plant failures, an unexpected change in the Net Transfer Capacity between zones or just the variability of non-programmable renewable resources.
The role of Terna is to **anticipate every possible imbalance** that may happen in the system and Terna does that by securing the right amount of power reserves at every hour.

Notice also that not every power plant is able to change immediately its production profile; especially thermal plants could require many hours to heat up and reach their maximum capacity. At the same time, since we are dealing with possibly great hazards in the system, Terna enables only a part of the Production Units to take part in the *Mercato dei Servizi di Dispacciamento (MSD)*, requiring that they satisfy some **specific prerequisites**[6]. These include constraints on the maximum capacity of the power unit (which should be at least equal to **10MW** of apparent power), technical minimum, ramping rates, start-up times, and Terna requires that the unit must also be powered by programmable energy sources.
All the Production Units operating in MSD have to be constantly connected to Terna's control system and must report to Terna every variation of the above cited technical details in advance.

All the units competing in MSD are also required to **make at least one offer** for the increase in production (till their maximum capacity), and one selling offer for all the quantity they may had in program to produce after MI. This is a major difference with respect to the previous markets MGP and MI, and is done in order to guarantee sufficient resources to cope with any possible hazard, at any time.
Furthermore, this market is **pay as bid** and not system marginal price, unlike the two previous ones, meaning that units are remunerated at the price at which they have bidded. This price, though, may not be the lowest available one, since Terna accepts bids on MSD in such a way as to minimize

---

[6] Terna does that according to the so-called *Registro delle Unità di Produzione*, which contains the list of the PUs needed by Terna to carry out ancillary services; it also contains their technical characteristics. Furthermore, the set of regulations issued by the TSO in order to specify all the technical constraints that a PU must satisfy to take part in MSD is specified in the *Grid Code*.

possible hazards to the Power System, taking into account only marginally the price of the bids.

The **unpredictability** behind the problem that generates the existence of this market, as well as the difference in the rewarding system for the operators, make MSD by far the most complex market to study, with bids from the Units that are not always easy to understand and justify.
At the same time, it is not always trivial even to appreciate how bids are evaluated by Terna and the reasons why those bids are accepted or rejected, since the TSO is not required to specify the reason why it calls a unit for a service.
Moreover, as reported in [7], it is often the case that a production unit is accepted for downward (or upward) bids in MSD ex-ante, only to reserve some power for the grid, but it is not guaranteed that this power reserve is then effectively used by Terna, since in real time we may not observe the actual need for it. When this is the case, the unit is then accepted on the last market, MB, with upward (or downward) bids respectively, with a process that takes the name of *opposite call* and was partly studied in [2]. This process is quite common, and contributes to the complexity for the PUs to define competitive strategies on MSD, as well as the difficulty to study the market for ancillary services at a statistical level.
In our work, we have tried to tackle some of these issues and we refer to Chapter 4 for major details on our research questions, as well as the models adopted to address them.

3.4 From the Production Units' Perspective

To conclude this chapter, we would like to assume the standpoint of a Production Unit (as opposed to a Consumption Unit) operating also in the ancillary services market and sum up every aspect of the previously presented topics: from the different market sessions to the bid types.

We start by specifying that in each hour of the day and for each market, a Production Unit (PU) can make two different kinds of bids, both of which must be accompanied by a related quantity of energy:

- **OFF Bids**:
  Upward offer for the PU, characterized by an increase in the energy produced and delivered to the market.

- **BID Bids**:
  Downward offer for the PU, which is requesting to produce less energy with respect to what was resulting from the previous markets. This type of bids only makes sense if the PU was active as a result of the markets prior to the considered one (i.e. it does not make sense for MGP, the very first market).

Related to this issue, it must be noted that OFF bids imply that the PU is remunerated, either at the system clearing price or at the adjusted price[7] of the original bid, according to the market of reference. BID offers, instead, are related to a "repurchase" of energy previously sold by the PU, therefore they imply that the PU uses its money to effectively purchase back from Terna that amount of energy that the TSO is requesting not to produce anymore.

We now analyse how a PU is able to operate within the markets, starting from the time-wise first one: **MGP**. In MGP each Production Unit can make up to a total of 8 different bids per hour (as usual, by means of a couple price-quantity), half of which must be of public type to compete in the market and the other half can be represented by bilateral exchange contracts. Since this is the very first market, the PU can only make OFF bids, as specified above. These offers will generally reflect the characteristics of the specific Production Unit, which has in its interest to make bids that will satisfy its ramp and technical minimum constraints. This is an incredibly important matter, especially for the thermal units operating with complex cycles (such as Rankine, steam-powered cycle), for which the initial ramp could take hours to fully heat up the machines. Finally, notice that the PUs are not obliged to bid in MGP, and this usually implies that the Units with a fixed amount of primary resources, such as pondage hydroelectric power plants, decide to make offers only in the busiest (and therefore highly remunerative) hours of the day, since otherwise they could not earn enough from the differential of price between daytime and nighttime hours.

After MGP, the PU analyses its production profile along the day, studying all the bids that were accepted in the first market. From this analysis (usually done by an algorithm), it will generally try to "smoothen" its production profile, offering upward or downward bids in **MI** in order to minimize the processes of switching on and off its turbines, since this is generally a very costly operation.

The **MSD ex-ante** market for ancillary services deserves a separate discussion, since the units that take part in it are **obliged to make bids** at every hour of the day. Within these bids, we identify four main different types that serve different purposes[8]:

- AS + OFF := Bids to switch on energy production, from zero to the minimal operating power of the PU
- AS + BID := Bids to switch off energy production, from the minimal operating power of the PU to zero

---

[7] The adjusted price is the price corrected by the TSO to meet the constraints of the market when they are not respected, otherwise it is equal to the original price of the bid.

[8] Notice that we are simplifying here, since these are not all the possible purposes. In fact, in Chapter 5 we will specify that there are also other ancillary services, namely secondary reserve, start-up, change of configuration (denoted, respectively, by RS, ACC, CA).

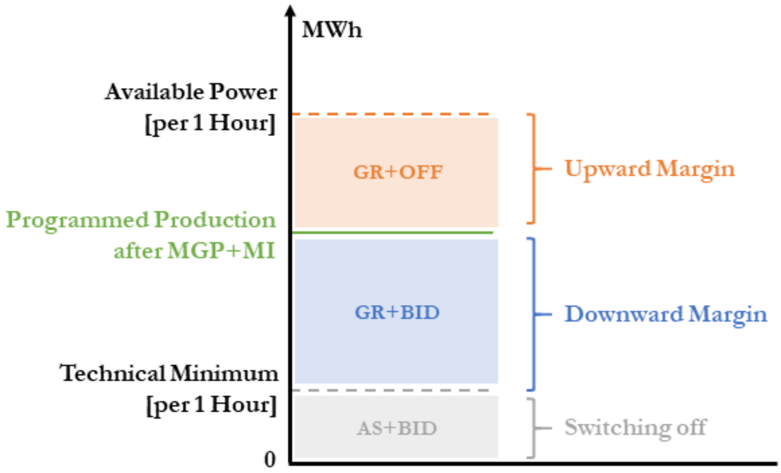


Figure 3.6: MSD ex-ante: visualization of **meaningful bids** (quantities), for a PU with positive programmed production after MI

- GR + OFF := *Step* bid to increase the amount of energy produced. They cover all the operating range, from the current operating point to the maximal power
- GR + BID := *Step* bid to decrease the amount of energy produced. They cover all the operating range, from the current operating point to the minimal power

"AS" bids (from the Italian *Accensione o Spegnimento*) are those related to the overall functioning of the PU, with Terna asking each hour to be able to control (upon the correct payment, of course) if the unit is producing or not electricity in that hour. Alongside these bids, the PU is also requested to make "step" bids (in Italian *Gradino*, hence indicated as "GR"), both in upward and downward directions. These bids give the **TSO** the right flexibility to cope with any kind of reserve-related problems.

In Figure 3.6 we have reported a visualization of possible bids in MSD from a Production Unit with a positive programmed production level after MI. In general, we could expect "GR+OFF" bid prices to be **at least as high as the zonal price of electric energy** on MGP, and "GR+BID" bid prices to be at most equal to the zonal price on MGP. This is not compulsory, but enforces the idea that the average accepted bid on MSD has a more convenient price for the Production Unit, with respect to the one on MGP.
Furthermore, in order to guarantee the feasibility of every offer made by the PU, Terna imposes a set of constraints over the bids in MSD, called **convexity constraints**, which are mainly based on the outcomes of MI. These serve the purpose to make sure that downward step bids are considered

sequentially before "AS+BID" offers, since clearly a PU can not have an "AS+BID" bid accepted if it is currently not at its minimum working range. In practice, this means that the price of the switching off ("AS+BID") bid must be lower than that for the lowest "GR+BID" bid, which in turn must be positive (we recall that for "BID" offers, the prices are virtually paid by the PU to Terna).
Moreover, the same applies for the minimum working point, since the price of the "AS+OFF" bids must be lower than the lowest price for the upward step bids ("GR+OFF"), indicating that Terna has to have an advantage in lighting up the PU first, before increasing its production from the minimum working point to a certain different level.
Lastly, the "GR" bids must satisfy the convexity constraint represented by the fact that the greatest price for the "BID" offers must be lower than the lowest price for the "OFF" type of bid; this request is effectively useful only for the PUs, since it requires that they do not "lose money for free" at any hour.

To conclude, we report that after MSD the Production Units know their *binding programs*, and they will be obliged to adhere to them in the actual production stage. Any possible imbalance will be evaluated a posteriori, on the basis of the procured difference in load, according to a mechanism of dual price, which is designed to be explicitly disadvantageous for the units. This makes MSD the last market on which the unit can adjust favorably its production program, stressing the importance of the aforementioned market. In the following chapter, we will then specify our research question, diving more deeply into the Ancillary Services Market MSD.

Studying the Italian Ancillary Services Market (MSD) 4

The aim of this chapter is to introduce the research question considered in this thesis, which will deal with a statistical analysis of the Italian **Ancillary Services Market (MSD) results**, focusing mainly on the North zone. In Section 3.3.3 we have already introduced the aforementioned market, and we have started to describe its technicalities and, most importantly, its importance for the correct and secure operation of the Italian Power System.

In Section 4.1, we will proceed with a brief **overview of the market**, where we will expand the topic of MSD, tackling in particular all the consequences related to the ever increasing penetration of renewable resources in the electricity systems. By doing so we will present the rising operational problems in maintaining the system balance and security, as well as the increasing attention on ancillary services, which are generally procured through short-term competitive market mechanisms.

Finally, we will present all the relevant **research questions** that we are going to answer within this thesis, as well as the general context of our work, which was carried out as part of a collaboration between Politecnico di Milano and the research institute RSE.

4.1 Introducing the Ancillary Services Market

In the past years we have observed a steady increase in the diffusion of power plants based on **renewable resources**, with many major countries committed to obtain a minimum share of renewable energy within 2030[1]

---

[1] The **European Union** makes no exception to this, with a commitment to cut by at least 40% the greenhouse gas emissions with respect to the levels of 1990, and the target of a minimum share of 32% for renewable energy. To obtain an overview of all the 2030 climate & energy framework in EU, we refer to the European Commission website: https://ec.europa.eu/clima/policies/strategies/2030_en

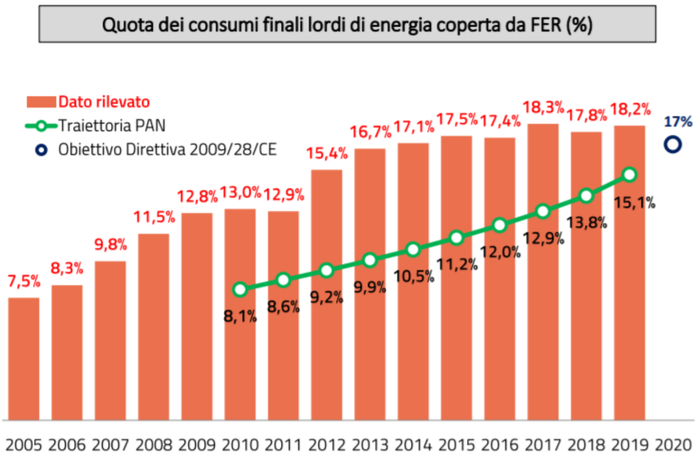


FIGURE 4.1: Percentage of load covered by Renewable Resources in Italy, in the years 2005 to 2019. Graph retrieved from [14]

and Italy adheres to this goal, as we can see from Figure 4.1.

These new infrastructures based on renewable resources are able to provide energy at a generally lower price than the ones based on traditional power sources, and they do so with considerably less harm to the environment.

However, they come with a greater cost for the management and general stability of electric Power Systems. This happens because renewable resources are generally of **non-programmable type**, meaning that we may have no control over the maximum output from a renewable power plant, for a specific hour of the day. This of course is linked to the unpredictability of the weather, which conditions, for instance, affect the maximum output of solar panels and wind turbines. Furthermore, many renewable production units present other problems for the stability of the grid, such as that of the absence of *mechanical inertia*, which we have already explained in Section 2.2.1.

All of the above observations imply the necessity to increase resources for **controlling the stability** of the electrical grid, which in the ancillary services market (MSD in Italy) are managed by the Transmission System Operator.

MSD has the primary scope to minimize imbalances on the power system and to guarantee its security in terms of power supply, meeting voltage and frequency qualities and congestion management. All these "services" are requested by Terna and are fundamental for the security and balance between supply and demand in the power system.

As mentioned above, ancillary services are acquiring more and more importance in the process of supporting the integration of renewable energy systems with the traditional ones. At the same time, conventional thermal power plants are facing a substantial drop of their contribution to the energy markets, with a consequent decrease of their economic margins[2].

Our work takes place in this framework, and is motivated by the importance to understand and to **forecast the electricity prices** of offers in the ancillary services market. Since MSD is a **pay-as-bid** market, and units participating in it are obliged to bid at any time, it becomes quite hard to predict the bids in this market, especially if we compare MSD to the more straightforward day-ahead market MGP.
Indeed, though MGP and MSD are in principle run separately, many operators play in both of them, which results in their bids on the last market being highly influenced by the behaviour on the first one.
Furthermore, the fact that all eligible operators are obliged to take part in every session of MSD, implies that their offer curves in this market will be generally characterized by a high variability, deriving from a somehow clear division between competitive bids and other bids presented with the specific desire not to be accepted.

With that being said, there is a second layer of complexity to add, derived by the fact that there exist **many types of ancillary services**, and given a fixed hour of the year, not every bid (nor unit) is qualified to compete for each of them. This implies that bids are not accepted according to a mere merit order (i.e. price) criterion, since Terna gives preference to the stability of the Power System, rather than to the economic convenience of the bids. To complicate the problem further, the TSO is **not required to specify** for which ancillary service it is accepting a specific bid in MSD, meaning that there is no official way for the units operating in the market to understand why an offer in MSD was accepted over another[3]. Overall, this complicates the definition of a "competitive" bid in MSD, since a PU operating in the market could potentially have made this bid with the idea to cope with different services.

To conclude, the **complexity** of the *Mercato dei Servizi di Dispacciamento* makes it a very hard problem to be studied with the help of statistics. At the same time, this becomes a strategic issue for all the actors partaking in it, with a specific mention for traditional thermoelectric power plants, which could gain a new strategic role by competing for the ancillary services and partly cope with their decreasing margins from MGP.

---

[2] For a global overview of the latest trends in the energy sector, we refer to the website of the International Energy Agency: https://www.iea.org/reports/

[3] It is interesting to notice how electricity operators are generally in disagreement with this decision, since it complicates a lot the process of price definition for bids on MSD. However, we can explain it in a purely "*Game Theory* fashion", by saying that in general the lack of information for the operators favors the final consumers, by lowering the prices of electric energy.

4.2 Studying bid profiles in MSD

The final scope of our work will be that of performing a statistical analysis of the ancillary services market MSD, **studying bid profiles** of the production units operating in it. This analysis was promoted jointly by the departments of Energy and Mathematics of **Politecnico di Milano**, together with the research institute **RSE**, which played a crucial role in providing the technical expertise to support our statistical analysis with more practical knowledge of the electricity markets.

In the beginning, Filippo Bovera (co-supervisor for this thesis and PhD from the Energy Department of PoliMi), proposed us a two-steps analysis to tackle the study of MSD. The first objective was to **investigate how external variables affected** the bids made on the market by some Production Units, considering variations in load forecasts, or calendar variables such as the temperature or the solar angle of incidence.
The second step was to **study the behaviour of Terna** (the Italian TSO) towards the ancillary services, with the objective of characterizing the underlying type of service for each accepted bid.

Once clarified the schemes regulating the ancillary services market and the operating modes by Terna, we could have two different goals. On one side, to replicate the optimization algorithm applied by Terna to *"solve"* the market (with prediction purposes over the expected outcomes), and, on the other side, to identify the best bidding strategies for the production units.

This first formulation is still very general and also quite ambitious, mainly because of the **lack of knowledge** around the acceptation mechanics for bids in MSD. Furthermore, the ever increasing penetration of renewable resources in the electricity systems, which is causing rising operational problems to MSD, is, as a matter of fact, only a quite recent issue.
In fact, there exist very **limited research publications** on the topic of the ancillary services market, with more analyses performed at a European level, but way less on the specific Italian market, which in itself is characterized by many peculiarities.

In this regard, we could mention the previous work performed by Professor Secchi with the thesis of two former Mathematical Engineering students, from which we managed to extract many useful ideas for our analysis. In particular, we recall the work of **Pietro Innocenti** [2], who organized a massive dataset for the study of MSD and then applied a Random Forest algorithm for the prediction of *opposite calls*[4].
Moreover, **Virginie Marchionni** [3] developed a new tool for multidimensional integrative analysis, named TAJIVE, and applied the tool to the study of the bids in MSD by the Combined Cycle Power Plants of two operators.

---

[4] An ***opposite call*** happens whenever a power reserve procured from a plant is immediately used in real time. This is in contrast with the general situation, since all the power reserve are usually just precautionary, hence not used in most cases.

Those thesis were helpful in studying two very specific questions about MSD, but still lacked a cross-sectional analysis of the **whole ancillary services markets**. Our thesis would like to build from the existing work and add some more general conclusions related to the whole set of units operating in a zone of the Italian MSD.

4.2.1 Refining the research question

As we have stated in the previous section, the preliminary goal was quite broad and ambitious. For this reason, we were glad to welcome the opportunity to develop the thesis as part of a **six-month internship** with a specialized team from RSE[5], who shared our aim to further explore the *Mercato dei Servizi di Dispacciamento*.
This was indeed an extremely good opportunity, as confirmed by the mission of RSE, which is to promote research programs in the electrical-energy sector, targeting the entire national electricity system.

While collaborating with RSE, we rephrased the research question in a more approachable way and **resized the problem**, focusing on a more specific analysis.
First of all, we decided to concentrate our analysis solely on the **North zone**, which (recalling Section 3.2) alone accounts for more than half of the Italian electrical load.
Secondly, we considered that each production unit submits bids for different kinds of service, and we decided to focus only on the one accounting for most of the energy exchanges, called **GR1**[6], that deals with upward and downward bids by the PUs. This of course was quite a limitation, but it was a necessary one, due to the complicated setting behind our research problem.

Eventually, we also chose to **investigate the bidding formulation** by power plants, without studying in depth the acceptance behaviour of the Transmission System Operator. We did so after a suggestion from the experts from RSE, who made us realize the actual complexity of a study in this direction, and that we therefore decided to leave for future researches.

Summing up:

---

[5] Including **Dario Siface**, **Silvia Canavese**, **Dario Piloni** and **Antonio Geracitano**, from the Development of Energy Systems Department of RSE.

[6] Units can make bids for different services in MSD, which are denoted by the labels "AS" and "GR" in the market, following what we have explained in Section 3.4. Focusing only on the "GR" (or *step*) type, units can make more than one bid, to refine bids refused in prior market sessions. What happens in practice, is that the vast majority of the accepted bids are of the GR1 type, with many operators that do not even make bids for the subsequent steps. To be even more precise, the energy exchange on the ex-ante MSD market stage in the years 2014-2019 made by GR1 bids accounted for around $78\% - 86\%$ of OFF-type bids and $64\% - 75\%$ of BID-type bids.

Our final aim will be to interpret and predict the price at which Production Units submit their bids[7] on MSD GR1, taking into account their scheduled production programs resulting from previous markets, their characteristics and the possible influence of external factors.

---

[7] In the following chapters we will sometimes use the term *offer* as a synonym for the more accurate term *bid*, since we are actually dealing with auctions.

Dataset building for Italian Electricity Markets

5

In this chapter we will provide a complete description of the **collection of data** from different sources and the final dataset to which we applied our algorithms. We will also explain how we constructed new customized variables that will be at the basis of the analyses conducted in Chapter 6.

Before diving more in-depth into the topics, a first clarification should be done about the software used. For the data gathering and the further analyses we relied on **R**, and all the collected datasets were stored in *.RData* format (134 different files, for a total of 7.4 GB).

Our datasets rely on information previously collected by Professor Bovera (especially for features from the structural class), on RSE databases (for market data), and on publicly accessible data sources such as GME [10], Terna [9] and Entso-e Transparency platform website [8], which cover most of the remaining variables.

In Chapter 4 we introduced the topic of our analysis and we stressed its complexity, due to the interconnection of MSD with other markets, the partially hidden mechanisms that drive the Transmission System Operator's decisions, the presence of many actors that play independently on the market and the unpredictability of external factors, such as the production of renewable energy systems. Since we **aim to forecast the offered price on the Ancillary Services Market**, we had to take into account all its dependencies and influences that refer to different sectors.

Indeed, the collected data can be divided into five macro classes: we first gathered a set of **Baseline Data**, which include few *calendar variables* that characterize our statistical units in the time, as we explain in Section 5.1.

In Section 5.2 we describe how we gathered all the relevant information for the **energy markets MGP, MI and MSD** from RSE databases, explaining how we organized the data and the final set of variables.

The third class, which is listed in Section 5.3, is labelled **"Structural Variables"** and provides information about technical characteristics of the

Power System, together with all the variables that affect its functionalities.

Section 5.4 is dedicated to **"Exogenous Data"** that include external variables that could influence the results of MSD, such as data related to weather conditions.

Eventually, we conclude the chapter in Section 5.5 with an in-depth description of the actors on the markets. As already explained, we decided to focus our attention on MSD bids by **Production Units in the Italian zone NORD** and in Section 5.5.2 we will show how we selected the set of relevant PUs to be considered in the analyses, together with a graphical overview of their technologies and of the operators in charge of them.

5.1 Baseline Data

In this section we begin by reporting baseline data, referring to them as *calendar variables*. These were constructed after we set the reference period for our analysis, as we have shown in Section 4.2.1, where we also defined the goal of our work. Indeed, price offers on MSD constitute our target variable and are univocally determined by date and time.

Since the final goal is to better understand today's bidding strategy of the production units and since there are many data sources available, we put our attention on a period of three years and we decided to exclude the most recent year 2020, due to its uniqueness caused by the COVID-19 pandemic situation. Therefore, our time interval spans from January 2017 to December 2019.
The relevant features for this set of data are listed below:

- **DATE**: date of the bid (YYYYMMDD, Year Month Day);
- **HOUR**: hour of the bid (e.g. The value 1 corresponds to 00:00 → 00:59);
- **WEEK_DAY**: day of the week for the bid (from Monday to Sunday);
- **IS_HOLIDAY**: boolean variable, 1 for public holidays and Sundays, 0 for working days;

5.2 Market Variables: MGP, MI, MSD

This section is dedicated to **endogenous data** of the energy markets.
Since our scope is to analyse bids submitted on the Italian Ancillary Services Market and, as explained in Chapter 3, MSD is strictly connected to MGP and MI, we decided to collect all bids presented in these three markets for the selected time span. Furthermore, even if we will only deal with the North zone for the analysis part, we decided to retrieve the information of all the Italian bidding zones and we will explain how we restricted the focus of our analysis in Section 5.5.

The data about energy markets are downloaded from RSE database[1] and, because of the high dimensions, were rearranged in different datasets depending on geographical zone and year. We got **90 datasets in total** and called them MGP_ZONE_YEAR, MI_ZONE_YEAR and MSD_ZONE_YEAR. The ZONE part of the name refers to the bidding zone for the offer, which is one of the geographical zones into which Italy is divided, as we already explained in Section 3.2 (see Figure 3.1). The YEAR part could be one of the years of the reference period, namely 2017, 2018 or 2019. This subdivision was necessary to handle the **high volume of data**[2].

We now list the selected variables, anticipating that this baseline will be the same for the three markets MGP, MI and MSD.

- **DATE**: date of the bid (YYYYMMDD, Year Month Day);
- **HOUR**: hour of the bid (1 corresponds to 00:00 → 00:59);
- **UNIT_REFERENCE**: coded name of the Production Unit that submits the bid;
- **OPERATOR**: name of the company that owns the Production Unit's plants or "Bilateral" in case of private agreement[3];
- **PURPOSE**: since in our analysis we are dealing with Production Units, the variable indicating the bid's purpose will assume the value "OFF" when the intent of the bid is to sell energy or "BID" if the unit would like to buy energy. For Consumption Units, the reverse holds. In short, "OFF" means that the bid is for an upward service, "BID" means that the bid is for a downward service;
- **STATUS**: it can assume only two values and it points out if the relative bid is accepted ("ACC") or rejected ("REJ") by the Transmission System Operator. Other values of the STATUS label ("REP" - replaced, "REV" - revoked, "INC" - inconsistent, "SUB" - submitted) are not considered here;
- **QUANTITY**: quantity of energy offered for the specific hour, it is expressed in MWh;
- **ADJUSTED_QUANTITY**: given a bid, the adjusted quantity is the actual offered quantity corrected by the TSO. Indeed Terna could accept a bid but, to guarantee the security of the power system, it could request a different amount with respect to the offered one [MWh];

---

[1] This was a convenient choice, since the data are publicly available also in [10] but in a completely impractical format, whereas RSE had a copy of them in a complete SQL database.

[2] To give a feeling of the dimension of our datasets, we report the number of lines for the NORTH zone for year 2019: in MGP we collected a total of 7173911 lines, in MI 6890225 and in MSD more than 9 millions.

[3] Which, as we will explain, is possible only for MGP bids.

- **AWARDED_QUANTITY**: it corresponds to the real quantity of energy to be exchanged by the PU, therefore remunerated, for an accepted bid [MWh];
- **PRICE**: original price of the bid; it is expressed in €/MWh;
- **AWARDED_PRICE**: it corresponds to the price at which an accepted bid is remunerated [€/MWh];
- **ID_GRID_SUPPLY_POINT**: it is a numerical code that indicates the "relevant *exchange point* associated to the unit bidding in the markets" [10]. Production Units that share the same supply point are those for which, regarding the dispatching activity, the difference in the power grid point where the energy injection/withdrawal exchanges are realized is negligible[4].

Having presented all the variables of the dataset, we give some more specifications about OPERATOR, PURPOSE and AWARDED_PRICE ones, especially regarding the values they span over.

As to MGP, producers have two possibilities to buy and sell energy: they can follow the procedure and rules of the GME's platform explained before, and in this case OPERATOR assumes the name of the owner company as its value, or enter into bilateral agreements, whose terms are not public, and in this case the variable OPERATOR is set to "Bilateral". In this latter case, even if the contracts are private, Terna needs to register the transactions to ensure the functioning of the system. Thus, when there is a bilateral energy exchange, it is represented by fictitious values on the market. This procedure, on one side, makes the public aware of the transaction and, on the other side, preserves the confidentiality of the agreement.

Regarding the PURPOSE variable, we should add that in MGP it can only assume the value "OFF", since we are considering Production Units.

Finally, considering the pricing mechanism in MGP (see Section 3.3.1), AWARDED _PRICE corresponds to the value of the PUN or to the zonal price and, for a given hour and zone, it will be the same for every bid of every PU.

During the seven sessions of the Intraday Market, Production Units can modify their energy programs determined in MGP. Unlike what happens in MGP, during MI producers do not have the possibility to stipulate private contracts, but could have the necessity to reduce the scheduled energy

---

[4] During our research work, we found it hard to further interpret this definition. Indeed, one might think that two units that are considered *close* on the electrical grid are also geographically near one to the other. This, however, is apparently false and it is common to find GSPs that include units from different Italian regions.
Nevertheless, the concepts of *Grid Supply Point* and *Relevant Exchange Point* appear to be of paramount importance to understand the acceptation of offers by Terna on MSD, as we will highlight in the conclusions of this work.

injection. Therefore the variable OPERATOR can not be "Bilateral" and PURPOSE assumes both values BID/OFF.

5.2.1 The peculiarities of MSD

Eventually we proceed by specifying the **peculiarities of MSD (ex-ante)**, which deals with bids finalized to congestion resolution and the creation of power reserve. The variables are the same listed above for the Day-Ahead market and all the considerations given for MI still apply in this case. However, we have to take into account that the price is established according to a "Pay-as-bid" mechanism, thus AWARDED_PRICE is equal to PRICE and it is applied to the accepted quantity. Furthermore, we find two additional variables in MSD datasets:

- **SCOPE**: indicates the type of service offered;
- **ADJUSTED_PRICE**: it is the price corrected by the TSO to meet the constraints of the market when they are not respected, otherwise it is equal to PRICE [€/MWh];

The SCOPE variable can assume values "AS", "GR1", "GR2", "GR3", "GR4" and the PUs must present at most one bid for each type of scope, in each direction (i.e. for each value of PURPOSE). The meanings of the acronyms depend on the value ("OFF"/"BID") of the PURPOSE variable, as already explained in Section 3.4.
If the purpose is "OFF" (that is the producer is bidding to increase the amount of energy injection), AS stands for switching on the plant and start producing, while GR1 corresponds to a first increase in energy production. This can be accepted by Terna if the plant is already active after MI or if Terna has accepted also the "AS+OFF" bid. Furthermore GR2, GR3 and GR4 symbolize a second, a third and a fourth augment of production, whose acceptability is subjected to the acceptance of all the previous step bids.
If the offer purpose is "BID" (i.e. the Production Unit is bidding to reduce its production), the meanings are reversed and AS stands for the possibility to switch off the plant.

The variable ADJUSTED_PRICE [€/MWh] refers to the corrected price and it will be our **reference variable** for the analysis in the following chapters.

In the last part of this section, we would like to point out another issue: on MSD there are also rules that do not allow to submit bids at any price, since there exists an **upper bound of 3000 €/MWh** [13]. This threshold is actually reached only by a negligible portion of "OFF" type bids on MSD and made us reflect on the actual meaning of such offers, since in the period of our analysis we observed that Terna never accepted bids with value higher

than 500€/MWh[5].
Discussing the issue with our tutors from RSE, we agreed that such offers are a **clear signal that the PU does not want to take part in the market MSD** (at least for the "OFF" purpose), even if, we recall, it is obliged to bid anyway. Therefore, we changed any value $\geq$ 500€/MWh to 500€/MWh during our analyses, thus obtaining smoother bid profiles of the units.

Eventually, we would like to describe also how we handled the few **missing values** in features related to the price. Remembering that it is compulsory to present hourly bids on MSD, *N.A.* values correspond to periods when the Production Unit is unable to operate on the market, such as during extraordinary maintenance periods. In this cases, the PUs do not participate in the market, thus we decided to replace the missing values with values having the same meaning. As explained above, for "OFF" bids we fixed them at the value 500 €/MWh, while for "BID" bids we used 0 €/MWh. Indeed, when the PUs are called to make bids to decrease their planned production (namely they buy energy from Terna) 0 €/MWh is the least attractive price for the TSO.

5.3 Structural Features

Structural features refer to variables that describe characteristics of the structure of the energy system. Some of these variables are obtained via forecast methods and we decided to include only the features that are available, i.e. known, to the Production Units **before** they present their bids on MSD.

- **LOAD FORECASTS**: a day-ahead forecast of the total load [MW], per market time unit, per bidding zone. It is forecasted and communicated at least two hours prior of the closure time of the day-ahead market in the bidding zone or at $d-1$, 12:00 in local time zone of the bidding zone.
  It represents the foreseen total amount of power requested by end users.
- **TOTAL GENERATION FORECASTS**: it is an estimate of the total scheduled net generation [MW] per bidding zone, per each market time unit of the following day;
- **GENERATION FORECASTS FROM RENEWABLE RESOURCES**: a forecast of wind and solar power net generation [MW] per bidding zone, per each market time unit of the following day. We

---

[5] To be precise, of the 117 PUs we selected for the statistical analyses (Section 5.5), only 63 made at least one bid greater or equal to 500€/MWh and the percentage of such bids on the whole dataset is around 4%. This is still a considerable number (in the order of tens of thousands), but only 2 bids were accepted by Terna among this group.

further distinguished between onshore and offshore wind generation forecast;

- **WATER RESERVOIR**: aggregated weekly average filling rate of all water reservoir and hydro storage plants [MWh/week] per bidding zone. The information is publicly available on the third working day following the week to which the information relates. It represents the potential energy associated to water basins;

- **COMMERCIAL EXCHANGE**: hourly scheduled commercial exchange [MWh] between Italy and TSOs of neighbouring countries (Austria, France, Corsica, Greece, Montenegro, Slovenia, Switzerland) and between the principal zones into which Italy is divided (North, Centre-North, Centre-South, South, Sicily, Sardinia). In our dataset we have introduced a column for each possible direction of the exchange, for example we have a column of data that represents the exchange North → Centre-North and one for Centre-North → North, being aware that when the program is scheduled for a direction, the reverse one receives a null value;

- **NET TRANSFER CAPACITY**: hourly day-ahead forecast NTC (Net Transfer Capacity) [MW] per direction between bidding zones. The Net Transfer Capacity is the maximum capacity for exchange of power between two areas, compatible with security standards applicable in both areas and taking into account the technical uncertainties on future network conditions.

5.4 Exogenous Variables

In this last part, we introduce what we called **exogenous variables** that are data extraneous to the electricity markets, but possibly correlated to it. They consist of weather features, which come from *METEORED Italia website* [17], and natural gas market data, published on the GME website [10]. Weather strongly affects the production of renewable energy plants and, as a consequence, it has an impact also on traditional power plants. Thus, we included data on temperature and wind speed (which are both strictly linked to solar and wind energy production), together with the atmospheric pressure, which is a more general indicator of good weather conditions. Data regarding the natural gas price have been introduced as well, because gas is the fuel of many traditional power plants, therefore it is plausible that the supplying cost could affect the energy price.

- **TEMPERATURE**: we downloaded the mean, minimum and maximum values of temperature [°C] registered in the city of Milan, which is assumed as the middle point of the Italian North bidding zone, and also accounts for the greatest portion of energy consumed;

- **WIND**: refers to raw wind speed data [km/h];
- **PRESSURE**: it is the atmospheric pressure [hPa];
- **NATURAL GAS SPOT PRICE**: it indicates the spot price of natural gas on the respective day-ahead market.

5.5 Production Units' Selection and Features

Electrical power stations have production systems that differ in dimension, installed capacity and technology, and they can be considered the real actors of the electricity markets. In addition, each market complies with specific participation rules and only PUs that can guarantee a power of at least 10 MW are allowed to take part in the Ancillary Services market[6].

We recall that in this thesis we will focus on the Italian bidding zone NORD[7], which accounts for Northern Italy and comprehends the regions of *Valle D'Aosta*, *Piemonte*, *Liguria*, *Lombardia*, *Trentino-Alto Adige*, *Veneto*, *Friuli-Venezia Giulia* and *Emilia-Romagna.*

With this consideration in mind, we began by selecting a **segment of peculiar Production Units** in the NORD zone that made offers in each year of the considered triennium.

The final dataset consisted in 117 Production Units and in Section 5.5.1 we proceed to explain the features collected for each of them.

Before doing so, we should mention that the final number of considered units is reduced quite a lot from that of the units operating in the Day Ahead Market (MGP), since in this market take part lots of smaller production units, that are not allowed to enter MSD due to their limited installed capacity or other structural reasons.

To **quantify the proportion of this analysis**, it is interesting to study the number of active units with at least one offer in a specific year, which we have reported in Table 5.1. From this table we can observe that the number of active PUs in the market grows a lot in the three years, reaching in 2019 a number that is more than the double of the units we have considered in the analysis.

Year	2017	2018	2019
Number of PUs	137	217	277

Table 5.1: Active Production Units in MSD, NORD Zone

---

[6] To be even more precise, the Production Units are divided into "Relevant PUs" and "Non-Relevant PUs", depending on their apparent power, and only the former ones are allowed to take part in MSD. For further details on the topic we refer to [15].

[7] This choice is conservative, but still rational, since this zone is by far the biggest one in the nation, accounting for more than half of the total electricity load during the selected period of the analysis (as shown in Section 3.2).

This might appear concerning at first, but this trend is easily explained by the intention of Terna to align to the European directives, allowing gradually the entrance of smaller production units in the Ancillary Services Market (MSD).
This said, those units affect only in minor part the dynamics of the market, since they do not have the ability to be accepted for big quantities of energy and will act as *market followers* in a *Game Theory fashion*, adapting their bidding strategies to those already in place by the biggest units.

5.5.1 Collecting relevant features for the selected PUs

For each of the **117 selected Production Units**, we proceeded to collect a set of relevant features from RSE databases and the ENTSO-e Transparency Platform [8]. The data describe properties of the plants mainly from a technical point of view, and are listed as follows:

- **`NAME`**: it is the name of the power plant;
- **`UNIT_REFERENCE`**: it is a coded name for the Production Unit and works as the linking variable between this dataset and those about energy markets.
  In some cases, power stations incorporate more Production Units that bid separately on the markets. Thus, differently from `NAME`, this variable distinguishes possible internal sections of each power plant;
- **`OPERATOR`**: it is a coded name for the owner company and corresponds to the variable we already introduced for market datasets;
- **`SOURCE`**: this variable indicates the power source and it can assume four values. *"Idroelettrica dispacciabile"* refers to those hydroelectric plants that are able to reply to a request of energy production of the Power System operator; *"Idroelettrica da pompaggi"* is the label of the hydroelectric Production Units that have a pumping system; *"Termoelettrica"* is used when the energy generation relies on a heat source; *"Rinnovabile"* indicates the PUs that use renewable energy (hence geothermic, wind and solar);
- **`TECHNOLOGY`**: given the source, this feature specifies the kind of technology used and follows a classification set by RSE.
  Hydroelectric plants, labeled with both *"Idrolettrica dispacciabile"* and *"Idroelettrica da pompaggi"*, could be of technology *"Bacino"*, *"Serbatoio"*, *"Puro"* and *"Asta Idrolettrica"*. The first label indicates those PUs that rely on a natural water reservoir as a basin, while the second one identifies the PUs that use artificial water reservoirs. The technology *"Puro"*, instead, is only used for Hydroelectric Pumped-Storage plants, and indicates that the natural water intake of the power plant is limited. Eventually, the last one labels complexes of

hydraulic interconnected systems, built on the same water flow and managed by the same company.
Thermoelectric plants could have combined cycle technology (called *"Ciclo Combinato"*) or *"Tradizionale"* that stands for steam systems. Ultimately, *"Idrico Fluente"*, the only type of technology for source *"Rinnovabile"* active on MSD, indicates those hydroeletric plants that utilize flowing water, instead of water basins;

- **`TYPE`**: this is a further specification of the technology based on the classification, founded on ENTSO-e Transparency Platform [8]. For example, in the case of wind farms it distinguishes on-shore wind farms (used for wind farms constructed on the mainland) from off-shore wind turbines (used for wind farms constructed in bodies of water, usually in the sea);
- **`INSTALLED_CAPACITY`**: it describes the maximum capacity that a productive system is designed to run at, in [MW];
- **`VOLTAGE`**: voltage level of the transmission line to which the PU is connected, measured in [V];
- **`PU_LOCATION`**: it is an ISTAT [16] code indicating the exact geographical location within Northern Italy.

Some new **customized features** were then constructed to characterize the Production Units from a business and economical point of view, mainly by aggregation of other already introduced variables. These new features will be crucial for the clustering analysis conducted in Chapter 6 and we will further explain their importance in that occasion.

- **`TOT_PROD`**: this variable provides the total production during the reference period [MWh]. It counts the production as an algebraic sum over all the accepted bids;
- **`TOT_MGP_OFF`**: it denotes the total income gained on the Day-ahead market through accepted bids of "OFF" type. It is calculated as the product between the accepted quantity and the price at which the bid is remunerated (i.e. PUN or zonal price);
- **`TOT_MI_OFF`**: it indicates the total income gained on the Intra-Day market through accepted bids of "OFF" type. It is computed as the product between the accepted quantity and the price at which the bid is remunerated (i.e. PUN or zonal price);
- **`TOT_MI_BID`**: it indicates the total amount of expenses on the Intra-Day market linked to accepted offers of "BID" type. It is computed as the product between the accepted quantity and the price at which the bid is remunerated (i.e. PUN or zonal price);

- **`TOT_MSD_OFF`**: it indicates the total income gained on the Ancillary Services market through accepted bids of "OFF" type. It is calculated as the product between the accepted quantity and the adjusted price;
- **`TOT_MSD_BID`**: it indicates the total amount of expenses on the Ancillary Services Market through accepted offers of "BID" type. It is calculated as the product between the accepted quantity and the adjusted price.

5.5.2 An overview of the the main characteristics of the selected PUs

Considering [7] and the suggestions coming from experts at RSE, we expected that the **technology** and **operator** variables would have had an important role in the process of the definition of bidding strategies on MSD and other electricity markets. For this reason, we decided to report an overview of these two variables, analysing their distribution across the 117 units selected for the analysis in the previous section, trying to understand the most common types of plant operating on MSD.

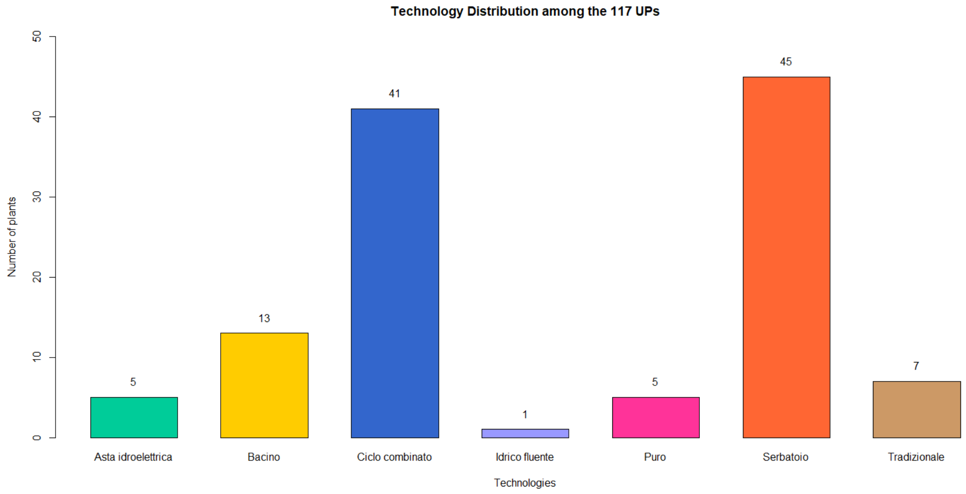


FIGURE 5.1: Distribution of the technologies among the selected PUs

In Figure 5.1, we can observe that the **main technologies** on MSD during the period 2017-2019 were **Combined cycle** (CCGT) and what we denoted *"Serbatoio"*, indicating an artificial hydroelectric reservoir.

As regarding the **owner companies**, the most widespread is ENEL, followed by Edison by a large margin (Figure 5.2). This is quite interesting to notice, and in the following sections we will see how the relevant position of the operator **ENEL** in the markets is reflected by its peculiar bidding strategies on MSD.

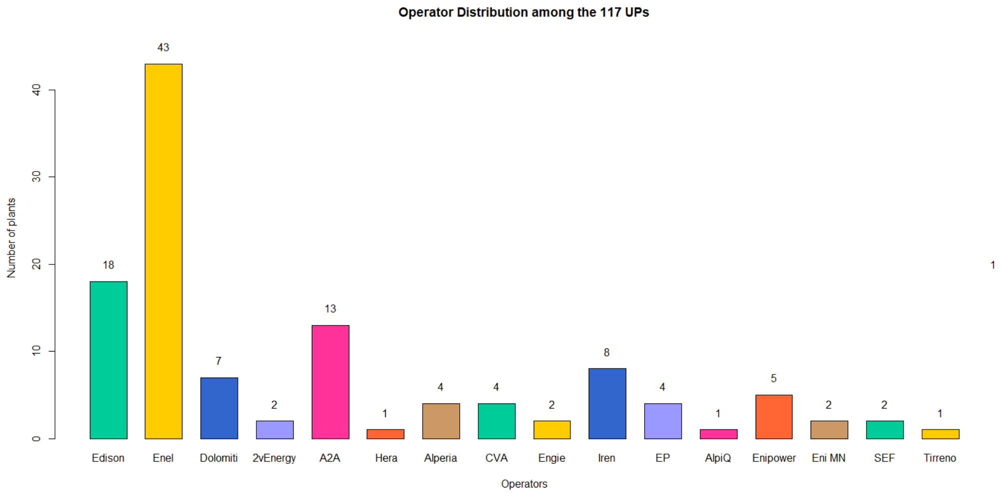


FIGURE 5.2: Distribution of the operators among the selected PUs

Clustering the macro-behaviour of Production Units in the North zone

6

The following chapter marks the beginning of the **statistical analyses on the Ancillary Services Market (MSD)** and has the objective to describe the second portion of the work I conducted during the internship at RSE, which me and Sara Martucci[1] have introduced in Chapter 4.

The first portion of the analyses was conducted by Sara Martucci and will be illustrated in her master's thesis [1] later on. Her work was focused on the implementation of a **functional clustering algorithm** with the aim to classify the PUs based on their bid profiles on MSD. This methodology provided us with some useful results but highlighted the difficulty to interpret and cluster the offer curves by Production Units, suggesting that a *broader approach* was necessary[2].

Nevertheless, due to the high mass of data, I felt the necessity to segment the power plants identified in Section 5.5 into smaller groups of units with similar behaviour on the electricity markets. Consequently, I decided to move from Martucci's methodology in favor of a slightly different one, which I will describe in Section 6.1 and is based on the **macro behaviour of the Production Units in the three considered markets**: MGP, MI and MSD.

The motivation behind this alternative approach lays in the idea that it is quite hard to *evaluate*[3] a single bid made by a PU at a certain hour of the year, whereas it is way easier to study the "general trend" of the outcome of bids across the three-year period. In these regards, the final aim will be to distinguish **peculiar long-term behaviours of the Production Units**,

---

[1] I recall, once again, that the introductory chapters 2, 3, 4 and 5 were written as a *four-handed* project together with Sara Martucci.

[2] This happens for many reasons that will be explained in [1], the most relevant one being the long time span chosen for our analyses, which imply a great difficulty in the definition of a powerful notion of distance in the space of the bids.

[3] By *evaluate* I mean in first instance to assess the competitiveness of the single bid on the market with respect to other bids (as explained in Chapter 4), but I also refer to the prediction of its outcome, in terms of acceptation from the TSO.

such as a particularly high activity for a certain type of bids, on a specific market.

In Section 6.2 I describe how I built the key variables for the aforementioned *novel* approach, together with a preliminary visual analysis of some of them. The final clusters were obtained using the **K-Means algorithm**, which, however, ended up showing some of the problems identified by Martucci in her work and therefore required a **second layer of interpretation directly on the offer profiles** of the PUs in each cluster. Indeed, the K-Means algorithm was beneficial in identifying certain patterns within the PUs, but the final clusters presented in Section 6.4 are the result of a further cleaning process that was done manually at a later time, also considering information over the **technology and managing operator** of the PUs.

In the following pages, I will describe in-depth the reasons behind the identification of each refined cluster, listing the units that compose it and highlighting its main peculiarities. By doing so, I will pose particular attention to the offer profiles on MSD of each cluster, since they are the final target of this work.
I also include in **Appendix A** a **detailed description of the adopted K-Means methodology**, including the specifics of the tuning process of the algorithm and the related unrefined results.

Eventually, Section 6.5 **concludes the chapter with a set of general considerations** on the approach used and the main clusters I have identified, posing the basis for the prediction analyses of Chapter 7.

6.1 Motivating the clustering on the macro-behaviour of the PUs

In Section 5.5 me and Sara Martucci have explained how we identified a set of 117 Production Units from Northern Italy, which have in common the eligibility on the Ancillary Services Market (MSD). Nevertheless, the high mass of data[4] linked to the bids of these PUs, make it quite impractical to perform a joint analysis of the offers from these units on MSD and imply the **necessity to organize the selected power plants into clusters** of units with similar characteristics[5].

As explained in the introduction of this chapter, the analyses conducted by Sara Martucci highlighted the difficulty to perform the clustering directly on the offer profiles on MSD of these production units (therefore in a *functional* way), implying that a different approach was required. In

---

[4] Indeed, for each Production Unit in our dataset we have 2 type of bids for every hour and for three years, resulting in a total of 52560 bids per PU.

[5] In these regards, the easiest option would be to simply segment the power plants based on their main structural characteristics, such as technology and operator. This procedure provides reasonable clusters in many cases, as we will see in the following sections, but it is arguably the best option, since it completely discards the bids done by the PUs.

the following lines I will motivate this *novel* approach, which in particular considered also the *behaviour*[6] of the units on other electricity markets.

The first thing to keep in mind is that the electricity markets **MGP, MI and MSD are not remunerated in the same way**. In fact, as explained in Chapter 3, MSD is effectively composed by pay-as-bid auctions, with accepted bids remunerated at the prices declared by the PUs. This implies that the prices of accepted offers in MSD are generally different than those on MGP[7], which instead are common to all the accepted bids in that market (i.e. equal every hour to the market-clearing price).
For this reason, supposing invariant the production costs across the markets[8], each unit qualified to operate in MSD would like to **maximise its production in this market for ancillary services**, in order to maximise its overall gains. This implies that a production plant could operate the tactic to hold part of its productive capacity on the first two markets, with the idea to make competitive bids in MSD and thus increase its overall profits in the case those offers end up being accepted by Terna.

However, what emerged from the preliminary analysis conducted by me and Sara Martucci is that different production units generally decide to operate in the electricity markets in **different ways**, with reasons that are often driven by their structural characteristics (such as the technology of their power plants, or their installed capacity) or by their geographical position in the power grid, which determines how well their are connected to the transmission lines and how well they can serve for Ancillary Services.

For instance, this implies that there is a **profound difference** in the behaviour of thermoelectric power plants and hydroelectric power plants, since units of the first type generally have a greater capacity, and are characterized by very long *ramps* to light up their turbines to full capacity when they are switched off. On the opposite, hydroelectric power plants have shorter reaction times and can adapt faster their working range. Furthermore, production costs for CCGT power stations are higher than those for renewable hydroelectric power plants.

Moreover, during the thesis work, it came clear that Terna tends to **"trust" more the biggest production units** (which, in general, are the thermoelectric power plants) for the Ancillary Services, accepting on average more offers from those power plants in MSD, further expanding the differences between these power plants and those fuelled by renewable sources. This might sound counter intuitive[9], but it is a direct consequence of the concept of Ancillary Service, as I have explained in Section 3.3.3, where I have described how all the offers accepted in this market are associated with

---

[6] As I will show, I considered the bids from the PUs on MGP and MI, as well as their overall results in terms of acceptation by the TSO.

[7] In particular, we usually observe higher OFF bids and lower BID bids on MSD.

[8] Which is, of course, a reasonable assumption if we fix a specific time of the year.

[9] E.g. Since thermoelectric power plants have higher production costs and are more harmful to the environment than renewable power plants.

the idea to cope with potential hazards occurring to the power system.

With all these considerations in mind, I thought that it would be more appropriate to cluster the selected units by also taking into account their **bids in the markets occurring prior than MSD**, as well as some **structural characteristics** of the PUs.
Unfortunately, it is unfeasible to perform a separate clustering for every hour considered in our time span, since this approach would imply more than 25000 iterations. At the same time, clustering the units based on only few specific hours could give inaccurate results, due to the high variability of the bids on MSD, which, for instance, are highly affected by maintenance periods of the PUs and other external factors.

What I have eventually thought, then, was to look at the problem from a broader perspective, switching the attention towards the observation of **general trends in the outcome of bids across the three markets**. By doing this, I fixed the three-year period 2017 to 2019 and looked at an *overview* of the earnings of the Production Units across this time span.
As I will show, the results of this approach were incredibly promising, but still implied a great manual effort[10] to clean the clusters given by the adopted algorithm K-Means.

6.2 Selecting the variables for the macro-behavioural clustering

Having introduced the reasons behind the macro-behavioural clustering, I move forward to the identification of a set of peculiar variables for all the units in consideration, in order to define the dataset for this statistical analysis. By doing so, I refer to Chapter 5 for a better explanation of the meaning of each variable.

I first retrieved the **structural characteristics** of the PUs (as from Section 5.5), then moved towards the definition of other variables describing their **macro-behaviour on the electricity markets**. These new features were mainly obtained by aggregating the economical outcomes (in terms of acceptance by the TSO) of the bids reported in our datasets, which I recall include the years 2017 to 2019.

The final table, in which each row was represented by one of the **117 units** operating in the Italian bidding zone `NORD`, was characterized by the following variables:

1. **Reference Code and Unit Name** := Two variables used to index the dataset, which were taken from the PU dataset and were helpful in referencing to each power plant.

---

[10] Which consisted mainly in a visual comparison of the price curves from the extracted clusters, with special mention to the bids in MSD, the final target of our study.

2. **Technology** := One of *Hydroelectric (Conventional, from ponds and water basins)*, *Hydroelectric (Pumped Storage)*, *Thermoelectric*, *Renewable (Run-of-the-river)*. This variable was adapted from the PUs' dataset and resulted very helpful in the interpretation of some results, even if it was not directly fed to the clustering algorithm.

3. **Installed Capacity** := It expresses the production capacity of the power plant, in MW. This variable was taken from the PU dataset and resulted being one of the most important variables in the clustering.

4. **Voltage Level of the Connection** := It quantifies the quality of the connection of the Production Unit to the power grid. A brief preliminary analysis of this variable highlighted how all the biggest and most relevant PUs are connected to the Italian grid at a high voltage level. Eventually, due to the low variability of this feature, I decided not to include it in the final cluster analysis, as I will explain later on.

5. **Total Production** := It is the total production of electrical energy by the PU in the three-year period of our analysis, across all considered markets. It corresponds to the variable `TOT_PROD` presented in Chapter 5.

6. **Total Income** := It is the net income from all the accepted offers on electricity markets in the three-year time span of the analysis (thus it does not take into account any production cost). It considers the three markets MGP, MI and MSD. Indeed, for each Production Unit j, I denote with O_j the set of all its OFF bids that were accepted in our time span, with B_j the set of the accepted BID bids, and I get the value of Total Income for the PU j as:

$$\begin{aligned}\texttt{TOT\_INCOME}^j = \sum_{i\in O^j} \texttt{ADJ\_PRICE}_i \cdot \texttt{ADJ\_QUANTITY}_i + \\ - \sum_{k\in B^j} |\texttt{ADJ\_PRICE}_k| \cdot \texttt{ADJ\_QUANTITY}_k\end{aligned}$$

7. **Total Income from MGP** := The portion of Total Income derived from the Day Ahead Market (MGP), thus via only upward OFF offers. It corresponds to the already introduced variable `TOT_MGP_OFF`.

8. **Total Income from MI** := The portion of Total Income derived from the Intraday Market (MI). Eventually, I have decided to divide this variable in two different ones: `TOT_MI_OFF` and `TOT_MI_BID`, as explained in Section 5.5.1.

9. **Total Income from MSD** := The portion of Total Income derived from the Ancillary Services Market (MSD). Eventually, I have decided to divide this variable in two different ones: `TOT_MSD_OFF` and `TOT_MSD_BID`, as explained in Section 5.5.1.

6.2.1 Preliminary analysis of the selected features

Before moving on to the actual clustering, I begin with a qualitative study of the above defined variables, with the objective to understand the significance of each of them.

The first issue I studied was the **relationship between the Total Production and the Total Income** for each of the 117 power plants operating in MSD.

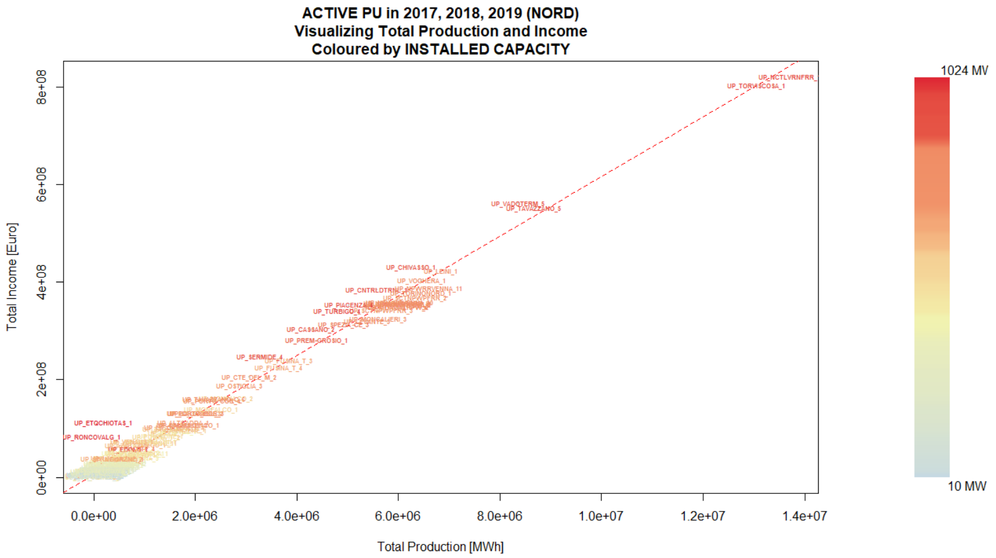


Figure 6.1: Total Production and Total Income of each PU, years 2017 to 2019, data coloured according to the Installed Capacity of the plant

In Figure 6.1 I have depicted the position of each Production Unit (via the variable `UNIT_REFERENCE`) with respect to the two variables **total income** and **total production**, with each power plant coloured on a gradient according to their installed capacity (the more dark red the name, the bigger the Production Unit).

It can be noted how there is almost a perfect **linear relation** between the two variables[11], identified with the dashed red line cutting the graph diagonally. Indeed, having fitted a simple linear regression model on **R**, I found that the angular coefficient of the estimated regression line is about 61 €/MWh, with the intercept being close to zero if compared to the order of magnitude of the data. This indicates that a singular MW of energy produced by a PU at a certain hour of the year, is on average remunerated at the price of 61€.

Indeed, there are mainly two reasons behind the high correlation of production and income: the first one being the fact that all the PUs make the **vast majority of their earnings** in the Day Ahead Market MGP,

---

[11] Actually, the correlation coefficient between the two variables is equal to 99.44%.

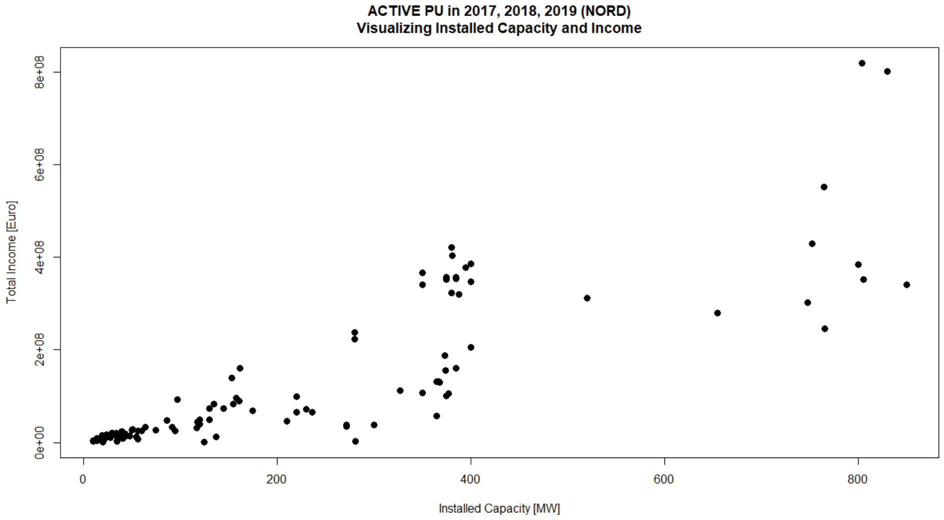


FIGURE 6.2: Installed Capacity and Total Income of each PU, years 2017 to 2019, NORD zone

which at a fixed hour is remunerated in the same way for all the active units. Secondly, since we observe this variable across three different years, we effectively eliminate almost all the inequalities in earnings caused by the periodic oscillation of the MGP zonal price[12].

In Figure 6.2 I have depicted the relation between the two variables **Installed Capacity and Income**. From that graph, we can clearly notice how the gross earnings done on the electricity markets increase with the dimension of the Production Unit, which is an expected behaviour.

Furthermore, it is interesting to notice that it also seems to exist a relation between the **"effectiveness"** of the bids by different PUs and their **installed capacity**, with the biggest units (in Figure 6.1 characterized by a darker red colour) that seem to be the most "effective" ones on average. It is necessary to specify that by most "effective" I denote the production unit that receives the greatest amount of money from Terna, within a group of units that produced the same fixed amount of energy in the considered time span. This concept only involves the gross revenues from the TSO and is very far from the actual earnings that the unit might make from its production activities, which instead will be highly related to other variables, such as the production technology. Indeed, from Figure 6.1 we can notice how the smaller production units lay below the estimated regression line, while the biggest ones are generally located above it, thus indicating that the biggest PUs tend to be more active on MSD and, more generally, in the

---

[12] To be even more precise, the weighted mean of the MGP zonal price by the requested load is equal to 57.895 €/MWh for our considered time span, which is slightly lower than the angular coefficient of the above estimated regression line. This is explained by the fact that we are only considering a sub-group of units eligible for MSD, which generally is characterized by much higher prices and lower volumes than the day-ahead market MGP.

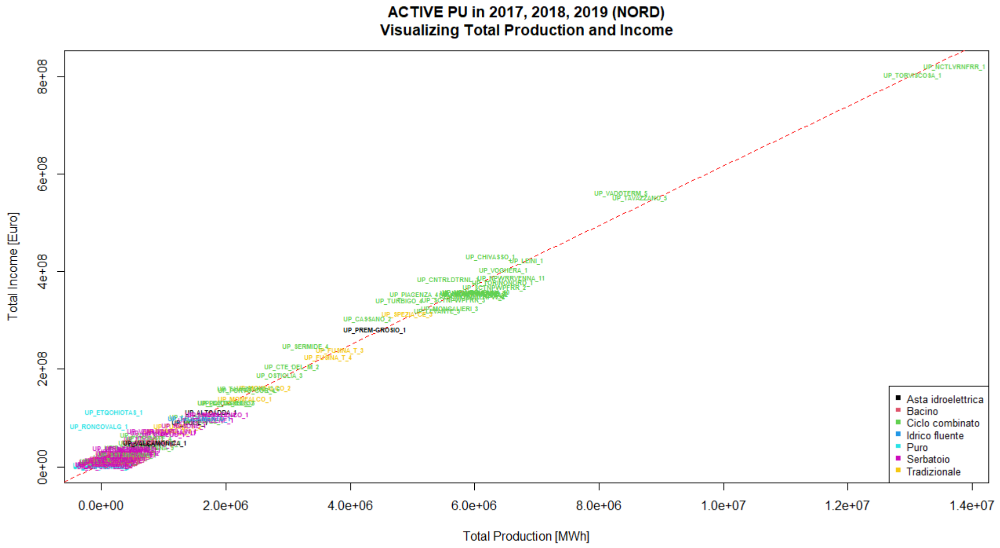


Figure 6.3: Total Production and Total Income of each PU, years 2017 to 2019, data coloured according to the Technology of the plant

hours where electricity prices are higher.

To further interpret the relationship between the total production and the total income, I have reported in Figure 6.3 the same graph, coloured by the **production technology** of the power plant.
This adds a second layer of interpretation to the previous considerations and it highlights how all the units that make the greatest gross earnings from the electricity markets are indeed the Thermoelectric ones, which dominate the upper-most part of the picture.

This plot is also helpful in the interpretation of the two outliers in the bottom-left part of Figure 6.1. In fact, these two units are **Pumped-Storage Hydroelectric power stations** of a very precise kind, labelled in our dataset as *Puro*[13]. In general, the behaviour of Pumped-Storage hydroelectric plants is quite different from that of the other units in the market, and later on in this chapter I will show how they are classified in a unique cluster also by the K-Means algorithm.

Regarding the remaining variables in the dataset, there are only few other results that are worth mentioning. One of them is the interdependence between the **Voltage Level of the connection** to the grid and the overall earnings from the markets, since all the biggest units are connected to the electrical grid at high Voltage Levels. This variable helped us explaining

---

[13] With the term *Puro*, which translates to "Pure" Pumped-Storage Hydropower, I denote a power plant for which the intake from natural resources that feed the upper water reservoir is on average less than 5% of the water volume *turbined* in a year of activity. More details on this classification can be found in [9].

the anomaly in the macro-behaviour of a group of Thermoelectric plants with limited activity on MSD, but was effectively a categorical one, since it presented about 95% of the data on only three different levels, that are specific to the Italian Power Grid. For this reason, I have decided not to include this variable in the final clustering analysis.

6.2.2 Preliminary analysis on the Equivalent Operating Hours of production

The last interesting topic before moving on to the identified clusters, was the idea to scale all the *aggregated* variables (such as, for instance, the Total Production) dividing them by the variable Installed Capacity, in order to obtain a representation that was coherent among all the units under investigation. This is a common practice in Energy Engineering, especially when dealing with production quantities, since it allows to obtain a new feature called **Equivalent Operating Hours (EOH)**, an indicator that has the same scale for every power plant, regardless its size.

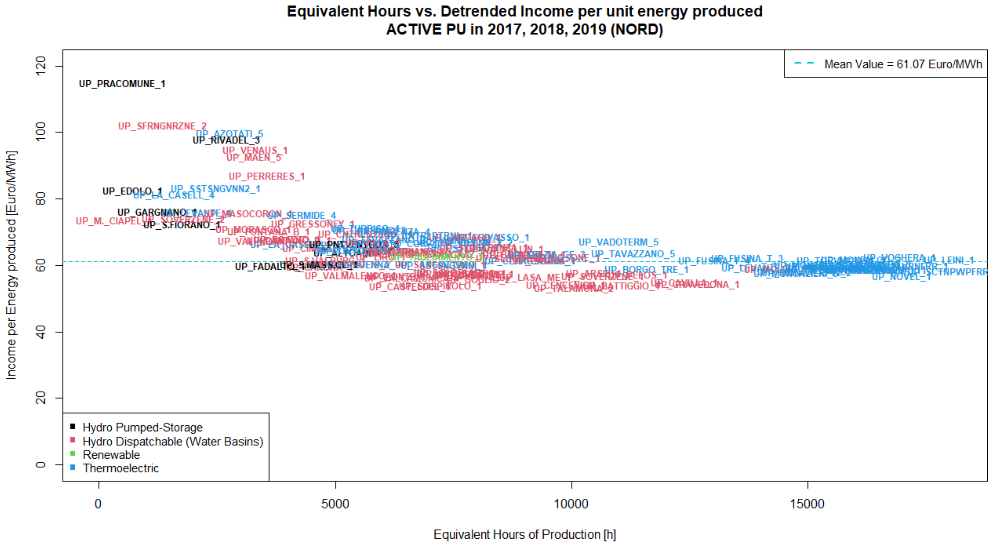


FIGURE 6.4: Equivalent Operating Hours vs. Income, coloured according to the technology of production

Just to give a brief overview of the new variable, I have reported in Figure 6.4 the **relationship between the EOH and the Income per unit of energy produced**, that is effectively a detrended version of the plots from Figures 6.1 and 6.3.

It can be seen that there is quite a strong division in the behaviours of Hydroelectric and Thermoelectric plants, with units from the second group that polarize towards high levels of Equivalent Operating Hours, while the

Hydroelectric units are generally less active, or at least tend not to be active at their maximum capacity.

It is interesting to notice how all the units that present a big amount of EOHs end up levelling their average price per unit of energy produced, which overall looks reasonable to believe, since they will end up working in the same hours and being remunerated mostly at the same prices, due to the rewarding mechanisms of MGP. On the opposite, **units that have lower values for the Equivalent Hours of production are able to work only in the best periods of the day** and are rewarded on average at a higher price per unit energy produced, if compared to the Thermoelectric plants located in the right-most part of this graph.

One last consideration regards the fact that **the most active units have around 16000 EOH** in the three-year period 2017 to 2019, out of a possible maximum of 26260 hours. This is quite staggering, and it becomes even more evident if we suppose that those units are always active at their maximum capacity, since this value imply that they would be active (at their maximum) for around 60% of the total available time.

In general, though, it is highly unlikely that a production unit is active at its maximum capacity, since on one hand the PUs tend to never reach their maximum as a result of the markets and on the other hand they generally decide to operate slightly below their maximum limit for safety concerns. However, even with this approximation, it means that there are some (mainly thermoelectric) units that are almost always active at very high levels.

6.3 Performing K-Means Clustering on the set of PUs active in the North zone

Having briefly studied the relevant variables, I proceed with the cluster analysis of the PUs, through many iterations of the K-Means algorithm.

However, since the very beginning of this work, it became clear how it would be difficult to interpret the results given by the algorithm without a **second layer of interpretation directly on the offer profiles** of the PUs in each cluster. This was not directly imputable to the selected algorithm, but more so because the **underlying problem is quite hard to study**[14], as explained in the introduction.

Eventually, I decided to proceed with the following seven variables, extracted from those identified in the previous section:

- **Equivalent Operating Hours** in the three-year period 2017 to 2019;
- **Total Net Income** on the three-year period, weighted by dividing for the Installed Capacity of the PU and measured in €/MW;

---

[14] For instance, even if it is reasonable to assume the existence of clusters within the selected units, we do not have prior information on their number. In this contest, the K-Means algorithm is mainly adopted as a tool to to identify relevant patterns in multiple dimensions.

- **Earnings from Day Ahead Market (MGP)** [€/MW], considering only the OFF bids (the only relevant ones for Production Units in this market) and dividing this value by the Installed Capacity of the PU;
- **Intraday Market Total Expenses from BID bids** [€/MW], divided by the Installed Capacity of the PU;
- **Intraday Market Total Earnings from OFF bids** [€/MW], divided by the Installed Capacity of the PU;
- **Expenses from Ancillary Services Market - BID bids** [€/MW], divided by the Installed Capacity of the PU;
- **Earnings from Ancillary Services Market - OFF bids** [€/MW], divided by the Installed Capacity of the PU .

In particular, I have found that it was extremely important to obtain comparable values of the selected variables for all the PUs in our dataset and therefore I have decided to weight all the relevant variables by the Installed Capacity of the power plant. By doing this I obtained a set of 5 variables related to the **total income per unit of capacity** (measured in €/MW), that were all referred to the same time span.
Notice also how this dataset does not contain any categorical variables, especially those indicating the technology of the power plant. This choice was made on purpose, with the idea to extract common patterns possibly across different technologies, in order to identify at a later time all crucial relations between the identified clusters and the production technology of units in the cluster.

In **Appendix A** of this work can be found all the specifics of the clustering algorithm, including the **tuning process** and the **clusters identified** in first instance. This said, in order not to loose to much focus from the core of this work, I have decided to present directly the manually refined clusters, in the following section.

6.4 Analysis of the identified clusters

Some of the groups identified by the K-Means algorithm were extremely interesting and allowed the recognition of some hidden patterns in the macro-behaviour of the PUs in the electricity markets. However, not every cluster was extremely clear, with some groups that did not present a distinct division from the others, and some units that behaved like outliers inside their reference cluster. This, though, was partly expected, since it was known from the preliminary analyses the difficulty of the underlying problem.

In this section, I will directly describe all the **major cleaned clusters**, which were obtained as a refinement of those identified by the K-Means

algorithm. By doing so, I will report the coded names of all the units in the cluster, together with an explanation of the main reasons behind the existence of the group.
I will also **focus on the bid profiles on the Ancillary Services Market (MSD)** of the units within each cluster, since this represents the goal of my research work and will also be the main topic of Chapter 7. This choice is also justified by the fact that MSD is not based on a merit order, thus its results alone are not sufficient to describe the behaviour of the PUs in it.

Concerning the precise results of the K-Means clustering, I remind that they can be found in **Appendix A** of this work, where I pinpoint the division in the original groups, interpreting in particular the position of clusters' centroids.

Here are listed all the major clusters I identified, in the same order they are reported in the following sections:

1. **Cluster of Thermoelectric Power Plants from ENIPOWER and S.E.F.**
2. **Cluster of Dispatchable Hydroelectric Power Plants from ENEL Produzione**
3. **Cluster of "ENEL to Alperia" Dispatchable Hydroelectric Production Units**
4. **Cluster of Thermoelectric Power Plants from ENEL Produzione**
5. **Cluster of Thermoelectric Power Plants active with bids of BID type**
6. **Cluster of Thermoelectric Power Plants active with OFF bids**
7. **Cluster of Pumped-Storage Hydroelectric Power Plants from ENEL Produzione**
8. **Cluster of Hydroelectric Power Plants from CVA - *Compagnia Valdostana delle Acque***
9. **Final Remarks over the remaining Production Units**

6.4.1 Cluster of Thermoelectric Power Plants from ENIPOWER and S.E.F.

UNIT REFERENCE CODES

- ENIPOWER (Mantova, Ferrera Erbognone [PV], Ravenna):
  `UP_NPWRMNTOVA_2, UP_NPWRMNTOVA_3,`
  `UP_NPWRFRRRRB_8, UP_NPWRFRRRRB_9,`
  `UP_NPWRRVENNA_9, UP_NPWRRVENNA_10,`
  `UP_NPWRRVENNA_11`
- S.E.F. - Società EniPower Ferrara (Ferrara):
  `UP_SCTNPWPFRR_2, UP_SCTNPWPFRR_3`

The first cluster I would like to introduce is by far one of the most interesting ones, and it is composed by a group of units from the two operators: **ENIPOWER S.P.A.** and **S.E.F. SRL**.
The cluster includes every unit from these two energy producers that are active on MSD, accounting for a total of 9 different Thermoelectric Power Plants of the type **Combined Cycle Gas Turbine (CCGT)**.

In the following paragraphs, I will explain how these units make very similar bids in MSD, even if their geographic position differs a lot, with units located in different provinces of the regions *Lombardia* and *Emilia-Romagna*.

Before moving on with our analysis, though, I have to specify that after a brief research on the internet, I found that the acronym S.E.F. stands for *Società EniPower Ferrara*, and therefore **all the PUs in this cluster probably share the same technical expertise**[15]. This should explain the similarities between the two operators, and confirms the goodness of our approach, since from solely the labels of our data it would have been impossible to infer a connection between the two operators.

This group of units was identified in first instance during the functional clustering analysis by Sara Martucci [1], who identified a cluster with **bid profiles on MSD that were extremely close one another**, as we can observe from Figures 6.6 and 6.7, included at the end of this section.
Furthermore, these units were also identified by our clustering algorithm K-Means[16], which highlighted how they all have in common the **macro-behaviour** across the other major electricity markets, in terms of accepted offers by GME.

---

[15] Intended as the team in charge for the definition of bids on electricity markets.

[16] To be more precise, the final K-Means algorithm divides these units in two distinctive groups, where it mixes PUs from ENIPOWER and S.E.F., as it is explained in **Appendix A** of this work. This, however, happens only in the final version of the clustering, where I have decided to increase the parameter K in order to obtain better overall results from the clustering, whereas in all the previous versions the units were part of a single and well-isolated cluster. All things considered, I have decided to stick with a unique cluster for these units, since I believe that it was the best choice for our purposes.

The Production Units in this cluster are **among the most active ones when considering the Equivalent Operating Hours** in the years 2017 to 2019, which means that their bids are usually very competitive, and a great portion of them is accepted by GME and Terna. The K-Means clustering also highlights the fact that units from ENIPOWER and S.E.F. are more active than the average on the Intraday Market (MI).
However, the peculiarity of this cluster is that it results in being by far the **most active one on the Ancillary Services Market (MSD)**, with high revenues from accepted bids of both downward and upward type.

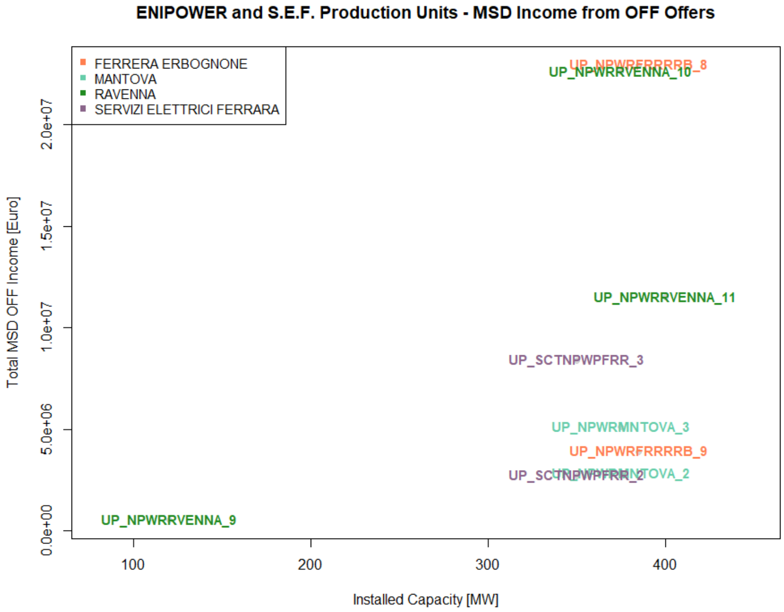


FIGURE 6.5: Cluster of ENIPOWER and S.E.F.: Total earnings from upward MSD bids (OFF type, years 2017 to 2019), against installed capacity

To better describe the units in the cluster, I have reported in Figure 6.5 an overview of these PUs, depicting how the capacity of each power plant affects the overall revenues from the activity with MSD bids of OFF type. The units are coloured by their geographical position, and we can appreciate how there is a pattern in the overall capacity of the identified units[17], since for the great majority of them is around 300 to 400 Mega Watts. The only exception is the center of *Ravenna* by ENIPOWER (coloured in dark green), which has three different units instead of two, with the last one having a capacity limited to 120 MW.

---

[17] This is probably a strategic decision by ENI, that prefers to own similar units in order to ease their management.

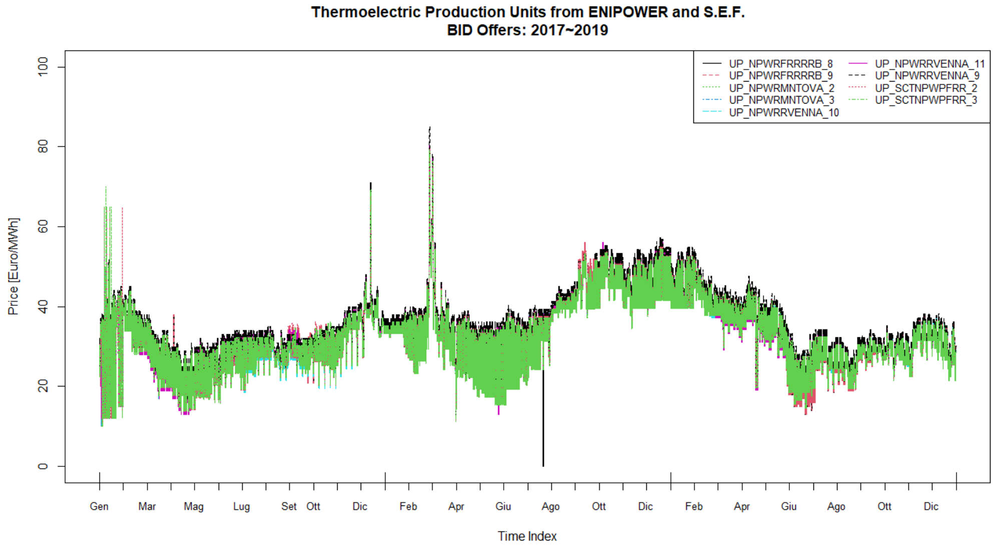


Figure 6.6: Cluster of ENIPOWER and S.E.F.: MSD GR1 bids of BID type

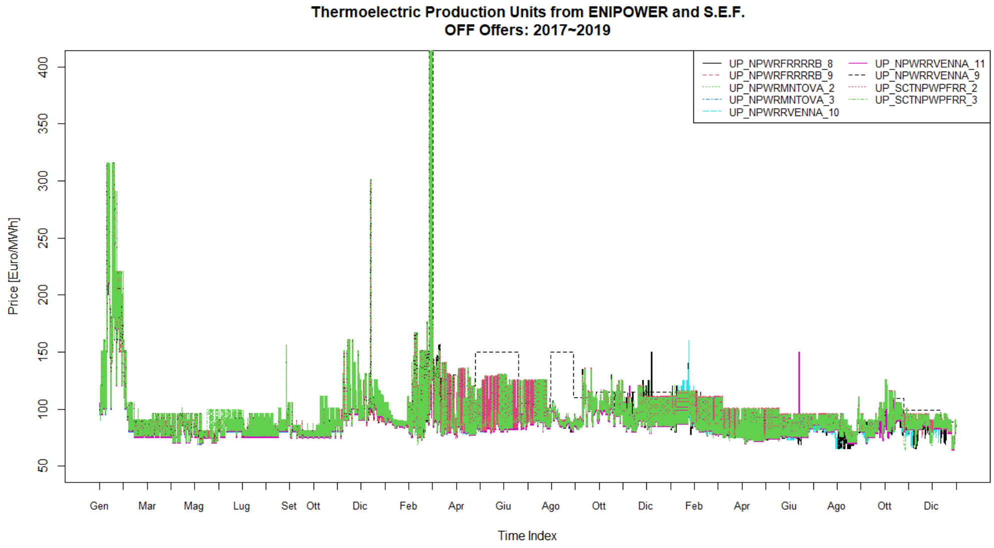


Figure 6.7: Cluster of ENIPOWER and S.E.F.: MSD GR1 bids of OFF type

As already stated, these group of units made very high revenues from upward bids in MSD, if compared to their average competitors. To better understand the reasons behind this fact, I have reported in Figures 6.6 and 6.7 their hourly bids on MSD for the years 2017 to 2019.

From those graphs we can appreciate how the bids from ENIPOWER and S.E.F. present an incredibly high **variability**, trying to adapt to the market's conditions on a hourly basis and also **following the trend of the zonal price on MGP**.

I will carry on the analysis of this cluster in Chapter 7, where I will fit with incredibly accurate results a Random Forest model to predict their MSD bids of upward and downward type.

In the following sections I will present the remaining clusters, and I anticipate that it will be clear how not all the competitors of ENIPOWER and S.E.F. bid on MSD like these two operators, since most of them present offers that appear to be *plateau* for many days (or even months) during the years 2017 to 2019.

6.4.2 Cluster of Dispatchable Hydroelectric Power Plants from ENEL Produzione

UNIT REFERENCE CODES

- ENEL PRODUZIONE:
  `UP_ARSIE_1, UP_CARONA_1, UP_CASTELDEL_1, UP_CAVILLA_1, UP_CENCENIGH_1, UP_GEROLA_1, UP_GOGLIO_2, UP_M._CIAPEL_1, UP_MORASCO_1, UP_PANTANO_D_1, UP_PELOS_1, UP_PONTE_1, UP_ROVESCA_1, UP_SFRNGNRZNE_2, UP_SOSPIROLO_1, UP_TALAMONA_2, UP_VALMALENCO_1`

The second cluster I would like to introduce is composed by a group of hydroelectric units from the operator ENEL PRODUZIONE and, as I will show, it has a very different behaviour from the first identified cluster.
ENEL is by far the biggest operator in the North zone, accounting for 36.75% of the active units in MSD[18]. This gives ENEL a **predominant position in the electricity markets**, enabling it to act as a *market leader* in a Game Theory fashion.
My work identified that **ENEL shares its expertise across all the units it controls**, and does so in a very explicit way, with tactics and bids that are usually very similar across power plants of the same Technology.

This group comprehends a set of **17 Hydroelectric Power Plants of Dispatchable type**, a production technology that indicates units operating with **water basins and dams**, excluding all the Pumped-Storage hydroelectric ones, that will be tackled later on in the analysis.
This cluster accounts for almost all the units from ENEL with the aforementioned technology, excluding only few units: two PUs from *Soverzene* that seem to operate as a couple in a slightly different way, the unit from *Venaus*, which has a very big installed capacity and operates in a more complex way in the markets, and finally a smaller group of 4 units that will be the topic of the following section.

Having introduced this cluster and its operator ENEL, I will now talk in details about the reasons why I have decided to report these units as a unique group. The first hint in this direction came from the algorithm K-Means, that identified a group of units of installed capacity smaller than the average, which were quite active on the **Intraday Market**, while reported **minor revenues from the Ancillary Services Market**. Visualizing graphically the bids from those units, I found incredible similarities in the curves of some of them, since they reported almost no activity for MSD bids in the downward direction, as depicted in Figure 6.8.

---

[18] With a total of 43 units out of the considered 117, **ENEL** is by far the biggest operator in the market of ancillary services. To put this in perspective, just consider that EDISON and A2A, the second and third operators by market size, only account for 13 and 12 units respectively, less than a third of those by ENEL.

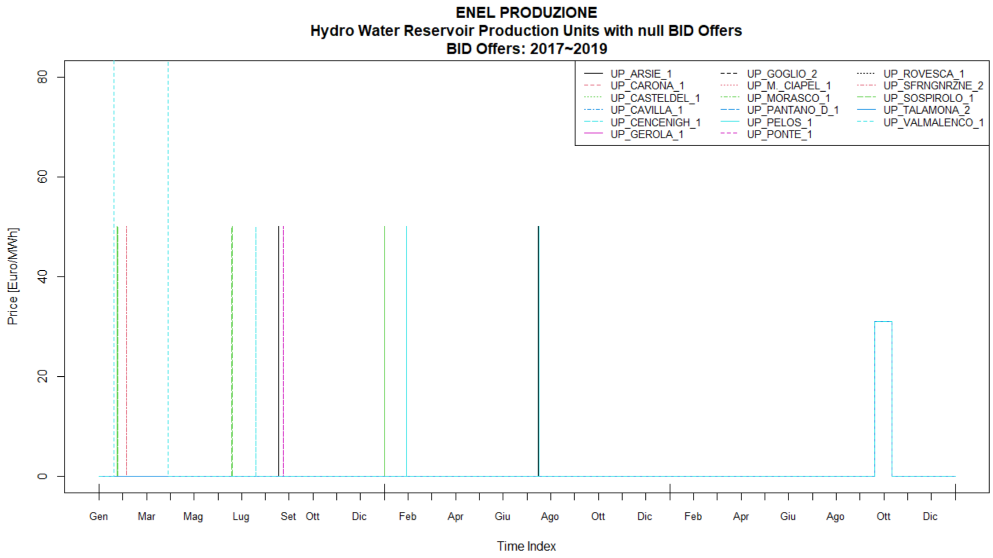


Figure 6.8: Cluster of Dispatchable Hydroelectric units belonging to ENEL PRODUZIONE: MSD GR1 bids of BID type

From this plot, we can clearly see that all these units present quite a peculiar behaviour for MSD bids of BID type, having a value that is constant to zero at almost every hour of the three-year period 2017 to 2019. It also emerges an anomaly from the 20^{th} of September 2019 to the 11^{th} of October 2019, when all the units fix their bids at the same value of 31 €/MWh. This behaviour is proper only to the units in this cluster, suggesting that ENEL adopts the **same pricing strategy** across most of its units of Dispatchable Hydroelectric type.

Regarding the value of the bids, we can trivially infer that fixing it to zero for almost all the MSD GR1 bids of BID type, implies that the **units are willing not to take part in the market**, at least for downward bids. In fact, BID offers are related to a request from the unit to decrease its energy production, effectively "buying back" energy from the TSO. Fixing to zero the price of this request, means that the unit would like to earn the money made in prior markets, yet without producing any amount of energy at all, and therefore without consuming its primary resources.

At first sight, this looks quite strange and sub-optimal in its own, since theoretically the best approach of each PU would be to try to optimize the overall gains from electricity markets, placing competitive bids in *all* the various sections, including MSD BID. This said, it is necessary to frame this consideration by saying that ENEL is the biggest operator in the market, and has in its own interests to **maximise its gains across *all* the units it owns**, with a particular mention to those of different technologies.
This observation could explain such a peculiar behaviour[19], which is definitely

---

[19] Indeed, this is only one of the many possible reasons, which are also quite hard to

not comparable to that of any other operator in MSD.

Nevertheless, from the K-Means algorithm it emerges that these units from ENEL are incredibly active in all the other markets, with a notable notion for MI, where they report the biggest gains with respect to the units' capacities. Considering this second point of view, it could be possible that the units from ENEL adopt a sub-optimal behaviour in MSD due to their major efforts in prior markets.

Looking again at Figure 6.8, it is impossible not to notice the **peak at the beginning of autumn 2019**, common to all the units in the cluster. I have questioned myself about the possible reasons for this anomaly, and came to the conclusion that it could be linked to one of the variables in our dataset: the **water reservoir in basins of Northern Italy**.

In Figure 6.9 I report on the same graph the curves of MSD BID offers

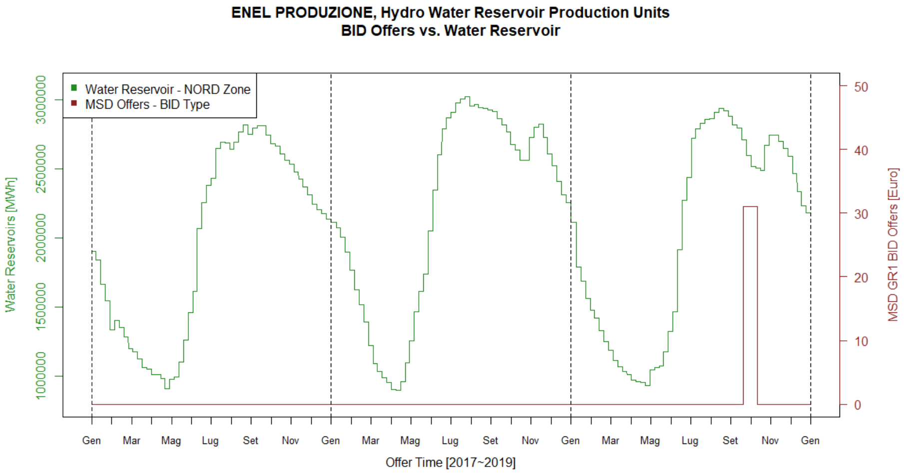


FIGURE 6.9: Cluster of Dispatchable Hydroelectric units belonging to ENEL PRODUZIONE: Comparing the peak in BID bids of autumn 2019 and the variable Water Reservoir.

from the considered cluster (coloured with a brown colour and with the related scale on the right part of the picture), together with the trend of the variable *Water Reservoir* (coloured in dark green and with the related scale on the left part of the picture). From this graph we can infer that there is a relation between the two curves, with the price for BID bids that increases during the autumnal trough of the Water Reservoirs in Northern Italy.

This could mean that in a period of scarce primary resources (at least compared to the expected value) for the considered units, ENEL decided to make more competitive downward bids, with the desire to be accepted more

---

formulate due to the known complexity of the electricity markets.

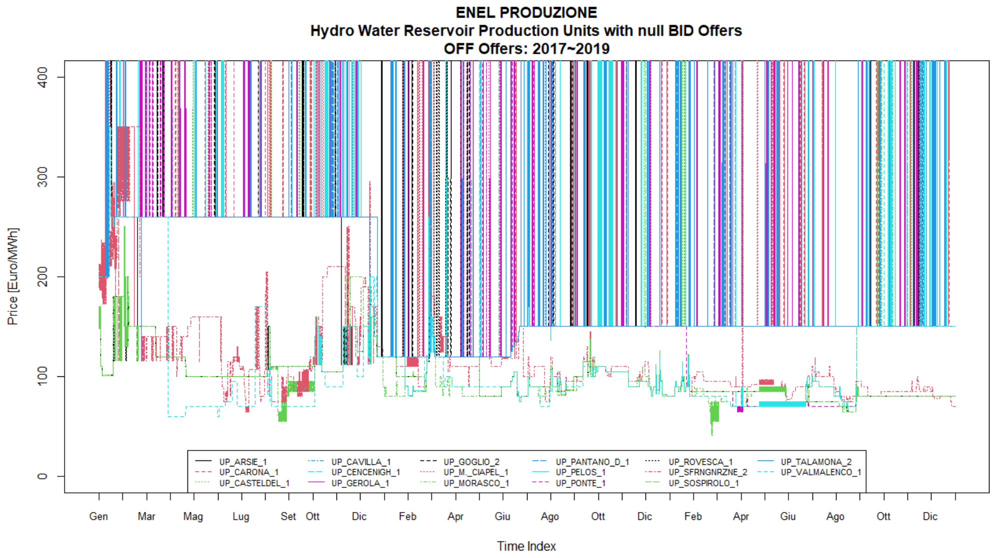


Figure 6.10: Cluster of Dispatchable Hydroelectric units belonging to ENEL PRODUZIONE: MSD GR1 bids of OFF type

often. Indeed, this tactic could be related to the desire to save water from being used in a shortage period, or, similarly, to recharge the water basins of the hydroelectric power plants.

Finally, I would like to analyse the **MSD bids of OFF type** from the units in this cluster, that are reported in Figure 6.10. The first thing that catches the eye is the biggest variability of these bids when compared to the BID offers of the same cluster. However, it is still possible to appreciate very clearly some interesting patterns inside the reported graph.
In particular, I have found that the units in the cluster could be divided into **two further clusters**, one with bids presenting huge steps alternated to constant values for long time, the other with more variable, as well as competitive, bids.
These two sub-clusters are depicted in Figure 6.11, where the two different strategies are reported in different pictures. From the graphical representation of these bids, it looks like that the bids from the second cluster (those reported in the lower part of the graph) are more competitive, since they are, on average, at a lower value and closer to the MGP zonal price. This idea is partially confirmed by the fact that these units made slightly greater revenues from MSD OFF bids. However, I could not find a practical reason for the ulterior division in the management of OFF bids by ENEL[20].

---

[20] Indeed, I discarded many options for this division, including: installed capacity of the power plants, voltage level of the connection to the grid and, finally, a geographical motivation. I also checked if the division could be linked to specific technologies of the power plants, but could not confirm nor reject this hypothesis from solely our data.

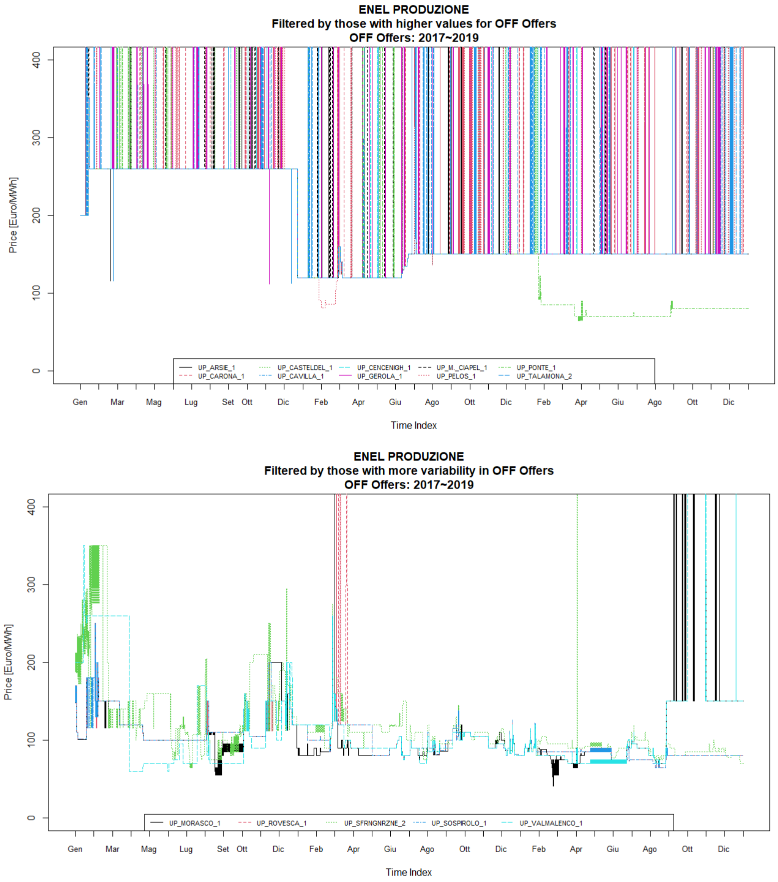


FIGURE 6.11: Cluster of Dispatchable Hydroelectric units belonging to ENEL PRODUZIONE: visualization of two sub-clusters for OFF Bids

In any case, the identified division is quite "fuzzy" and does not seem to hold in an extremely precise way, since, for instance, we can observe how two units from the second cluster seem to change their behaviour in the final months of 2019.

Eventually, I also found an interesting correlation related to the bids of the first sub-cluster, which I have reported in Figure 6.12. In the two reported pictures, I have included the MSD OFF curves for the units of *Carona* and *Pelos Nuova* (in the brown color), plotting on each graph also the programmed energy production after the market MI (dark green). This variable, which I shall refer to as **`MI_PROFILE`**, is evaluated hour-by-hour, by considering the signed sum of all the accepted bids of all type (OFF and BID), from the markets that precede MSD; therefore MGP and MI.

It is very common, in fact, that the **units consider their programmed production up to the already concluded markets** MGP and MI before placing their bids in MSD, and this looks exactly the case for this sub-cluster.
In particular, we can appreciate how the two reported units have the same

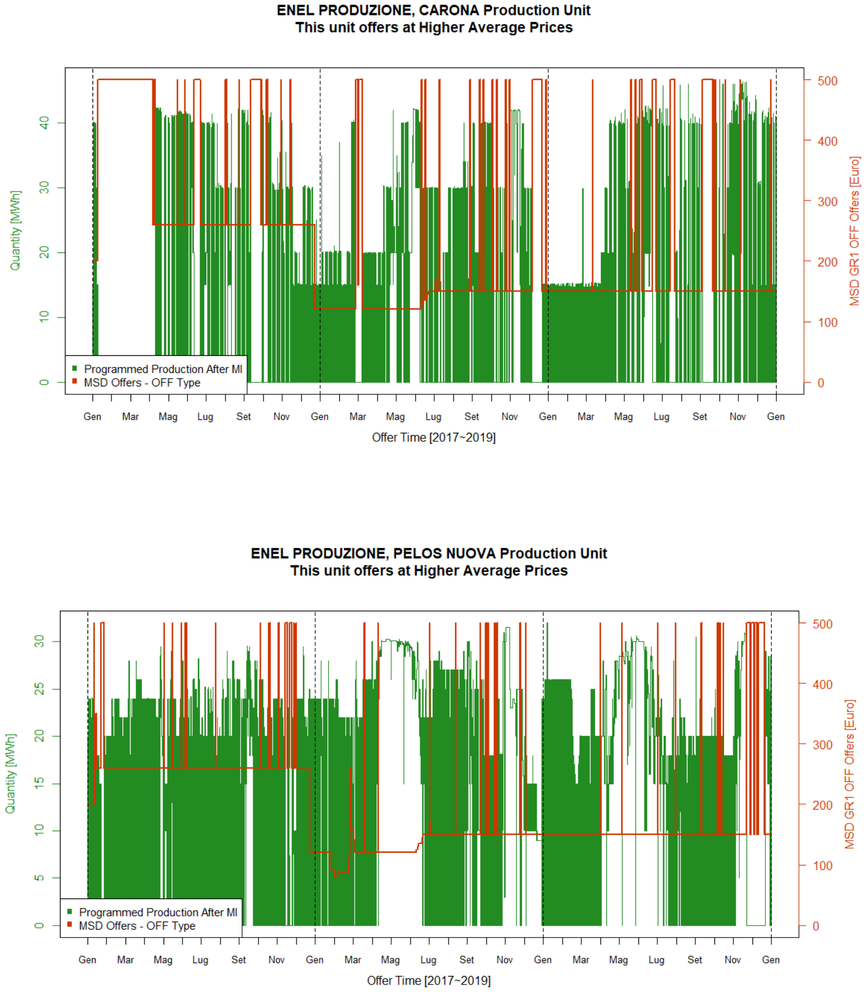


FIGURE 6.12: Production Unis of **Carona** and **Pelos Nuova** (ENEL PRODUZIONE), visualization of OFF Bids (dark brown) against their programmed production after MI (green)

"baseline" for their bids, from which they deviate only if they *leave* MI with a null programmed production, in that case they decide to shift upwards their bids to the value 500 €/MWh, which effectively means they would not like to start producing only in MSD.
Eventually, talking with experts from RSE, we also inferred that the **different values for the "baseline"** of bids in this sub cluster could be related to a concept known as *water value*, which the operators evaluate only at very specific times in the year, and it then remains unchanged for long times.

To conclude, I have identified a very **peculiar cluster of units from the operator ENEL**, that present quite similar bids on the MSD market. A preliminary study of this cluster highlighted few interesting topics, such as

the correlation between the peak in the MSD BID offers and the trough in the variable Water Reservoir, or, at a different level, the connection between the peaks in OFF bids and the programmed production of energy after MI, for a sub-group of units in the cluster.
Nevertheless, the **bids from this cluster result being quite "naive"**, if compared to the cluster of ENIPOWER and S.E.F., since they are often fixed at **integer values** that remains unchanged for long periods of time. Therefore, it seems reasonable to assume that those units do not change their offers according to a complex algorithm that accounts for many different variables, hour-by-hour.

For all these reasons, **I will not explore further the activity on MSD of this cluster**, maintaining the focus on those units with a more complex behaviour. In future analysis, though, it could be interesting to recover this cluster in order to study its behaviour on the markets MGP and MI, which, on the opposite, see these unit being extremely active, as identified by the K-Means algorithm.

6.4.3 Cluster of "ENEL to Alperia" Dispatchable Hydroelectric Production Units

UNIT REFERENCE CODES

- ENEL to ALPERIA: `UP_S.PANCRAZ_1`, `UP_FONTANA_B_1`, `UP_S.VALBURG_1`, `UP_PRACOMUNE_1`

As anticipated in the previous section, I did not include all the hydroelectric units from ENEL PRODUZIONE in the previous cluster, since I found that 4 of them presented quite a strange behaviour in their MSD bids, especially in the second half of 2019.

This consideration pushed me towards a very interesting discovery, that eventually lead towards the definition of a **cluster of 4 units that switched management from ENEL PRODUZIONE to the operator ALPERIA TRADING**, during the period of our analysis.
The geographical position of these units is depicted in Figure 6.13 and it corresponds to the *Val d'Ultimo*, in South Tyrol. Most interestingly, even if our dataset for Production Units indicates them as belonging to ENEL, I found that their offers on the electricity markets were effectively made by Alperia starting from March 2019.

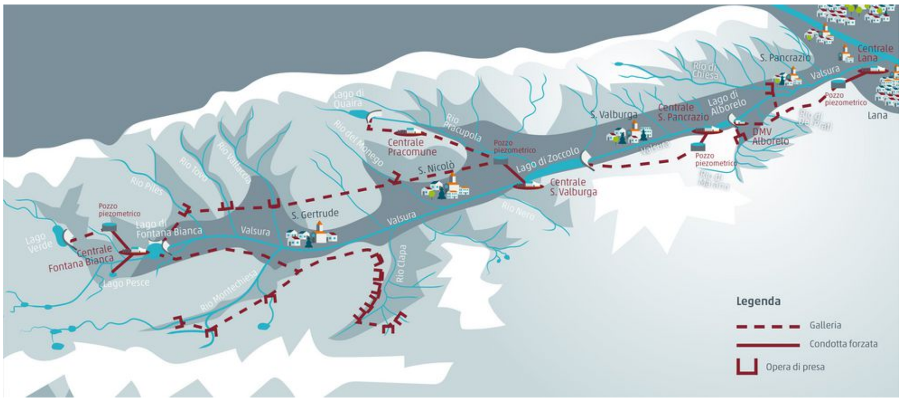


Figure 6.13: Alperia: Hydroelectric Power Plants in *Val d'Ultimo* (South Tyrol), image taken from **Alperiagroup** website.

Indeed, this observation could be extremely important for the research work on energy markets, since it allows to compare how two different operators decide to manage the same Production Units, which could be a very good indicator for the strategies of two different companies.

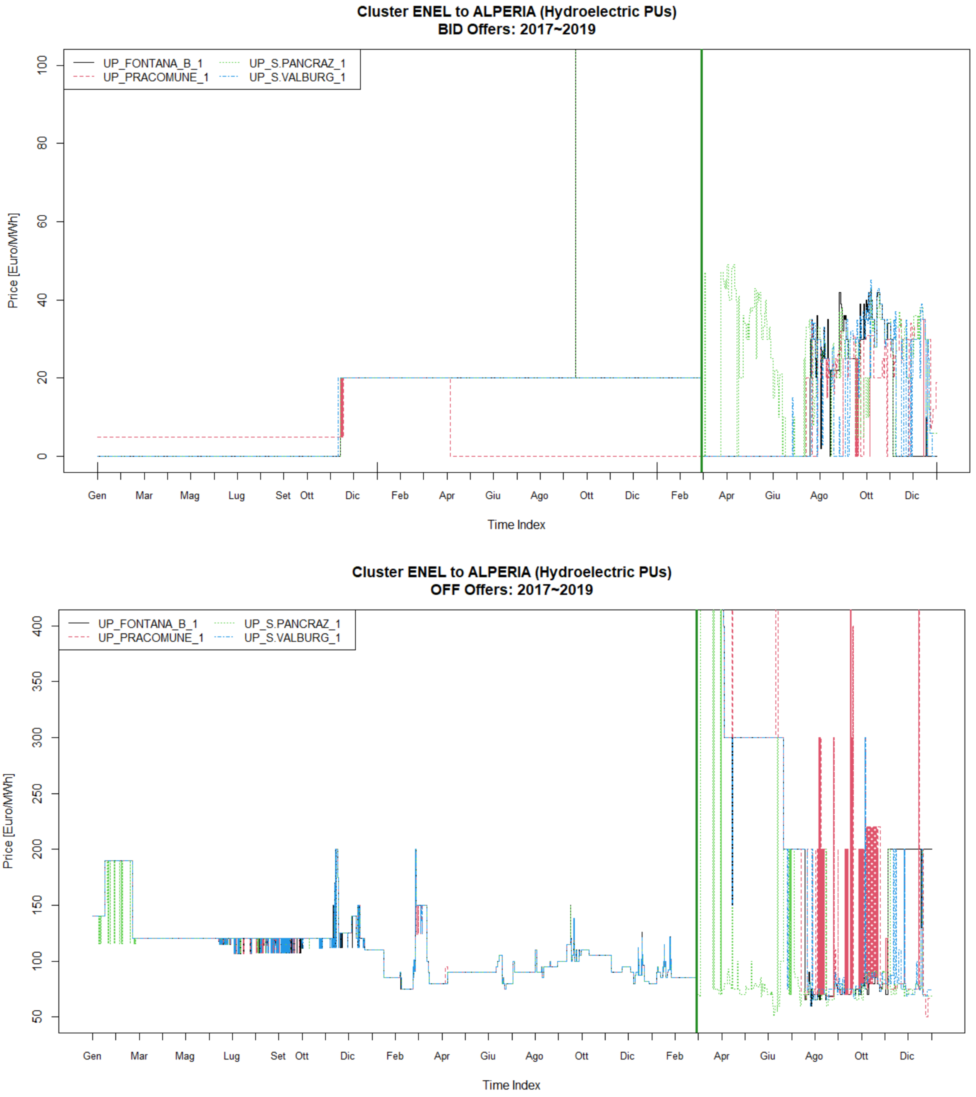


FIGURE 6.14: Cluster ENEL to Alperia, visualization of MSD bids of BID type (above) and OFF type (below). In both pictures the vertical dark green line marks the switch in management.

In Figure 6.14 I have reported the MSD BID and OFF bids from units in this cluster, highlighting with a vertical dark green line the **date of the change: the 1^{st} of March, 2019**.

From this picture, it is possible to see how the bids of these units in MSD under the operator ENEL are quite close to those of the previous cluster, with values that are *plateau* for long periods of time, and are generally common to all the units in the cluster.
On the opposite, after the green line and under the new management of Alperia, we can see the bids changing drastically, with a greater variability and a more visible differentiation of the units within the cluster.

In particular, we can appreciate that the **BID bids under Alperia are**

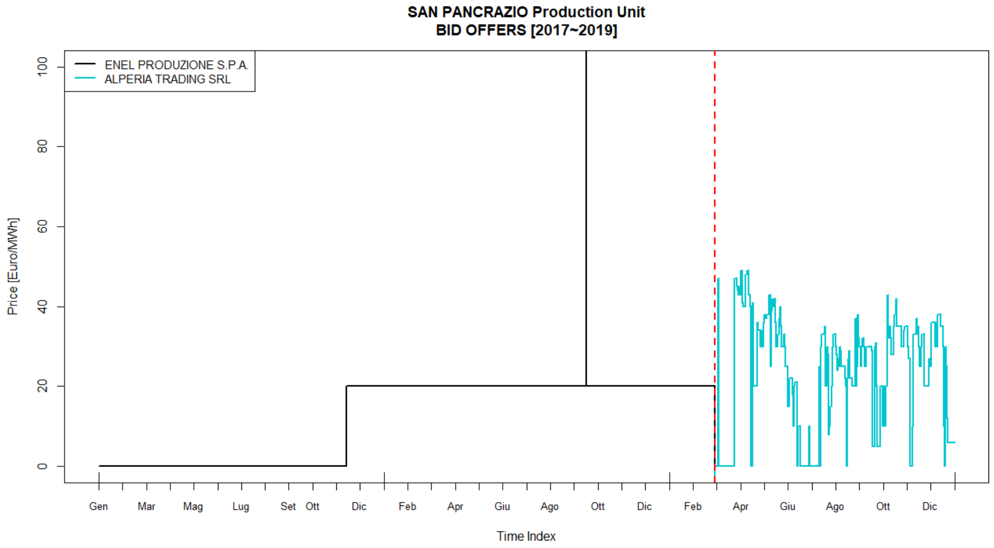


FIGURE 6.15: Cluster ENEL to ALPERIA, BID bids from the power unit of San Pancrazio (South Tyrol)

on average higher than those from ENEL, effectively implying **higher price competitiveness**[21].

Basically, the same principle applies for the upward OFF bids: the new operator Alperia introduces greater variability in them, but we can clearly see that on average the new bids have a lower price[22] than the ones from ENEL, effectively making them more appealing to the TSO.

To further support this hypothesis, I have reported in Figure 6.15 the curve of the BID offers from the unit of ***San Pancrazio***, which is a Dispatchable Hydroelectric power plant within the selected cluster. In this picture, the dashed red line divides the bids from ENEL (in black) to those from Alperia (in light blue) and we can appreciate a **great difference in the strategies from the two operators**.

It would be very interesting to study how the different strategies affect the **overall earnings** on the Ancillary Services Market, however this is not so trivial to do, since in the time span of this work we only observe few months with the new operator. Furthermore, the **earnings of hydroelectric power plants are highly affected by seasonality**, which

---

[21] I recall, once again, that BID bids are related to a request for a reduction in the energy produced by the PU, virtually corresponding with a remuneration from the PU to Terna. In this framework, higher BID bids imply that the unit is willing to give Terna more money if those are accepted, thus we can consider those offer as more attractive for the TSO.

[22] I remind that OFF bids work in the opposite of BID bids, since Terna is rewarding the power plants if those are accepted.

complicates a lot this task. This said, even a decision to extend the observation to the subsequent years could result in a failure, since during our analyses we observed how **Terna partially stopped the acceptation of the offers from smaller hydroelectric units in February 2019**, just before the management change of the units in this cluster[23].

To conclude, I believe that the group of units switching from ENEL to Alperia could be extremely useful for future analysis, since this cluster could be a **great indication for the different management choices** of the same Production Units, especially in a market characterized by high exchange in energy volumes such as the Day Ahead Market (MGP).
This, however, is quite far from the research question of this work, as I have decided to remain more focused on the MSD market. For this reason, I decided to treat these units as outliers, and I anticipate that I will not take them in consideration for the future predictive analysis.

---

[23] I will go back to this topic in the conclusions of this work, in Chapter 9.

6.4.4 Cluster of Thermoelectric Power Plants from ENEL Produzione

UNIT REFERENCE CODES

- ENEL, Thermoelectric power plants:
  `UP_FUSINA_T_1, UP_FUSINA_T_2, UP_FUSINA_T_3, UP_FUSINA_T_4, UP_LA_CASELL_1, UP_LA_CASELL_2, UP_LA_CASELL_3, UP_LA_CASELL_4, UP_PORTO_COR_3, UP_PORTO_COR_4, UP_SPEZIA_CE_3`

I now go back to the analysis of thermoelectric power plants, moving to a group of **11 PUs that belong to a total of 4 power plants controlled by ENEL PRODUZIONE**: *La Casella*, *Fusina* (also known as *Andrea Palladio*), *Porto Corsini* (also known as *Teodora*) and *La Spezia*.

This cluster includes all the thermoelectric power plants by ENEL in the dataset, and the algorithm K-Means originally classified them into 3 different clusters, mainly following their division into separate power plants[24].

All the units selected in this section have in common a **very high activity in the Intraday Market (MI)**, especially for bids of BID type. This is a similar trend to what was already shown for the Dispatchable Hydroelectric power plants by ENEL, and confirms the willingness of the operator to be considerably active in this market, in contrast to many of its competitors.

In Figure 6.16 I have reported the **offer curves on MSD** of the units in the cluster. Looking at the first plot, we can see once again that ENEL adopts similar strategies in the definition of downward bids by its Production Units. However, in contrast to what I have shown in Section 6.4.1 about the thermoelectric units from ENI, we can see that **the bids from ENEL present way less variability**, and for many PUs are fixed at stable values for long periods of time. This may be an indication of the absence of a complex algorithm that works at a hourly basis, at least for MSD offers of BID type.

Conversely, the second plot from Figure 6.16 presents much more variability than the upper one, with bids that are rarely invariant for more than a couple of days. However, even in this case, it looks like that there exists a common pattern in the offers, since the units exceed the value of 250 €/MWh only in the period between November 2017 and July 2019, where all the units increase the maximum value of their upward bids to 500 €/MWh.

---

[24] Indeed, the algorithm grouped as a single cluster the units from *La Casella* and *Porto Corsini*, while separated into two different clusters the units from the power station of *Fusina*. Eventually, after a manual analysis of all the thermoelectric units from ENEL, I have decided to include them in this section as a single cluster.

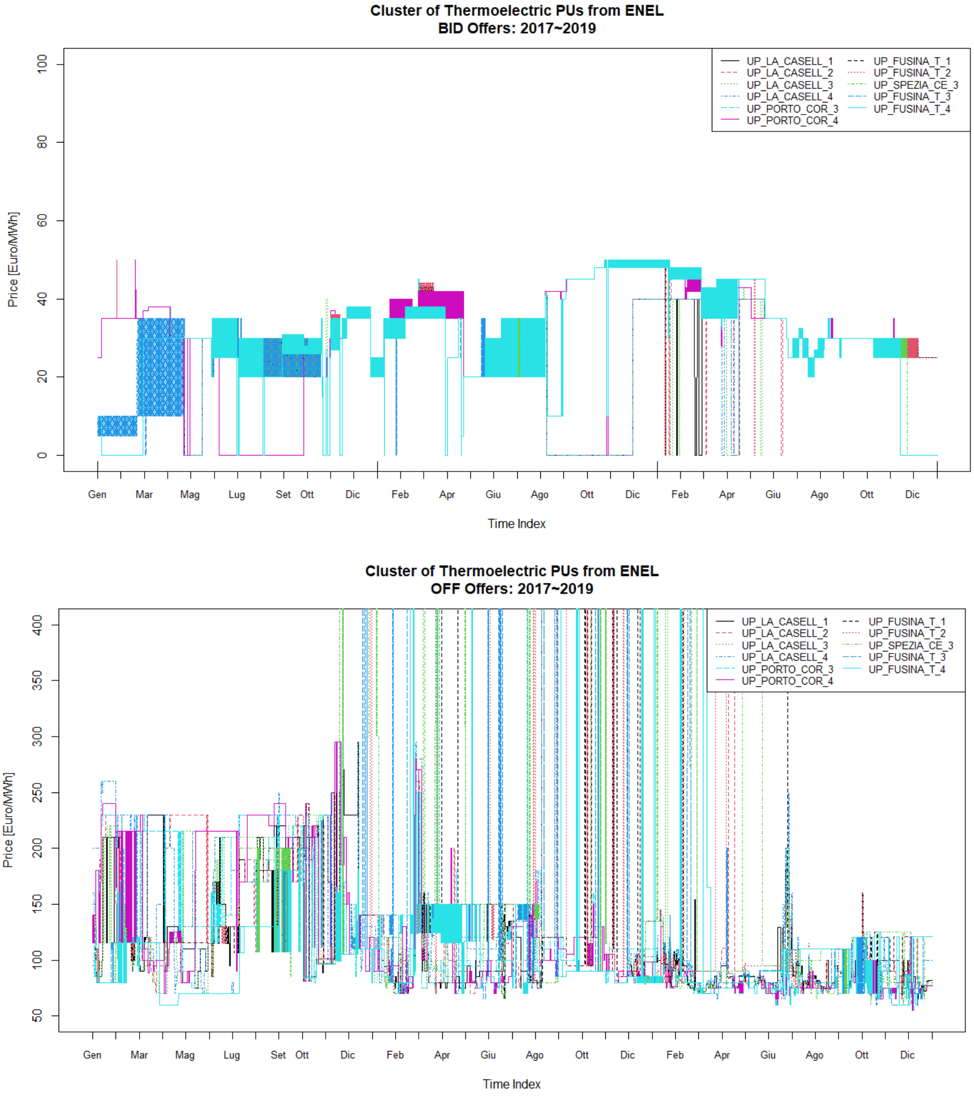


FIGURE 6.16: Cluster of Thermoelectric PUs from ENEL, visualization of MSD bids of BID type (above) and OFF type (below)

Having introduced the cluster, I would now like to study in details two of the power plants that compose it, namely *Porto Corsini* and *La Casella*.

With this purpose in mind, I have reported their bids in the plots of Figure 6.17 and Figure 6.18. From both of them it is quite interesting to notice how the **MSD bids of BID type** (those on average below the MGP zonal price, depicted in grey) are generally **equal across the PUs of a specific power plant**, whereas there is much more differentiation for OFF bids.

This confirms a trend already seen for other clusters of units by ENEL, who seems to define similar downward bids across production units of the same technology. Moreover, the **higher differentiation in the definition of upward bids**, could imply a major effort to place more competitive OFF

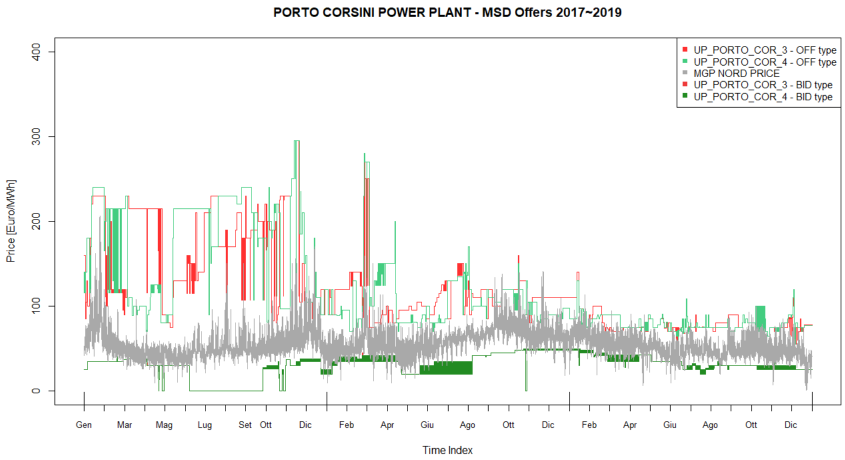


FIGURE 6.17: ENEL, Thermoelectric power plant of *Porto Corsini*: visualization of MSD bids in the years 2017 to 2019

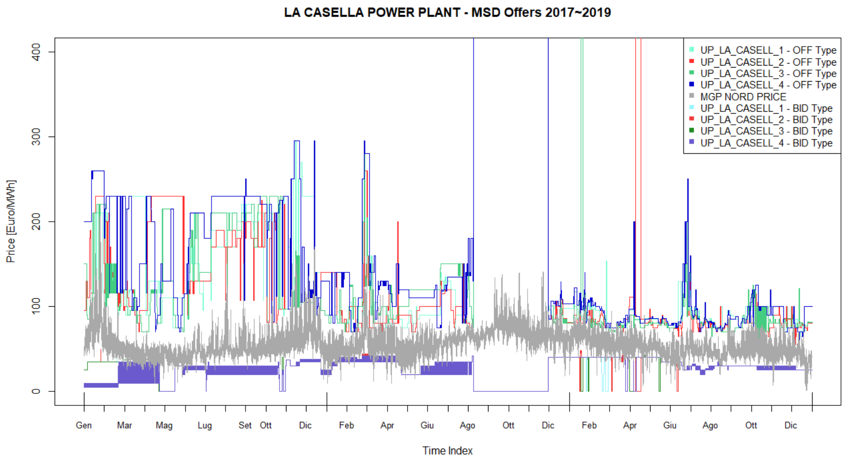


FIGURE 6.18: ENEL, Thermoelectric power plant of *La Casella*: visualization of MSD bids in the years 2017 to 2019

bids, which may be perceived as more important in order to increase the overall profits of the Production Unit.

Another interesting fact that we can appreciate from Figure 6.18 is related to the final months of the year 2018. In fact, **all the PUs from the power plant of *La Casella* seem to uniform their upward and downward offers to the same value**: 500 €/MWh for OFF bids and 0 €/MWh for BID bids. As me and Sara Martucci have explained in Section 5.2.1, these are very specific values, signifying that the Production Unit would rather to be rejected on MSD. Since this behaviour is prolonged between August and December 2018, we can infer with great confidence that the power plant was out of order in those months or, similarly, was

undergoing a period of major maintenance operations[25].
Most interestingly, though, it looks like that the maintenance period coincides with a change in the bids placed by the power plant. This is particularly clear for MSD OFF bids, which, starting from December 2018, seem to increase in variability and look on average closer to MGP zonal price, thus more competitive.

In conclusion, in this section I have illustrated broadly how the company ENEL operates its thermoelectric Production Units in Northern Italy, highlighting some **common strategies and recurring patterns across the offer curves**.
Eventually, in Chapter 7, I will analyse again this cluster, focusing on MSD bids placed by its units. In that occasion, I will pinpoint how the offers made on MSD by ENEL's thermoelectric units are quite peculiar, since they seem to be constant on a daily basis and defined without taking into account the hourly changes of the *Prezzo Unico Nazionale* on MGP.

---

[25] This is very common for thermoelectric power stations, due to the fact that these plants usually operate at very high temperatures, which may create stress in the materials composing their structure, requiring frequent controls and adjustments.

6.4.5 Cluster of Thermoelectric Power Plants active with bids of BID type

UNIT REFERENCE CODES

- Thermoelectric Units (BID bids):
  `UP_LEINI_1`, `UP_NCTLVRNFRR_1`, `UP_TAVAZZANO_5`, `UP_TORVISCOSA_1`, `UP_VADOTERM_5`, `UP_VOGHERA_1`

I will now introduce two additional clusters, identified by the algorithm K-Means and composed entirely by **Thermoelectric plants of CCGT production technology, from various operators**. These two clusters are quite interesting for the analysis of the Italian Ancillary Services Market, since they comprehend the units that were most active in MSD, respectively with BID and OFF bids. For this reason, I will present a brief overview of the two clusters in the following sections, but I will study them in greater details in Chapter 7.

I decided to name the first cluster **"Thermoelectric: MSD - BID"**, since it comprehends six units that reported the highest amount of money exchanged with the TSO Terna, for MSD bids of BID type, in the period of our analysis[26].

As anticipated, this cluster was originally identified by the algorithm K-Means, which grouped together the units of *Leini*, *Livorno Ferraris*, *Torviscosa*, *Vado Ligure* and *Voghera*, while left alone in a different cluster of a single observation the power plant of *Tavazzano*. Eventually, after a visual analysis of the bids presented by the units of the two original clusters, I decided to report them together in my final work, since they **presented lots of similarities**, as I will explain in the following paragraphs.

Unit Reference	**PU Name**	**Operator**	**Installed Capacity**
`UP_LEINI_1`	Leini	Engie Italia	380 MW
`UP_NCTLVRNFRR_1`	Livorno Ferraris	EP Produzione	804 MW
`UP_TAVAZZANO_5`	Tavazzano	EP Produzione	765 MW
`UP_TORVISCOSA_1`	Torviscosa	Edison	830 MW
`UP_VADOTERM_5`	Vado Ligure	Tirreno Power	760 MW
`UP_VOGHERA_1`	Voghera	Engie Italia	381 MW

Table 6.1: Cluster of Thermoelectric power plants active with BID bids: overview of the variables Operator and Installed Capacity

In Table 6.1 is reported **an overview of the PUs in the cluster**, including their operator and installed capacity. We can notice how all the

---

[26] I recall that the data used by the clustering algorithm were scaled by the installed capacity of each power plant, hence it would be more appropriate to state that the units in this cluster reported the highest amount of money earned *per MW of installed capacity*.

units in the cluster are quite big, with an installed capacity that ranges from 380 MW for the two units from *Engie Italia*, to the value of approximately 800 MW for the four remaining units, which in turn belong to three different operators. Geographically, the units are located in different parts of Northern Italy, inside the regions of *Liguria*, *Piemonte*, *Lombardia*, *Friuli-Venezia Giulia* and *Toscana*.

However, the clustering algorithm suggests that **all the units in this group have a similar behaviour on the electricity markets**, and, among other characteristics, their downward bids on MSD are accepted quite often by Terna, possibly implying that the prices of their offers are often very competitive.
Indeed, recalling that MSD bids of BID type are associated to a request made by the PU to "buy back" energy from Terna, this means that the units in this cluster are often willing to reduce the amount of energy to be produced, with attractive offers for the TSO.

I decided to pose particular attention to this cluster, since I found their behaviour to be quite peculiar, because it should involve the definition of more complicated tactics also on prior markets. In fact, the decision to **systematically renounce to produce great amount of energy** is rewarding only if the unit is precisely aware of its production costs, as well as the programmed production profile in hours close to the considered one.

Furthermore, it is quite reasonable to believe that a unit from this cluster should also take into account the amount of energy it agreed to produce via Bilateral contracts on MGP, prior to the decision to make competitive downward bids on MSD, in order to confirm the actual convenience of those bids.

To understand a possible reason for this peculiar result, I have reported

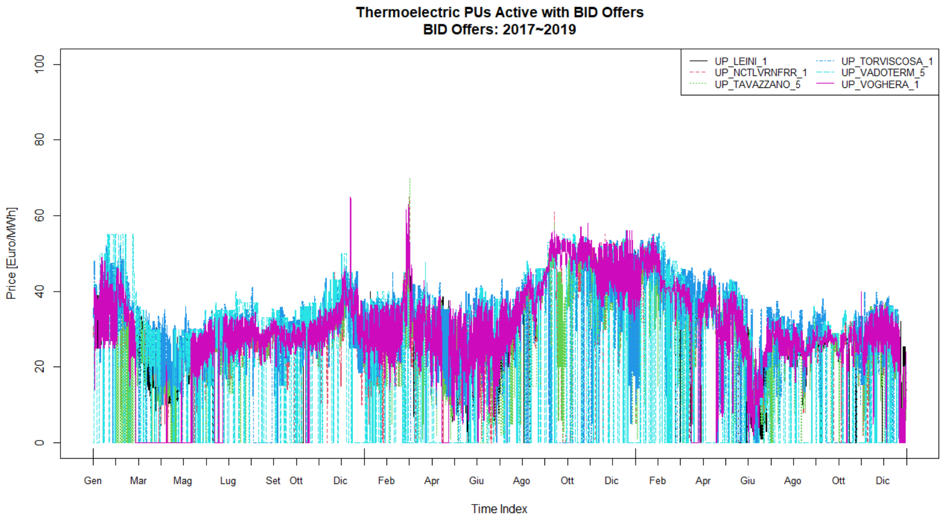


FIGURE 6.19: Cluster of Thermoelectric power plants active with BID bids: visualization of MSD bids of BID type

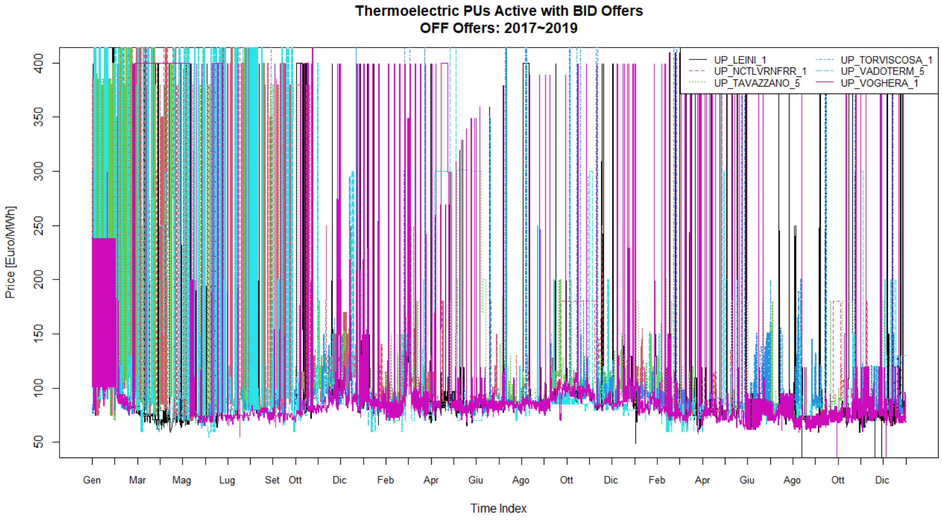


FIGURE 6.20: Cluster of Thermoelectric power plants active with BID bids: visualization of MSD bids of OFF type

in Figure 6.19 the MSD offers of BID type made by the units in the cluster. In that picture we can see how all the offers are visually quite similar one to another and they **tend to follow quite precisely the trend of the MGP Zonal Price**. This is a desired behaviour, since it confirms that all the units tend to place similar competitive bids of downward type.

Regarding the upward offers on MSD, I have reported them in Figure 6.20, from which it is quite clear that this kind of bids presents more variability than BID offers. In fact, it is quite obvious how the curves within this graph reach higher values with different frequencies, and are in general quite different one from the other.

To conclude, I would like to point out that in this cluster are grouped most of the units that can be found in the upper-right corner of Figures 6.1 and 6.3, meaning that **these units reported some of the highest gross revenues from the electricity markets** in the 3 years considered by our analysis.

This observation underlines the importance of an in depth study of this cluster, at least for BID bids, which are quite similar within the units composing the group. For all these reasons, I will return again on these power plants in Chapter 7.

6.4.6 Cluster of Thermoelectric Power Plants active with OFF bids

UNIT REFERENCE CODES

- Thermoelectric Units (OFF bids):
  `UP_AZOTATI_5, UP_CASSANO_2, UP_CNTRLDTRNL_1, UP_PIACENZA_4, UP_SERMIDE_4, UP_TURBIGO_4`

I will now introduce the second cluster of Thermoelectric power plants that were identified as active above the average on MSD, consisting in the group that I denoted with the label **"Thermoelectric: MSD - OFF"**.

This cluster comprehends a set of six different Production Units that earned **the greatest amount of money**[27] **on MSD through bids of OFF type**. Roughly speaking, it is reasonable to believe that these units make very competitive upward offers on MSD, which lead to an acceptance rate by Terna much higher than the average[28].

Unit Reference	PU Name	Operator	Installed Capacity
`UP_AZOTATI_5`	Azotati	Edison	236 MW
`UP_CASSANO_2`	Cassano	A2A	748 MW
`UP_CNTRLDTRNL_1`	Turano Lodigiano	Sorgenia	800 MW
`UP_PIACENZA_4`	Piacenza	A2A	806 MW
`UP_SERMIDE_4`	Sermide	A2A	766 MW
`UP_TURBIGO_4`	Turbigo	Iren Energia	850 MW

Table 6.2: Cluster of Thermoelectric power plants active with OFF bids: overview of the variables Operator and Installed Capacity

In Table 6.2 I have reported **an overview of the PUs in the cluster**, including information over their operator and installed capacity. It can be noticed how the most represented operator in this cluster is *A2A*, with three units, while the other PUs belong to *Edison*, *Sorgenia* and *Iren Energia*. It is also interesting to observe that three of the units are quite close to the city of Milan, which is the place with the highest request of energy in the whole bidding zone `NORD`.

Regarding the size, all the units in the cluster, except *Azotati*, have about 800 MW of installed capacity and result being some of the biggest power plants in the zone.

---

[27] As for the previous cluster, this holds true only if we remember that we are referring to money earned *per MW of installed capacity*.

[28] In Chapter 3 and Chapter 4, I have explained lengthy how it is not entirely true that the most price-competitive offer on MSD is always accepted by the TSO, due to the particular nature of this market. However, looking at this matter on the long term, we can reasonably assume that the PUs which earned the highest amount of money in MSD over the span of three years, were also those that placed, on average, the most convenient bids for Terna.

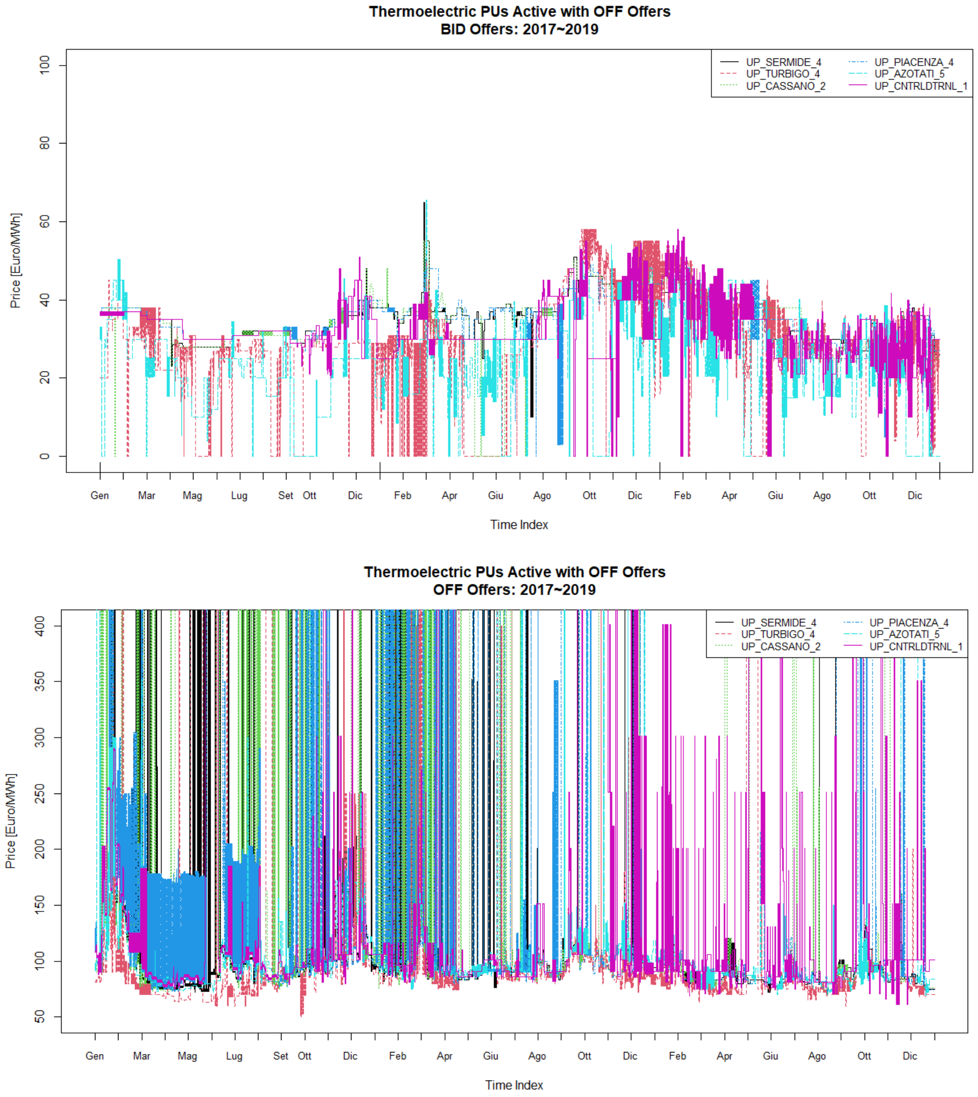


FIGURE 6.21: Cluster of Thermoelectric power plants active with OFF bids: visualization of MSD bids of BID type (above) and OFF type (below)

In Figure 6.21 I have reported the bids on MSD from all the units within this cluster. The two graphs present a much higher variability than the corresponding ones from thermoelectric units active with BID bids, which suggests that the cluster is more *fuzzy* than the previous one. In particular, even if these units have in common great earnings on MSD, it can be seen that they are characterized by quite different offer curves.

However, if we only **focus on the sub-cluster made by the three units from A2A** (that are *Cassano*, *Piacenza* and *Sermide*), we notice that these three have offer curves that are quite similar in the shape. To prove this, I have reported a visualization of the bids made on MSD in the year 2017 by these three power plants, in Figure 6.22.

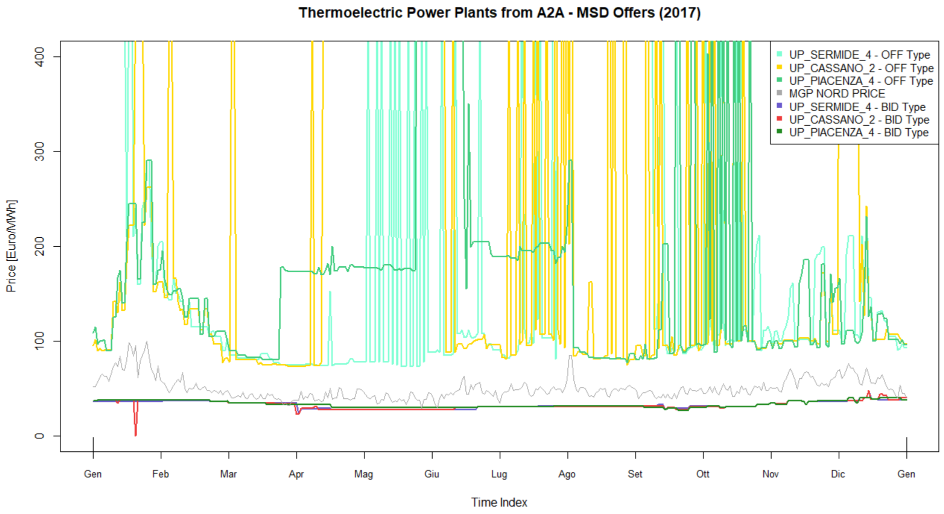


FIGURE 6.22: Sub-cluster of Thermoelectric PUs from the operator A2A, visualization of MSD bids presented in the year 2017 against the zonal price on MGP (in grey)

From this picture, we can notice how **the curves present almost no variability for MSD BID offers**, which are fixed at a price that is about 33 €/MWh for almost all the hours of the year. This value is quite interesting for two reasons: on one side, since it is constant across the whole year, **it implies a lack of interest from the operator to adjust its downward offers** according to the MGP price. On the other side, though, the value 33 €/MWh is quite high (especially if compared to the value 0 €/MWh of hydroelectric units from ENEL), therefore **these offers end up being very competitive in certain periods of the year**, where the zonal price on MGP is low. Conversely, **the units are quite active for OFF bids**, which present much greater variability and tend to mimic the trend of the zonal price on MGP.

The difference in macro-behaviour becomes even more evident if we compare the MSD BID offers by A2A to those of the previously presented cluster "Thermoelectric: MSD - BID", reported in Figure 6.19. This observation confirms how the units from A2A make compelling bids mainly of upward type, and therefore end up classified in the relative cluster by the K-Means algorithm.

Eventually, I have decided to **restrict to only the units from A2A** the analyses I will conduct in Chapter 7 on this cluster. In that occasion, I will show how the bids made by the units from A2A look quite similar to those by thermoelectric plants from the operator ENEL, since are mainly kept constant on a daily basis.

6.4.7 Cluster of Pumped-Storage Hydroelectric Power Plants from ENEL Produzione

UNIT REFERENCE CODES

- ENEL, Pumped-Storage Hydroelectric power plants: `P_BARGI_CEN_1`, `UP_EDOLO_1`, `UP_ETQ_ROVINA_1`, `UP_ETQCHIOTAS_1`, `UP_FADALTO_1`, `UP_GARGNANO_1`, `UP_RONCOVALG_1`, `UP_S.FIORANO_1`

This cluster concerns a group of 8 **Pumped-Storage Hydroelectric power plants from ENEL PRODUZIONE**, that present similar behaviour for MSD bids of BID type.

It was originally identified by the K-Means algorithm, which grouped together a set of 5 units with the same technology and operator[29]. Eventually, after a manual analysis of the offers on MSD from these PUs, I have decided to add also the units of *Bargi Centrale*, *Entracque Rovina* and *Fadalto* to the cluster[30], in order to include all the pumped-storage hydroelectric units from the operator ENEL in a single group.

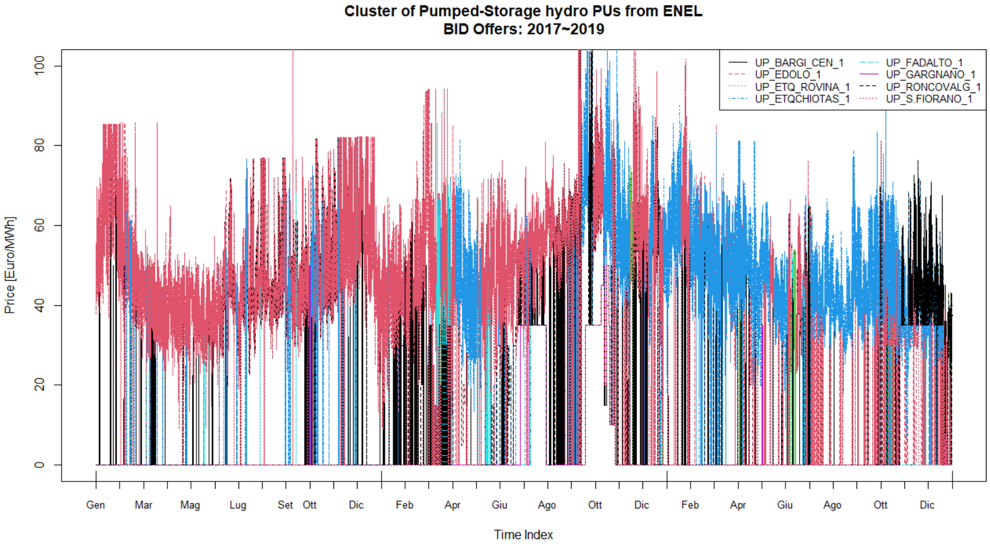


FIGURE 6.23: Cluster of Pumped-Storage Hydroelectric units belonging to ENEL PRODUZIONE: MSD GR1 bids of BID type

As we can see from Figure 6.23, the units in the cluster show great similarities in MSD bids of downward BID type, with offers characterized

---

[29] I recall that the algorithm was run without using the variables on technology and operator of the Production Units. Nevertheless, it was able to group together this set of power plants from ENEL, of pumped-storage hydroelectric type.

[30] These three additional units were originally classified by the algorithm into two separate clusters, mainly with other dispatchable hydroelectric power plants.

by the same variability and common patterns. Indeed, from the reported picture it is clear that the units tend to follow the zonal price on MGP in the definition of many of their bids, while sometimes they bid at a null price.

Moreover, we can observe that all the units present a **very peculiar behaviour around the month of October 2018**, when they change the minimal value of their offers from zero to the value 35 €/MWh. This is a similar behaviour to what was shown in Section 6.4.2 for the cluster of hydroelectric PUs from ENEL, confirming that **the operator adopts similar strategies for all its units of the same technology**.

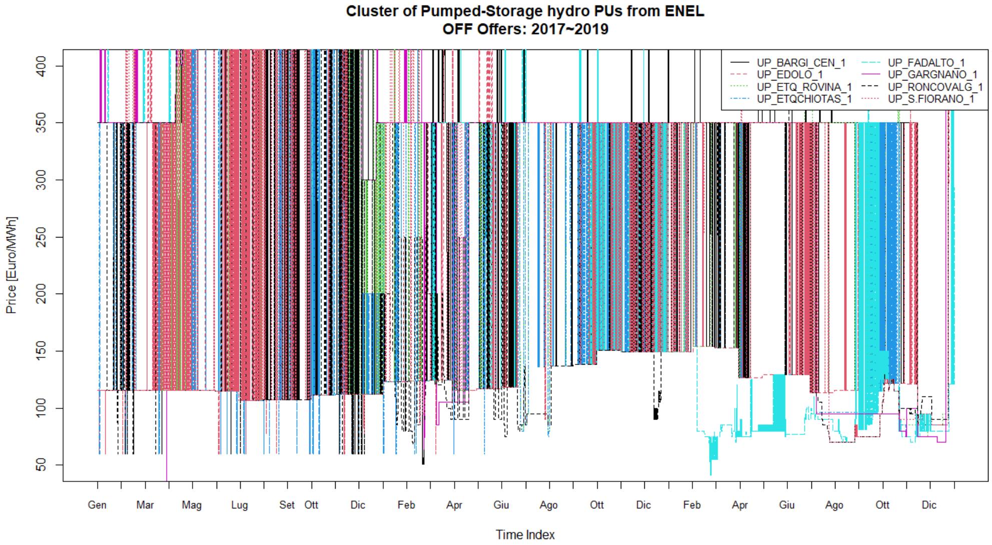


FIGURE 6.24: Cluster of Pumped-Storage Hydroelectric units belonging to ENEL PRODUZIONE: MSD GR1 bids of OFF type

In Figure 6.24 are reported the offer curves on MSD OFF of the Pumped-Storage hydroelectric power stations in the cluster, and we can appreciate that they are **quite significantly different in shape and variability** than those in downward direction reported in Figure 6.23. However, the curves present common patterns for what concerns minimum and maximum values of the bids, which change at the same times for most of the units in the cluster.

After this preliminary analysis, I have decided not to focus on pumped-storage hydroelectric units, since these type of technology has very peculiar behaviours on the electricity markets, which are usually not so trivial to interpret. This is a direct consequence of what was explained in the last paragraphs of Section 2.2, where I have introduced this type of technology.

Furthermore, this type of technology has seen a steady decline in use in Italy in the last 20 years[31], stemming from a decrease in the price differential

---

[31] For an overview on the subject I refer to an article on the blog **QualEnergia**.it

of energy between days and nights.

Nevertheless, that of pumped-storage hydroelectricity remains a crucial topic, since these units provide a way to store energy and could be of great importance for the further development of renewable power plants, maintaining the stability of the Power System.

6.4.8 Cluster of Hydroelectric Power Plants from CVA - *Compagnia Valdostana delle Acque*

UNIT REFERENCE CODES

- CVA Trading, Hydroelectric power plants:
  `UP_MAEN_5`, `UP_PERRERES_1`, `UP_VALPELLIN_1`, `UP_GRESSONEY_1`

In the second-last section of this chapter, I would like to introduce a cluster of minor importance, composed by hydroelectric Production Units that belong to the same operator.

The name of the operator is **CVA - *Compagnia Valdostana delle Acque***, and it is a small company that produces renewable energy through hydroelectric power stations, in the localized region of *Valle d'Aosta.* This operator does not have a particular economic relevance and only owns 4 power plants eligible on MSD, yet our algorithm is able to segment its units of *Maen* and *Perres* as a unique well-defined cluster.

For this reason, I thought that it would have been interesting to report briefly also the offers made by a different type of operator, of smaller dimensions than those analysed up to this moment.

In Figure 6.25 are reported the offer curves from the four units considered in this cluster, relative to the bids on the MSD market. From the two pictures, we can clearly notice how the smaller operator CVA adopts very *naive* strategies, with **bids that are slightly different among the PUs but often constant in both directions for long periods of time**. In particular, the unit of *Gressoney* is the one presenting the lower variability in the cluster, while that of *Maen* seems to be the most active one for OFF type of bids.

One more interesting thing to notice is related to the **presence of "hollows" in the bids of BID type**, which we can appreciate in similar shapes among the four units in the cluster, in the upper-most part of Figure 6.25. These happens in correspondence to the summer months and we could possibly interpret them by considering that all the units in the cluster are of hydroelectric type and located in the Alps. This fact implies that **these four PUs will have lots of primary resources available when the summer heat melts the surrounding glaciers**, increasing considerably the amount of water in the rivers of *Valle d'Aosta.*
It might be interesting to **compare this behaviour to that of Hydroelectric power plants from ENEL Produzione**, which I have presented in Section 6.4.2. In that occasion, I have inferred that the operator ENEL decided to place less competitive bids in periods of water shortages, especially in the autumn monts of 2019. What we can observe now is quite the opposite behaviour from CVA, that tries to "discourage" Terna to accept its downward bids on MSD, with less competitive offers during the summer,

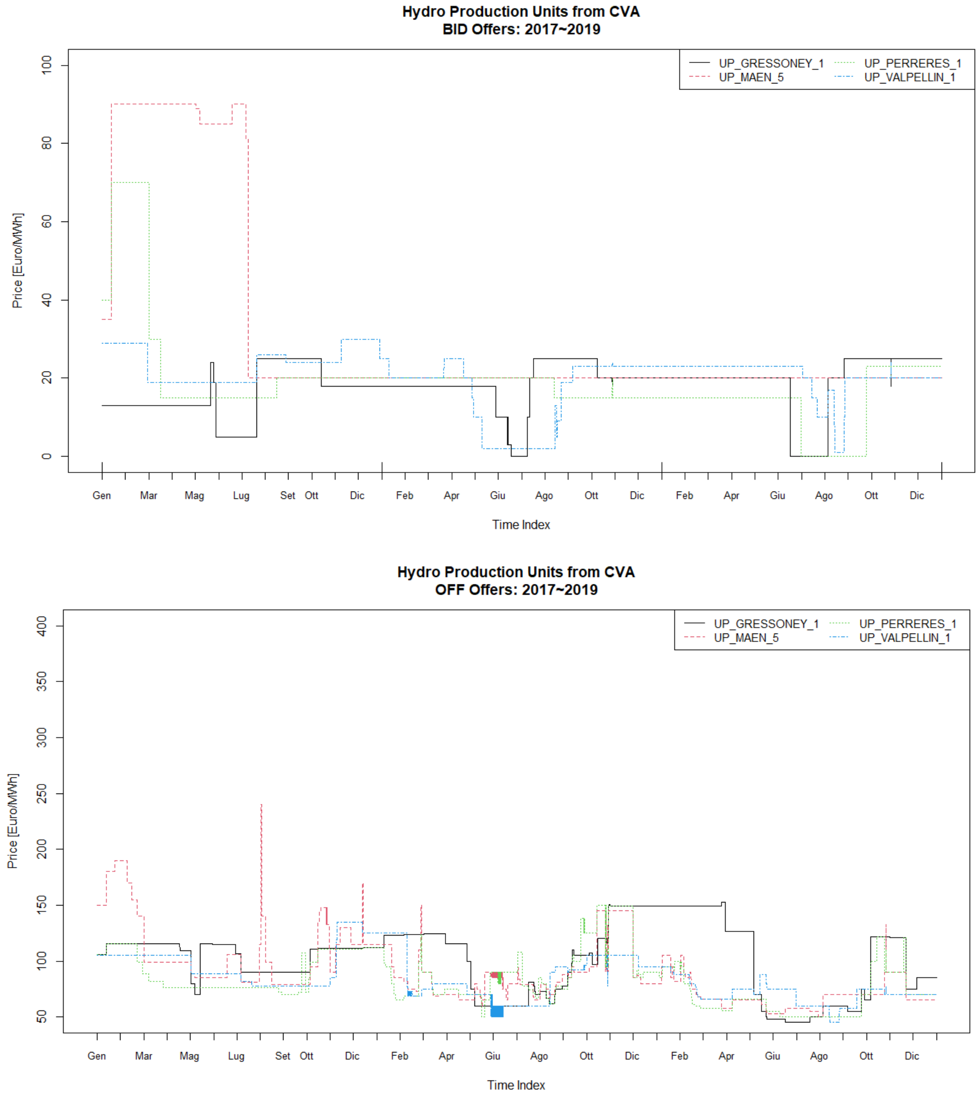


FIGURE 6.25: Cluster of Hydroelectric power plants from the operator CVA Trading: visualization of MSD bids of BID type (above) and OFF type (below)

when primary resources abound.

To conclude, I have presented a cluster of a small operator that bids on the Ancillary Services Market (MSD) with quite simple strategies. Indeed, from the available data we can infer that there exist some underlying strategies according to which CVA places its bids on the market, considering for instance the seasonality. However, the values of **the bids looks to be defined manually** and the offered prices always have integer values.

For this reasons, I will not analyze further this cluster and in Chapter 7 I will only focus on the most interesting power plants for MSD, which are often of thermoelectric production technology.

6.4.9 Final Remarks over the remaining Power Plants

In the previous sections, I have presented 8 different groups of power plants operating in similar ways on the Italian electricity markets, with a particular attention to their bids on MSD. This was the result of a major effort made to refine the results of the K-Means algorithm[32], and resulted in the classification of a total of 65 units out of the original 117.

The choice not to include all the PUs from the dataset inside this chapter is a conservative one, and is due to the **difficulty to include in well defined clusters each PU**. Conversely, I found it more interesting to focus on the few clusters of units that presented the most interesting behaviour, which are all included in this chapter.

Furthermore, those clusters include all the major units operating on MSD and comprehend the power plants on which I will focus in Chapter 7. In any case, I would like to spend few more lines on the units that I did not include, which I have left to future analyses.

Indeed, most of the non considered Production Units are hydroelectric power plants[33] that have limited installed capacity, thus they only play a minor role in the Ancillary Services Market.

Among these units, we can identify a group of 7 PUs from ***Dolomiti Energia***, which is a small operator active in a limited area of the Italian Alps, where it produces renewable hydroelectric energy. This operator is quite similar to *CVA - Compagnia Valdostana delle Acque*, which I have presented in Section 6.4.8, in both its structural characteristics and macro-behaviour on electricity markets. Eventually, since the bids on MSD from this operator are not particularly interesting for our purposes, I decided not to include it in this work.

Keeping the focus on hydroelectric power plants, there is another quite large group of 9 units from the operator ***Alperia***, with bids similar to those identified in the cluster "ENEL to Alperia", in Section 6.4.3. The algorithm K-Means struggled to group them into a single cluster, and tended to mix them up with other hydroelectric power plants, and a visual study of their offer curves confirmed that they do not seem to adopt common tactics.

Within the remaining hydroelectric units, we can identify a sub-group of **4 units with MSD bids of BID type that are constant to the value zero**, similarly to what I have reported in Figure 6.26 for the power plant of *Maso Corona*, by the operator *2V Energy*. Furthermore, these units present OFF bids that are constant for several months and do not seem to be particularly eager to take part in MSD, thus they result not being interesting for further analyses.

---

[32] Which, I recall, can be found entirely in **Appendix A** of this work.

[33] To be precise, a total of 36 units out 52, consisting in about two thirds of the remaining units.

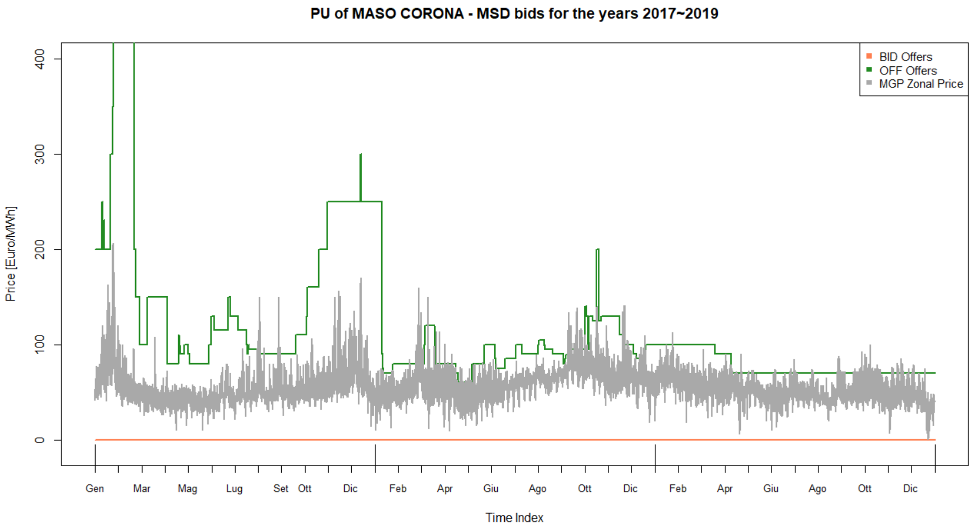


Figure 6.26: Production Unit of Maso Corona (*2V Energy SRL*), MSD bids in the years 2017 to 2019 (OFF offers in coral, BID offers in green) and MGP zonal price (in dark grey)

All the other PUs are mainly thermoelectric power plants, from various operators. Among them, we should mention the presence of few units from *A2A* which, however, behave quite differently from those presented in Section 6.4.6, thus end up being classified in different clusters. I have tried to analyse their characteristics, but I leave to future analyses to understand the reasons behind this different management choice by the operator *A2A*.

Eventually, I would like to conclude talking about the operator **EDISON**, which is one of the biggest in Northern Italy. Indeed, I managed to cluster two of its thermoelectric units in the clusters presented in previous sections, but I left unclassified most of its power plants. Indeed, a manual analysis of the offers on MSD by this operator, identified how it does not seem to exist a common pattern within its units.
This said, it is worth mentioning the presence of two power plants from EDISON with technology denoted as "*Asta Idroelettrica*", which in Italian stands for the union of many hydroelectric power plants that are located quite close one to the other, and therefore are related to the same water source. These PUs usually comprehend more than one power stations, that are aggregated and managed at the same time in order to be able to bid for greater amount of energy in the electricity markets. These two power plants consist in the units of *Taio* (in the region of Trentino) and *Venina* (on the river Adda, in Lombardy) and, due to their structural organization, reach a considerable value of installed capacity. For this reason, it may be interesting to study them in further analyses.

6.5 Final Considerations on the macro-behavioural Clustering

To conclude the chapter, I would like to sum up all the major discoveries on the macro-behaviour of the selected PUs in the bidding zone `NORD`.

My analyses identified how there are **profound differences in the strategies adopted by operators** while bidding on the electricity markets. In particular, the K-Means algorithm highlighted **great variability in the activities across the three considered markets**, with the operator ENEL that seems to be by far the most active one on the Intraday Market (MI).

Moreover, some operators decide to **manage all their Production Units of the same technology in a similar way**, with MSD bids that are almost equal across all the various PUs, as is the case for the thermoelectric units by ENIPOWER presented in Section 6.4.1. Conversely, other operators (like EDISON and A2A) behave in a different way, with MSD bids that may be similar within the Production Units of the same power plant, but **tend to differ from one power plant to another of the same technology**.

In this regards, we can also observe that the **operator ENEL behaves in a dual way**, as I have shown in Sections 6.4.2 and 6.4.4. On one side, it tends to manage all its hydroelectric units in the same way, with MSD bids that are almost identical among dozens of units. On the other side, though, we can appreciate a great differentiation of the bids from its thermoelectric PUs, which can vary substantially even inside a single power plant, as shown in Figure 6.18.

Regarding the production technology of the PUs, the clustering algorithm confirmed the initial hypothesis that **activity on the Ancillary Services Market is closely tied with the technology of the power plant**, with clusters of CCGT units that stood out for accepted offers of both upward and downward type.

Indeed, we often see **hydroelectric units bidding in a much simpler way** than thermoelectric units, with offers that are usually *plateau* for many weeks or even months. Of course, these "simple" bids could be the reason of their limited activity on MSD, however, it is even more likely that this behaviour is the consequence of the nature of the Market for Ancillary services and the **tendency by Terna to reject the bids made by hydroelectric units**.

Eventually, the macro-behavioural analyses laid the basis for some interesting observations on MSD activity by the PUs in the dataset. In the following chapter, I will dive deeper on this topic, analysing the BID and OFF bids from four of the identified clusters.

Predictive Analysis of MSD bids via Random Forests 7

In the past chapter, I have shown 8 different clusters of units operating on the Italian electricity markets, precisely in the zone `NORD`. While doing so, I have posed particular attention to their bids on MSD, due to the fact that the market is not based on a merit order and, therefore, it is not sufficient to describe the activity on MSD of a PU based solely on the acceptance results of its bids.

In the following chapter, I would like to **focus on four of the identified clusters**, with the scope to **analyse and interpret the bids on MSD** from their Production Units. For this reason, the clusters I chose are the most interesting ones for MSD bids, and comprehend only units of thermoelectric technology. Three of them comprises units from only one operator, respectively **ENIPOWER**, **A2A** and **ENEL**, while the last cluster is the one of **Thermoelectric PUs active with MSD bids of BID type**, introduced in Section 6.4.5.

For all the considered clusters, I will begin with a general overview of their offers on MSD, presented in Section 7.1. This preliminary analysis will mark substantial differences within the four clusters, since I will show that **some groups of units place their bids on a daily basis**, thus disregarding completely the "ramp" hours of weekdays (i.e. those associated to the highest electricity load on the grid), as well as the quotidian oscillations of the electricity price on MGP.

For that reason, I move on to the hourly prediction of MSD bids for only the two clusters of **ENIPOWER** and **Thermoelectric - MSD BID** units. To do so, I will introduce in Section 7.2 the **algorithm Random Forests**, which I will then apply to the problem under consideration. I will also describe the **feature selection process**, as well as the **tuning** of the algorithm, done mainly with the help of [5], [19] and [20].

Finally, the results of these procedures are presented in Section 7.3 for the cluster of ENIPOWER and Section 7.4 for the group of Thermoelectric units active with bids of BID type on MSD.

7.1 Preliminary considerations on MSD bids

As I have shown in Chapter 6, many Production Units of those eligible to operate in MSD, seem to be **effectively uninterested to place competitive bids at every hour** on the market for ancillary services. This holds true in particular for hydroelectric power plants, since most of them bid at a **constant price for subsequent weeks**, or even months.
That being said, a deeper analysis of the bids from **thermoelectric units** identified the fact that not all of them adapt their bids hour by hour, with some of them placing **bids that are constant on a daily basis**, and others that define **"recurring patterns"** of prices, with daily frequency, that are then left unchanged for many weeks.

In the following sections, I will dive deeper into the analysis of four major clusters of thermoelectric units, among those that were identified in Chapter 6, **highlighting the relevant characteristics of their MSD bids**.
The selected clusters are the following:

1. Cluster of Thermoelectric Power Plants from **ENIPOWER and S.E.F.**
2. Cluster of Thermoelectric Power Plants from **ENEL Produzione**, focusing on the power plants of ***La Casella*** and ***Fusina***
3. Sub-Cluster of Thermoelectric Power Plants from **A2A**, extracted from the group of units active with MSD-OFF bids
4. Cluster of **Thermoelectric Power Plants active with bids of BID type**

In Section 6.4 I have already reported the offer curves on MSD (of both BID and OFF type) from the units in each of these clusters, and **all of them seem to present a great degree of variability in their bids**, over the three-year time span of my analyses.
However, what I found out by looking more accurately at each cluster, was that a visualization of their bids on the large scale of three years was not sufficient to describe completely their behaviour on the market for ancillary services. Indeed, from Section 6.3, we can observe how all the selected units tend to **adapt the prices of their offers to the seasonal oscillations of the zonal price on MGP**, but these figures still give almost no information on the precise frequency of update of those bids, since the graphs related to the years 2017 to 2019 tend to compress quite a lot the bids of the production units.

To give a better visualization of this topic, I have selected **four specific weeks in our time span**, belonging to different months and about equally spaced in the range of our observations. Then, I have retrieved the bids from the above-mentioned clusters, during the four weeks under analysis, and I have plotted them in different graphs. The complete results can be found in **Appendix B** of this work[1], and confirm that there exist major

---

[1] The scope of this additional chapter is to give a more precise understanding of

dissimilarities between the behaviours of different operators.

In the following lines I illustrate these major differences and I will also pose the basis for the prediction analysis conducted in Section 7.2, which will be the main focus of this Chapter.

7.1.1 Cluster of Thermoelectric Power Plants from ENIPOWER and S.E.F.

The first cluster that I report is the one of thermoelectric PUs from the operator Enipower, introduced in Section 6.4.1. As already mentioned, this is one of the most interesting clusters to be analysed, since **its units submit bids on MSD that present a lot of variability**, as was firstly depicted in Figures 6.6 and 6.7.

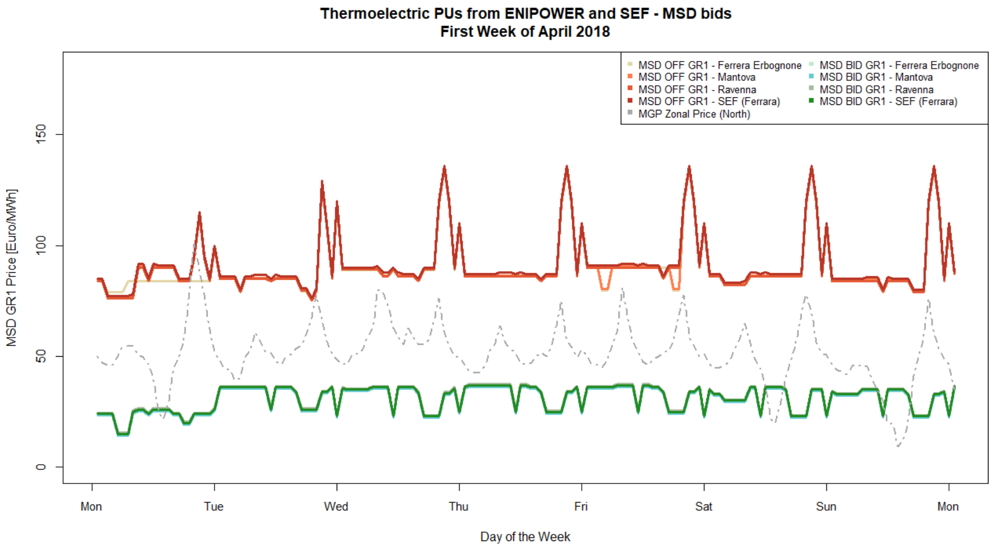


Figure 7.1: Cluster of Thermoelectric Production Units from ENIPOWER and S.E.F.: MSD GR1 bids submitted in the first week of April 2018. OFF bids are plotted in red tones, BID bids are in green tones, the zonal price on MGP is reported in grey.

In Figure 7.1 I have depicted the bids presented by one PU per each power plant in the cluster, relative to the specific week of April 2018, and I recall that three more weeks of bids from this cluster can be found in Figure 2 of **Appendix B**.

From Figure 7.1 we can notice that **almost all the units in the cluster submit identical bids at every hour**, regardless the fact that they are

---

how each cluster places its bids on a daily basis. Ideally, I would have liked to report a zoomed-in visualization of every week of bids in our dataset, but this was clearly unfeasible. In any case, the analyzed weeks were chosen completely randomly, and the reported results are in line with the data observed in the whole time span.

geographically located quite far one from the other. The bids present a **high degree of variability** also at a weekly scale, with OFF bids that tend to be above the price of energy on MGP and, in particular, seem to increase considerably during the evening ramp hours.
Even downward BID bids change quite frequently, although in this case it is more visible a recurring pattern in the offers, at a daily basis.

Overall, the offer curves on MSD - GR1 of the cluster of ENIPOWER and S.E.F. comply with the initial expectations and I will perform a predictive analysis of these bids in Section 7.3.

7.1.2 Cluster of Thermoelectric Power Plants from ENEL

The second cluster to be analyzed in details is the one of Thermoelectric PUs from ENEL, which was presented in first place in Section 6.4.4. I have decided to focus on the power plants of *La Casella* and *Fusina*, which present a very peculiar behaviour and together comprehend 8 of the 11 PUs in the cluster.

Indeed, the **bids on MSD of this cluster** are quite interesting to study, since they **seem to be defined on a daily basis, and are constant from midnight to 11 p.m. of almost every day**.

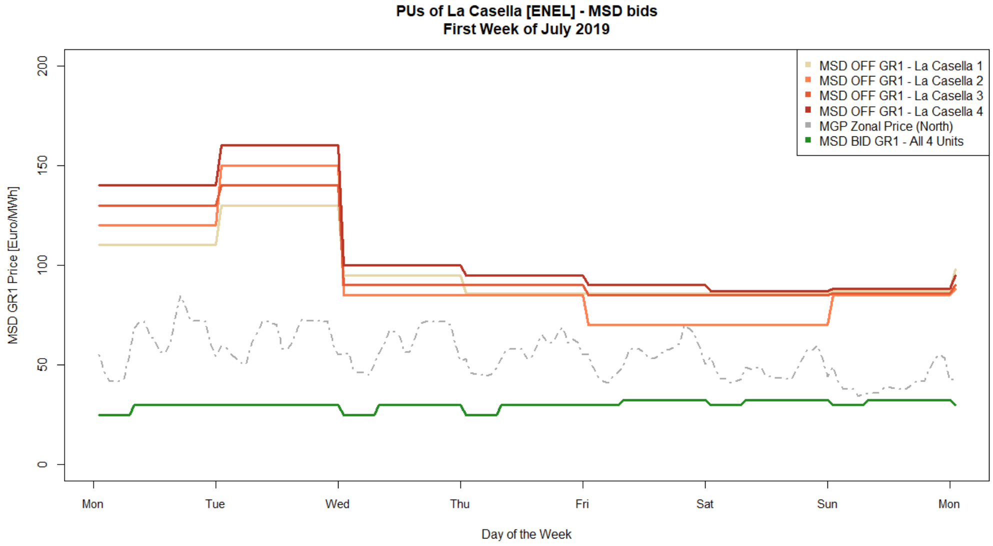


FIGURE 7.2: Cluster of Thermoelectric Power Plants from ENEL ***La Casella***: MSD GR1 bids submitted in the first week of July 2019.
OFF bids are plotted in red tones, BID bids are in green tones, the zonal price on MGP is reported in grey.

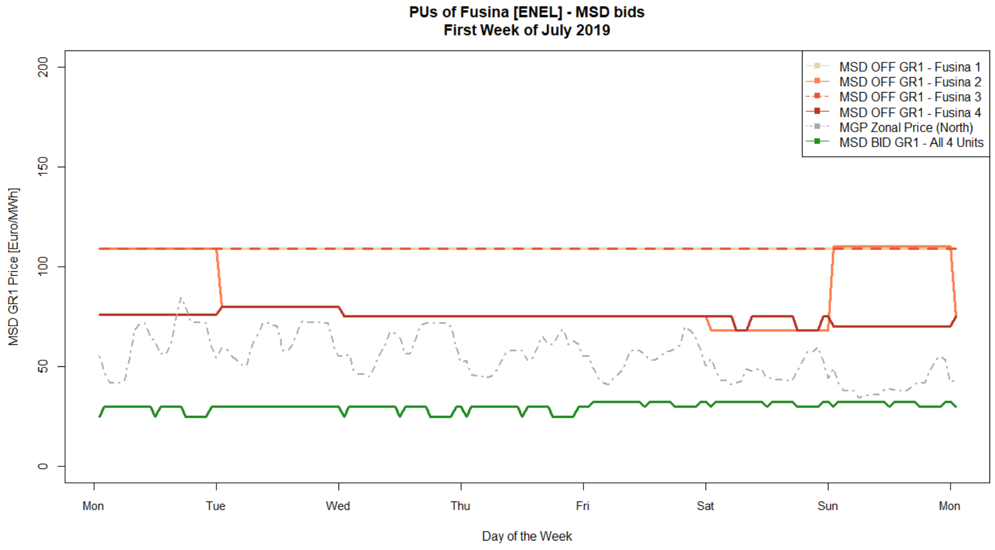


FIGURE 7.3: Cluster of Thermoelectric Production Units from ENEL ***Fusina***: MSD GR1 bids submitted in the first week of July 2019.
OFF bids are plotted in red tones, BID bids are in green tones, the zonal price on MGP is reported in grey.

In Figures 7.2 and 7.3 I have depicted the offer curves on MSD of the power plants of *La Casella* and *Fusina* respectively, related to the first week of April 2019 (1 April 2019 to 7 April 2019).
In the two pictures, we can see how **upward OFF bids remain constant for almost every Production Unit, throughout every day** of the selected week, regardless the hourly oscillation of the zonal price on MGP.

This behaviour is quite a surprising one, since it **differs considerably from that of the previously presented cluster** of ENIPOWER and S.E.F.. It also appears to be somehow "sub-optimal", due to the fact that at certain hours the upward bids on MSD almost reach the price of energy on the Day-Ahead Market (MGP). Moreover, the decision to update the bids at midnight, seems to be entirely arbitrary and not well founded[2].
To further support the idea that the above one is not an isolated behaviour, I refer to Figures 3 and 4, located in **Appendix B** of this work.

Regarding **downward bids on MSD of BID type**, in the selected weeks we can see that they are **always equal across the PUs of each power plant**. This, again, is not an isolated behaviour, with the power plant of *La Casella* that presents the same downward bids for the four PUs in 82% of the observations in our dataset, with this value growing to 93.5%

---

[2] Indeed, given the decision to present bids that are constant for 24 hours straight, it would probably make more sense to **follow the trend of the Day-Ahead Gas Market**, which, relative to a day d, goes from 6 a.m. of day d to 5.59 a.m. of day $d+1$.

when we consider the units from the power plant of *Fusina*[3].
That being said, MSD bids of BID type seem to be updated by ENEL more regularly than OFF bids, even though they seem to follow daily patterns, updated at midnight and constant for few consecutive days. In particular, this fact is visible for the power plant of *La Casella*, from the graphs reported in Figure 3 of **Appendix B**.

Eventually, I would like to point out one more interesting issue, related to the fact that **MSD bids from the operator ENEL have most of the times integer values**, as we can observe from Table 7.1.

Percentage of bids with integer value:	**ENEL** ***La Casella***	**ENEL** ***Fusina***
MSD BID	99.68%	99.68%
MSD OFF	97.14%	86.80%

Table 7.1: Thermoelectric PUs from ENEL: Percentage of bids on MSD-GR1 having integer values, in the years 2017 to 2019, relative to the power plants of *La Casella* and *Fusina*

This fact is in contrast to the majority of the other operators in our dataset, which generally tend to present bids exploiting also decimal factors. Furthermore, focusing only on **OFF bids** from thermoelectric units of the operator ENEL, I found that **about** 80% **of them have values that are multiple of 5**.
This behaviour indicates with great likelihood a **human involvement in the process of bid definition**, with offer prices that are often rounded to the closest multiple of 5.

Overall, it is quite easy to predict the bids from the operator ENEL, and this can be done just by studying the values of the offers in prior hours, which, to the operators competing on MSD, are generally available with little delay.
In my analysis, however, I found it quite **hard to predict with great precision the values of each daily constant threshold**, from solely the data presented in Chapter 5. Indeed, most of the available data change on a hourly basis, with values that differ quite significantly even within the same day, as is the case for the *Prezzo Unico Nazionale* (PUN) and the variables on generation and load forecasts.

This is the main reason why the Random Forest prediction of the bids by ENEL did not give meaningful results, other than confirming the fact that **the offers presented by the operator ENEL do not depend much from the variables that are updated with hourly frequency**, such as the zonal price on MGP or the Forecasted Load on the Italian grid.

---

[3] To put this in perspective, the submitted OFF bids are equal in only about 10% of the cases, for both clusters.

7.1.3 Cluster of Thermoelectric PUs from A2A

This section is dedicated to the analysis of the offers on MSD from four thermoelectric units, belonging the operator A2A. These units were identified as a sub-cluster of the PUs listed in Section 6.4.6 of Chapter 6, dedicated to the group of **Thermoelectric power plants active with OFF bids on MSD**. In that occasion, I introduced the units of *Cassano*, *Piacenza* and *Sermide*, together with their MSD bids, in Figure 6.22.
To these units, I then decided to add the one of *Chivasso*, which belongs to the operator A2A as well and seems to present similar bids on the market for ancillary services.

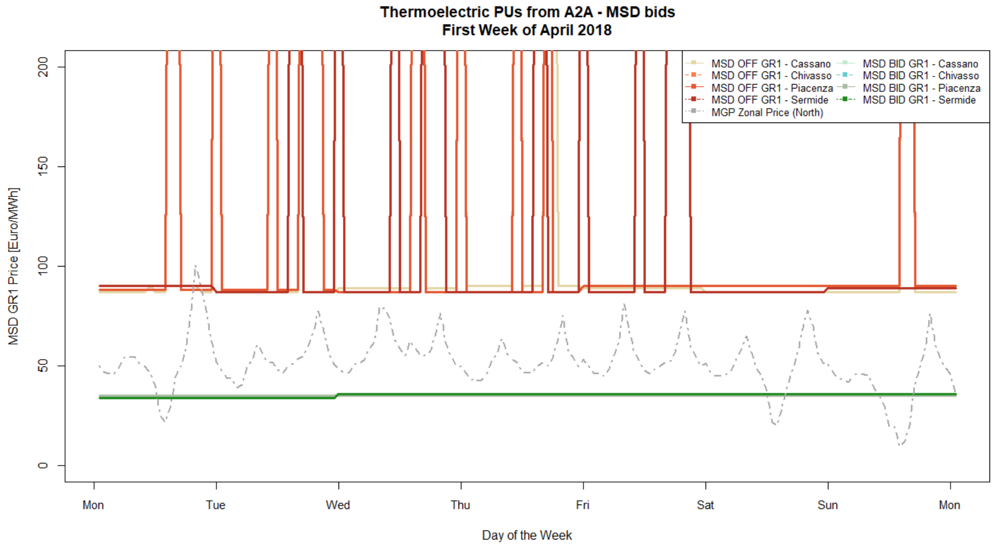


FIGURE 7.4: Cluster of Thermoelectric Production Units from A2A: MSD GR1 bids submitted in the first week of April 2018.
OFF bids are plotted in red tones, BID bids are in green tones, the zonal price on MGP is reported in grey.

In Figure 7.4 I have depicted the bids presented by each Production Unit in the cluster, relative to the specific week of April 2018, and I recall that three more weeks of bids from this cluster can be found in Figure 5 of **Appendix B**.

What becomes extremely clear from this picture, is the fact that **BID bids present almost no variability**, as we inferred in the previous chapter from Figure 6.22, which showed how MSD offers of BID type remain **effectively unchanged during the course of the years 2017 to 2019**. Indeed, this is probably the reason why these units from A2A are clustered by the K-Means algorithm in the group of units identified in Section 6.4.6, which segments PUs characterized by moderately high revenues from MSD OFF bids, but **low activity from downward MSD bids, of BID type**.

Percentage of bids with integer value:	***Cassano***	***Chivasso***	***Piacenza***	***Sermide***
MSD BID	100%	98.54%	100%	99.36%
MSD OFF	100%	100%	100%	100%

Table 7.2: Thermoelectric PUs from A2A: Percentage of bids on MSD-GR1 having integer values, in the years 2017 to 2019, relative to the power plants of *Cassano*, *Chivasso*, *Piacenza* and *Sermide*

As for the cluster of ENEL, it is interesting to notice how **MSD bids** from the PUs in this cluster **have almost every time integer values**, as it is shown in Table 7.2. This is a peculiarity of the operators ENEL and A2A, and confirms that also for the latter one there is probably a human involvement in the process of bid definition, although MSD-GR1 bids are not rounded as much as in the previous case.

Regarding **upward MSD OFF bids** from the operator A2A, these seem to present a **much greater variability than BID ones**. The variability, however, is given by sudden "jumps" to extremely high values, which are usually set to about 500€/MWh, and thus indicate a general will not to be accepted on MSD. This said, looking more closely at the OFF bids, we can notice that they are generally characterized by a **constant baseline during the majority of the days**, to which the bids deviate only for the "extreme" offers at very high prices. Moreover, similarly to the operator ENEL, this baseline is usually updated at midnight, and tends to be quite close to that of the previous day.

In my analyses, I have tried to interpret in many ways the peaks for MSD OFF bids, but I did not manage to link them with any of the variables in my dataset. In particular, they **do not seem to be related to the programmed production after the market MI**, as it was the case for the cluster of Hydroelectric Production Units from ENEL, presented in Section 6.4.2.
Furthermore, I have also tried to predict the OFF bids from this cluster with the algorithm Random Forest, in the same way as the one presented in Section 7.2, but I did not reach meaningful results[4].

I would like to conclude this section by saying that the bids on MSD from the thermoelectric Production Units of the cluster of A2A appear to be **quite difficult to interpret**. Indeed, the operator appears to be uninterested in BID type of offers but, on the opposite, it submits OFF bids that are much more complicated to explain than its competitors.
Overall, I believe that **further analyses will be needed for this cluster**,

---

[4] The final model only explained 38% of the total variability, with an Out Of Bag error that was still extremely high, effectively meaning that the model was unable to capture the reasons behind the upward peaks in OFF bids.

possibly considering also other services than the only GR1 and including even more variables than the ones adopted in this work.

Indeed, with a brief preliminary analysis on the accepted MSD OFF bids from the unit of *Cassano*, I found out that it experienced **a total of 3127 accepted "AS+OFF" bids**[5] in the period of our analysis, corresponding to a staggering 11.9% of all the hours under consideration.
Such bids are generally **remunerated at a higher price than the PUN on MGP**, and involve considerable quantities of energy (i.e. since the unit is then set at its minimum value of production), implying **great revenues**[6] **for the PUs in consideration**.

For this reason, if this tendency was confirmed also for the other units in the dataset, it would imply the necessity of a **broader study** on the bids on electricity markets from the operator A2A, considering every possible type of bid service and not only the GR1 considered in this work.

7.1.4 Cluster of Thermoelectric Power Plants active with MSD bids of BID type

The last cluster I would like to analyze, is the one of thermoelectric PUs active with MSD bids of BID type, introduced in Section 6.4.5. This one is the only cluster in the chapter that is composed by more than a single operator, and **it comprehends a set of 6 different units that reported great expenses from downward bids on MSD**, relatively to their installed capacity.

As mentioned in Chapter 3, MSD offers of BID type correspond to a request from the unit to "buy back" energy from Terna, usually at cheaper prices than those awarded in previous markets. Due to this fact, **these bids naturally involve complicate strategies across all electricity markets**, which is probably the reason why many operators (such as ENEL and A2A) tend to disregard them entirely, offering constant prices for long periods of time.

In Figure 7.5 I have depicted the bids presented on MSD by each Production Unit in the cluster, for the specific week of April 2018, and I recall that three more weeks of bids from this cluster can be found in Figure 6 of **Appendix B**.
From the reported graph, we can see that there are greater differences in the bids submitted by the PUs in the cluster, especially for those of OFF type. This, though, was expected, since we are aggregating together bids from different operators[7].

---

[5] As I have explained in Section 3.4, that kind of bids is related to the hours in which the unit is inactive as a result of MI and it starts producing on MSD.

[6] For the unit of *Cassano*, this is equal to 180 MW of capacity, implying that the unit earns, on average, little less than 17000€ each time such a bid is accepted.

[7] In particular, there are few similarities between the upward OFF offers of units from the same operator, such as the two units of *Leini* and *Voghera*, belonging to **Engie Italia**, and the two of *Livorno Ferraris* and *Tavazzano*, from **EP Produzione**.

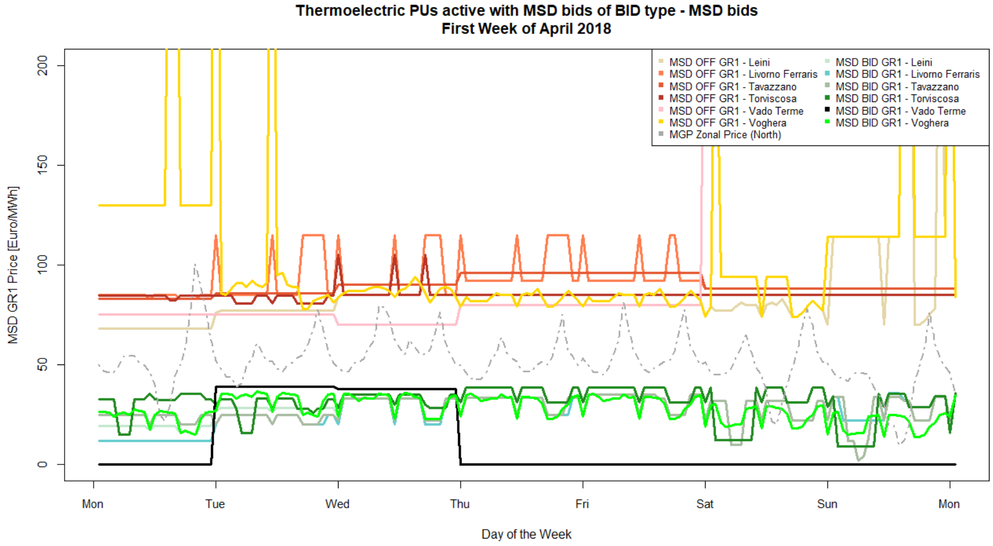


Figure 7.5: Cluster of Thermoelectric Production Units from A2A: MSD GR1 bids submitted in the first week of April 2018.
OFF bids are plotted in red tones, BID bids are in green tones, the zonal price on MGP is reported in grey.

Eventually, I have decided to **focus only on MSD-GR1 bids of BID type** for this cluster, since they are at the basis of the original segmentation of these PUs by the algorithm K-Means.

Looking at BID type of bids, five out of six units seem to present very similar offer curves, with prices that change almost every hour and having similar patterns of oscillation. The **only exception is the unit of *Vado Terme*, from the operator Tirreno Power**, which seem to behave quite differently from the remaining PUs. In particular, the unit of *Vado Terme* presents **MSD bids that are constant on a daily basis** and are updated at midnight, in an analogue way to the units of ENEL already presented in previous clusters.

Furthermore, this unit seem to alternate days in which it bids at the value zero, to others in which it submits really competitive prices, generally above those of the other units in the cluster. Indeed, this might be the reason why the unit reports similar results on the market for ancillary services across the three years, and is therefore grouped with the other five units.

To conclude, **the downward BID offers on MSD-GR1 of this cluster seem to be extremely interesting** and need to be studied more in depth. Therefore, I will perform a predictive analysis of their prices in Section 7.4. Finally, since the prediction analysis will be done at a hourly basis, I will exclude the unit of *Vado Terme* and consider only the remaining five.

7.2 Random Forest prediction and interpretation of MSD-GR1 bids

This section is dedicated to the prediction and interpretation of bids on the Italian market for ancillary services and will focus on the two clusters of **Enipower** and **Thermoelectric - MSD BID** units.
To predict and analyse the bids on MSD-GR1, I decided to use the algorithm **Random Forest**, since it combines **great predictive performances with a relatively high degree of interpretability**[8]. Indeed, since the final goal of this work is to understand how PUs place GR1 bids on MSD, it is necessary to use an algorithm that not only reaches great predictive performances on the available data, but is also quite easy to interpret.

The algorithm Random Forest was introduced in 2001 in a paper by Leo Breiman [19], in which it was described a novel *ensemble* method based on the construction of a multitude of decision trees, at training time[9].
Effectively, **the algorithm aggregates a collection of random decision trees**, which are not necessarily optimal in their own, but allow for an extensive exploration of all the possible tree predictors, thus resulting in **better predictive performances** than prediction trees.

Citing [5], we may define a Random Forest as follows:

Definition 7.2.1 (Random Forests)
Given a learning sample $\mathcal{L}_n = \{(X_1, Y_1), ..., (X_n, Y_n)\}$.
Let $\left(\hat{h}(., \Theta_1), ..., \hat{h}(., \Theta_q)\right)$ *be a collection of tree predictors, with* $\Theta_1, ..., \Theta_q$ *q i.i.d. random variables independent of* $\mathcal{L}_n$. *The Random Forest predictor* $\hat{h}_{RF}$ *is obtained by aggregating this collection of random trees.*
For ***regression*** *purposes, the aggregation is done as follows:*
$\hat{h}_{RF}(x) = \frac{1}{q}\sum_{\ell=1}^{n} \hat{h}(x, \Theta_\ell)$ *(average of individual tree predictions)*

Random forests may be used for both regression and classification purposes, but in this work I will only use them in the first way, to predict the price of MSD-GR1 bids from a selection of PUs in our dataset.

I performed my statistical analysis with the help of the programming language `R` and the package `randomForest` [20], since it has already built-in some extremely useful methods for the assessment of the results of the regression models.
In the following paragraphs, I will present an overview of the major technical details of the algorithm, including the concept of **Out Of Bag Error**, that will be used to evaluate the performance of each model, and I will also talk

---

[8] The algorithm Random Forest is one of the most interpretable among "black-box" models, since it is theoretically possible to study each of the trees used by the algorithm in its predictions. Furthermore, the algorithm is quite flexible, since it does not make any assumption regarding the underlying model.

[9] This was effectively a refinement of Classification And Regression Trees (CARTs), introduced in the mid-1980s.

about **Variable Importance** and **Partial Dependence Plots**, which, instead, are used to interpret the results given by the Random Forests.

7.2.1 Technical Considerations on the algorithm

As anticipated in the previous paragraphs, the algorithm Random Forest was born as an improvement of Classification And Regression Trees (CARTs). In particular, Random Forests would like to mitigate a well-known negative behaviour of CART trees, namely their instability [5]. In this context, by instability we mean the tendency of tree predictors to change, even considerably, if the original data are slightly perturbed.

Indeed, there exist different versions of Random Forests, defined on the basis of how it is handled the issue of instability[10], and in this work I will use and explain the one called **Random Forest Random Input**, introduced by Breiman in 2001 and currently the reference method.

The principle of the **RF-RI** algorithm is simple and consists in aggregating a set of **tree predictors that are built with random input variables and starting from a *bootstrap sample* of the data**.

Definition 7.2.2 (Bootstrap Sample)
Given a learning sample $\mathcal{L}_n = \{(X_1, Y_1), ..., (X_n, Y_n)\}$ *of size* n*, we can obtain a Bootstrap Sample of it by drawing, with replacement,* n *different observations from the original learning sample. Each draw has uniform probability over the observations* (X_i, Y_i) $[i = 1, ..., n]$ *in the learning sample.*

To further explain the **RF-RI** algorithm, I report its main steps as follows:

1. **Choose the parameters** of the algorithm, namely: the number of trees in the forest (`ntree`), the number of variables to try at each node (`mtry`) and a stopping criteria for each tree
2. For each tree in the forest, do as follows:
    a) Draw a **bootstrap sample** from the available data
    b) At each node of a tree, randomly **select a subset of variables**, in number equal to `mtry`
    c) Look for the **best way to split** the bootstrapped data, based on the selected subset of variable
    d) Proceed with the selection of a new subset of variables and the creation of a new split, until one of the stopping criteria is reached
3. **Aggregate all the `ntree` trees in the forest**, obtaining the final Random Forest predictor

Regarding the **stopping criteria**, many possibilities exist, and I decided to proceed with the most common one, based on the minimum number of

---

[10] Indeed, the most straightforward version of Random Forests is called **Bagging**, which stands for *"Bootstrap Aggregating"* and indicates an ensemble predictor that is obtained by an aggregation of many CART trees, each of them built on a different bootstrap sample of the data.

observations that should be contained by a leaf of every tree, controllable via the parameter `nodesize`.
Finally, in the regression case of RF-RI, the **best split is generally defined by minimizing the within-group variance** resulting from splitting a node t into two child nodes t_L and t_R, as explained in [5].

Focusing on Random Forest Random Input algorithms for regression purposes, it is interesting to spend few words on the concept of **Out-Of-Bag (OOB) Error**. With this term, it is indicated a **method to assess the predictive performances** of the aforementioned algorithm, with an estimate that is of the same type as cross-validation, effectively making unnecessary the creation of a test set.
Indeed, the OOB Error is an estimator of model's error, which uses the **observations that were not included in the bootstrap sample as test data**[11].

The package `randomForest` on `R` has also another interesting tool to assess the goodness of the fit, simply by printing an object of type `"randomForest()"`. What I am referring to is the **percentage of variability explained by the model**, which is evaluated automatically by the algorithm and has the aim to measure how well out-of-bag predictions explain the variance of the target variable.
More precisely, the percentage of variability explained is evaluated as follows:

$$\%\text{VAR} = 1 - \frac{errOOB}{Var(Y)}$$

Where $errOOB$ is the Mean Squared Error from the OOB sample and $Var(Y)$ is the variance of the target variable.

I will now move on to present two different tools that are commonly used to **interpret the results of the Random Forest algorithm**, namely Variable Importance and Partial Dependence Plots.

The goal of **Variable Importance** is the definition of a hierarchy of the variables in input, based on how each of them affects the output variable. To this extent, the most widely used (*non-parametric*) method to asses the importance of a variable in the model is to study the increase in error made by a tree, when the data from the variable under considerations are randomly permuted in the OOB sample.
Citing [5], we can define Variable Importance as follows:

Definition 7.2.3 (Variable Importance)
Let us fix $j \in \{1, ..., p\}$, with p being the number of predictors in the dataset, and calculate $VI(X^j)$, the importance index of variable X^j:

- *Consider a bootstrap sample $\mathcal{L}_n^{\Theta_\ell}$ and the associated OOB_ℓ sample, that is, all observations that do not belong to $\mathcal{L}_n^{\Theta_\ell}$.*

---

[11] On average, 36.8% of the available data end up not being included in each bootstrapped sample [5].

- *Calculate* $errOOB_\ell$*, the error made on* OOB_ℓ *by the tree built on* $\mathcal{L}_n^{\Theta_\ell}$ *(mean squared error or misclassification rate).*
- *Then randomly permute the values of variable* X^j *in the* OOB_ℓ *sample. This gives a perturbed sample, noted* $\widehat{OOB}_\ell^j$*.*
- *Finally, calculate* $err\widehat{OOB}_\ell^j$*, the error made on* $\widehat{OOB}_\ell^j$ *by the tree built on* $\mathcal{L}_n^{\Theta_\ell}$*.*
- *Repeat these operations for all bootstrap samples. The importance of the variable* X^j*, namely* $VI(X^j)$*, is then defined by the difference between the average error of a tree on the perturbed OOB sample and that on the OOB sample:*

$$VI(X^j) = \frac{1}{q}\sum_{\ell=1}^{q}\left(err\widehat{OOB}_\ell^j - errOOB_\ell\right)$$

Indeed, the prediction error should increase when we permute the values of one variable, and, **the higher the error increase, the more important is the variable**. Fortunately, the `randomForest` package in `R` has already built-in all the required methods to assess variable importance and it also provides the function `varImpPlot()` for a quick visualization of variables' hierarchy, in decreasing order of importance.

Eventually, I would like to introduce the topic of **marginal effects**. Indeed, Variable Importance is used to assess *which* variables are more responsible to influence the output of the model, but it does not take into account *how* those variables actually affect the output. To solve this issue, I have used the concept of marginal effects, which are a way to tell **how the output of a model (dependent variable) is affected when a specific independent explanatory variable changes over all its possible range, while other covariates are held constant**. This idea is commonly used in generic regression models and for Random Forests is often visualized via **Partial Dependence Plot**.

Once again, this procedure is done quite easily with `R`, employing the function `partialPlot()` of the package `randomForest`.

7.2.2 Specifics of the Adopted Variables

This section is dedicated to an overview of the variables that I used as input for each of the Random Forest models that will be discussed in the following sections. For each model, the **target variable was the price of the MSD GR1 bid of BID or OFF type**, in one hour of the three year time span 2017 to 2019.

Thus, each observation in the dataset was composed by a set of independent variables, related to the same hour as the target one, and **publicly available before the observation of the MSD GR1 prices**. Furthermore, some variables are referred to the specific Production Unit that is placing the bid on MSD (in that case, they are always normalized by the installed capacity of the PU), others are common to all the units operating in the zone `NORD`, such as the Total Forecasted Load.

Many of the reported variables were already introduced in Chapter 5, therefore I will only list the final set of predictors.

- **`HOUR`**: Hour of the day when the bid was submitted. It has the aim to track particular patterns in the offers that may happen in the ramp hours of weekdays, when electricity is highly demanded.
- **`MONTH`**: Month when the bid was submitted. It should bring out any seasonal pattern on the three considered years.
- **`WEEK_DAY`**: Day of the week when the bid was submitted, therefore it is a categorical variable with seven possibilities.
- **`IS_HOLIDAY`**: Boolean variable, it takes the value 1 if the bid was submitted during a holiday (i.e. Sundays and any other Italian festivity), 0 otherwise.
- **`MI_PROFILE`**: This variable is proper to the unit under consideration, it was retrieved by me from the electricity datasets presented in Section 5.2 and it has the aim to track the programmed production of the PU as a result of the markets prior to MSD. Indeed, **it expresses the functioning percentage of the PU after MI**, evaluated as the ratio of the net accepted energy quantity on MGP and MI, and the installed capacity of the PU. Indeed, I have decided to express it as a pure number, since in this way it becomes comparable between different PUs of the same cluster.
- **`TOT_PROD_BEF`**: It reports the Equivalent Operating Hours of production for the PU under consideration, relative to the day before the bid. It has the scope to track periods of inactivity for the PU, which usually lead to "extreme" bids, both upwards and downwards.
- **`MGP_NORD_PRICE`**: Zonal price of electricity on the Day Ahead Market MGP, relative to the Italian bidding zone `NORD`.
- **`DELTA_PUN`**: Price difference between the price of electricity on MGP in Northern Italy and the national *Prezzo Unico Nazionale* (PUN), for the specific hour under consideration.
- **`LOAD_FOR`**: Forecasted load by Terna, it corresponds to the homonymous variable presented in Section 5.3.
- **`GEN_FOR`**: Total Generation Forecast, it corresponds to the homonymous variable presented in Section 5.3.
- **`GEN_FOR_RNW`**: Portion of Total Generation derived from Renewable resources, it corresponds to the homonymous variable presented in Section 5.3.

- **`WATER_RES`**: Water Reservoir (in MWh) in basins of Northern Italy. It is updated weekly, as already introduced in Section 5.3.
- **`MEAN_TEMP`**: Mean temperature in the city of Milan, as presented in Section 5.4.
- **`WIND`**: Average wind speed, as presented in Section 5.4.
- **`PRESS`**: It tracks the atmospheric pressure, and was also presented in Section 5.4.

Regarding the dependent variable, this was chosen to be the MSD GR1 price of the bid in the hour under consideration, dividing BID (**`MSD_GR1_BID`**) or OFF (**`MSD_GR1_OFF`**) types of offers, depending on the model.

It is also worth mentioning that I was originally considering to use as target variable the **price difference** between the submitted price by the Production Unit and the zonal price of electricity on MGP. However, I discarded this option after the preliminary analyses presented in Section 7.1, which highlighted how many operators place their bids without considering the daily oscillations of the PUN on MGP.
Furthermore, in the first models I also included few ***autoregressive* variables**, such as the average of the offered prices between 10 and 20 hours prior to the bids under consideration or the price of the bid placed 24 hours earlier. These data are all available to the operators eligible for MSD at the time they place their bids, however, due to the low variability of the offers on a daily basis, these two variables ended up monopolizing the Variable Importance of every model. Eventually, even if they led to extremely accurate models, I decided to excluded them, trading predictive performances for interpretability.

7.2.3 Tuning the algorithm

The tuning of the algorithm was mainly done by following the suggestions of [5] and consisted in the choice of specific values for the parameters `ntree`, `mtry` and `nodesize`, which have to be specified when using the function `randomForest()` on the software `R`.
Since the problem under consideration comprehends a high amount of data, **I have first studied their optimal values by creating many simple models on single Production Units**, extending the study to cluster of units only in a second instance.

While doing this, I found that the **optimal value for the parameter `ntree`** (i.e. the number of trees in the forest) **strongly depended on the data under consideration** and it was therefore tuned with trial and error on each specific case. Indeed, we know from theory that the OOB error should decrease and stabilize with the number of trees in the forest, thus we would rather use high values for `ntree`. However, when studying clusters of units we usually have hundreds of thousands of observations, therefore

adopting too high a value for the number of trees would result in **extremely long computation times**.
Overall, I settled to **use between 35 and 70 trees in each forest**, depending on the number of observations in the cluster, as well as on the quality of the results.

Regarding the parameter `mtry` (i.e. the number of variables used in the split at each node), I tried many possibilities for its value, including taking `mtry` equal to the total number of predictors, in a *Bagging* setting.
Eventually, I found that **the best option was the one suggested in [5]**, that is to take `mtry` $= \frac{p}{3}$, where p is the total number of predictors. Therefore, I opted to use `mtry` $= 5$ in this setting.

Finally, I found that the **`nodesize` parameter did not affect much the final results of the algorithm**, and I have therefore decided to leave it at the default value for the function `randomForest()`. As explained in [20], in a regression setting this corresponds to a minimum of 5 observations in each leaf.

7.3 Random Forest prediction on the cluster of Thermoelectric Power Plants from ENIPOWER and SEF

The first cluster to be analysed is the one composed by **four power plants from ENIPOWER and SEF** (for a total of nine PUs), introduced in Section 6.4.1 and further examined in Section 7.1.1.
This cluster submits interesting bids of both upward and downward type on MSD, and I will try to interpret both of them, in the following paragraphs.

7.3.1 MSD GR1 Offers of BID type

I will begin with the study of **BID offers on MSD GR1**. To analyse them, I have fitted a Random Forest Random Input model, with a total of 50 trees.

In Figure 7.6 I have reported the **Out Of Bag error** (that was introduced in Section 7.2.1) as a function of the number of trees in the forest. From the graph, we can clearly see that the **error is decreasing** after the first few iterations of the algorithm, finally reaching the extremely low value of 0.51 $(€/\text{MWh})^2$ after 50 trees.
Furthermore, the **percentage of variability explained by the model is more than acceptable**, being equal to 99.31, meaning that the model is performing extremely well on the available data.

Moving on to the interpretation of the results of the algorithm, we can observe the **ranking of the importance of the variables composing the model** from Figure 7.7. In particular, we can observe that the two most important ones are the **Zonal Price on MGP** and the **Mean Temper-**

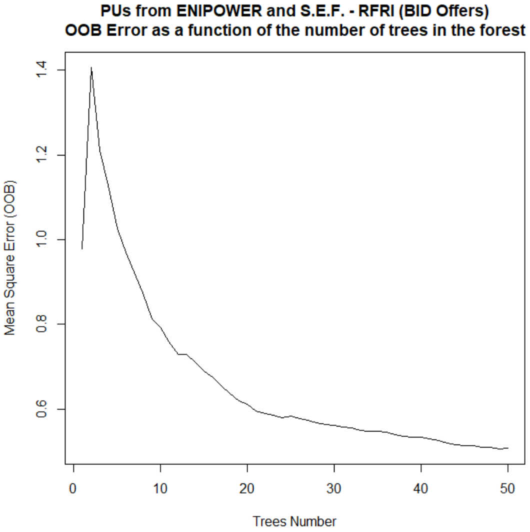


Figure 7.6: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., learning curve of the Random Forest algorithm for MSD GR1 **BID Offers** (OOB error as a function of trees in the forest).

ature in the city of Milan, followed by the variable indicating the **Water Reservoir** in artificial basins used by hydroelectric power plants, as well as the two variables for the **hour and month** of the considered bid.

We could comment these results by saying that the **units within this cluster place their downward bids on MSD considering mainly the price of electricity on the Day Ahead Market**, adapting their bids by looking at many indicators of seasonality, such as `MEAN_TEMP`, `WATER_RES` and `MONTH`, that generally defines the "average values" of the electricity price (daily ramps excluded). On top of this, as already inferred while looking at Figure 7.1, the model highlights the importance of the variable `HOUR`, meaning that the unit adapts its bids on a daily basis, adjusting the offered price in the hours of the day that are usually characterized by higher values of the PUN on MGP.

Moreover, it is also quite interesting to notice that the **units in this cluster do not seem to adapt their BID offers according to their percentage of work after MI**, since the variable `MI_PROFILE` has low importance in the model. This is for sure a peculiarity of this cluster of units, and it is a direct consequence of the decision from the operator Enipower to place almost identical bids for all its units, at every hour. This said, the units in the cluster are characterized by the **highest values for Equivalent Operating Hours**, meaning that they will almost always be active as a result of the markets MGP and MI.

Furthermore, it is interesting to notice how the two variables **Load Forecast** and **Generation Forecast** do not appear to be much useful in the definition of the final price of MSD GR1 bids of BID type.

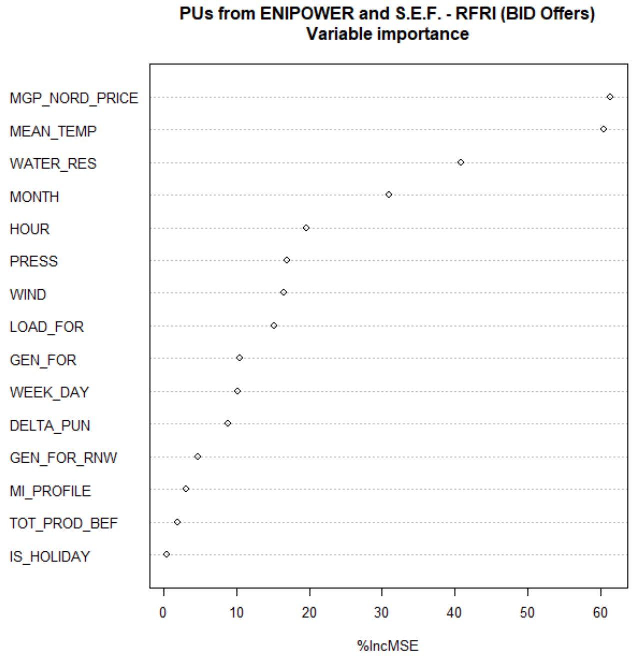


Figure 7.7: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., Random Forest algorithm for MSD GR1 **BID Offers**: Variable Importance.

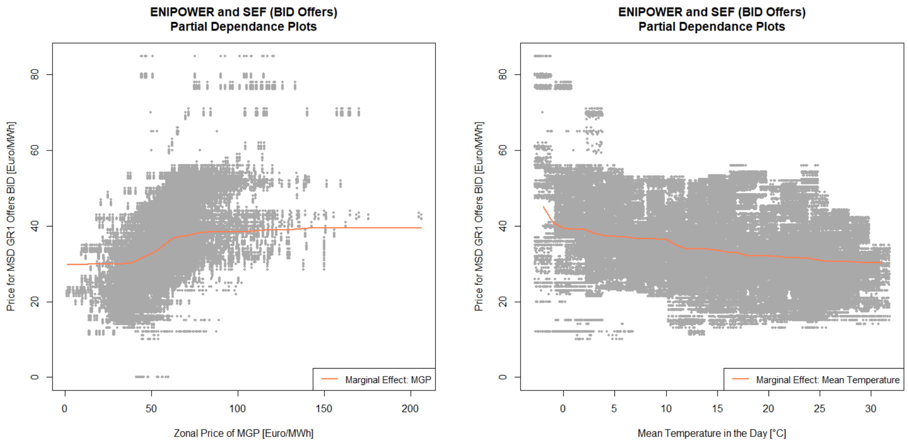


Figure 7.8: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., MSD **BID bids**: visualization of marginal effects via Partial Dependence Plots of the variables **Zonal Price on MGP** (left) and **Mean Temperature** (right).

In Figure 7.8 I have reported the Partial Dependence Plots for the two most important variables: the zonal price of electricity on MGP and the Mean Temperature for the city of Milan.
From the reported graphs, we can have a better idea on the importance of the two variables in the model. From the first one, we can notice that the **the**

price of the bids tend to increase with the zonal price of electricity on MGP, meaning that the operator Enipower tends to increase the prices of its BID offers when the zonal price of electricity on the Day-Ahead Market grows. This is effectively a direct consequence of the fact that the operator tends to keep a **somewhat constant price delta** between the price of electricity on MGP and its downward bids, effectively encouraging the TSO Terna to accept its BID bids, when the price of electricity on MGP is high. Furthermore, from the plot reported on the right-most part of Figure 7.8, we can see that **the price of bids presented by Enipower seems to decrease with the mean temperature**. This fact is particularly true if we consider the colder days in our time span, when we know that the price of electricity is generally higher due to an expected increase in demand. Interestingly, though, we do not observe a similar pattern for the hottest days in the observed three-years period.

7.3.2 MSD GR1 Offers of OFF type

I will now move on to the analysis of OFF offers on MSD GR1 from the cluster of thermoelectric units belonging to the operator Enipower. Also in this case, I have decided to fit a Random Forest Random Input model, with a total of 50 trees. From Figure 7.9 we can observe the value of the **Out Of Bag** error, as a function of the number of trees in the forest. Fortunately, we can observe even in this case that the OOB Mean Squared Error is decreasing with the number of trees, levelling at a value of about 15 (€/MWh)^2. This said, in this specific case the error seem to saturate somewhat earlier than

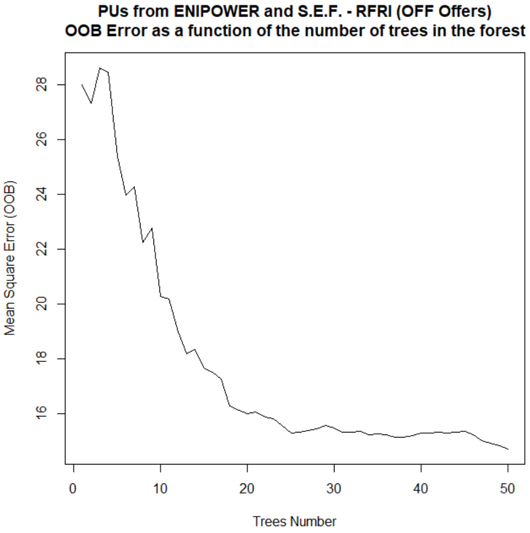


Figure 7.9: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., learning curve of the Random Forest algorithm for MSD GR1 **OFF Offers** (OOB error as a function of trees in the forest).

the previous case, after 35 trees.
The final model presents an **OOB Error** of 14.73 (€/MWh)^2, together with a **percentage of variability explained by the model** equal to 97.61%. These values are slightly worse than those of the model for BID offers, but this is mostly due to the fact that **OFF bids have a wider range than BID ones**, and thus we can consider the model to be highly accurate even in this second case.

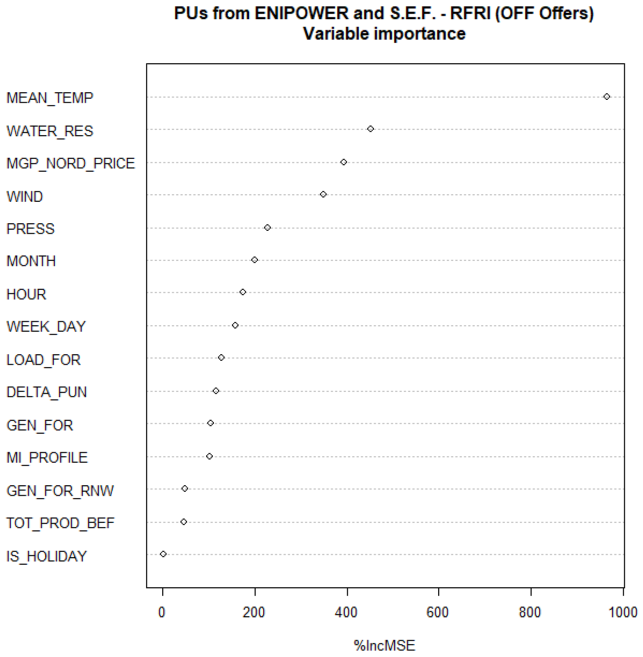


FIGURE 7.10: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., Random Forest algorithm for MSD GR1 **OFF Offers**: Variable Importance.

Looking at variable importance for the model with OFF offers as a target, we can observe from Figure 7.10 that the highest ranked variables are the same than the previous case, meaning that the operator is somewhat consistent for both types of bids. The main differences are represented by the fact that in this second case **the most important variable is the Mean Temperature**, followed by the **Water Reservoir** and the **Zonal Price of Electricity on MGP**. The two categorical variables indicating the month and hour of the bid are still important and are ranked just above the Day of the Week for the offer. Furthermore, there seems to be a much higher percentage increase in MSE, on average, in this second case. This is mainly due to the fact that we are dealing with bids at higher values, therefore the model is more sensible to permutations in the OOB sample[12].

---

[12] Indeed, the package `randomForest()` on `R` has a `scale` parameter to account for the magnitude of the target data for Variable Importance, as well as the standard deviation of trees' errors. However, as suggested in [5], it is considerably better to use the not-scaled version for permutation Variable Importance.

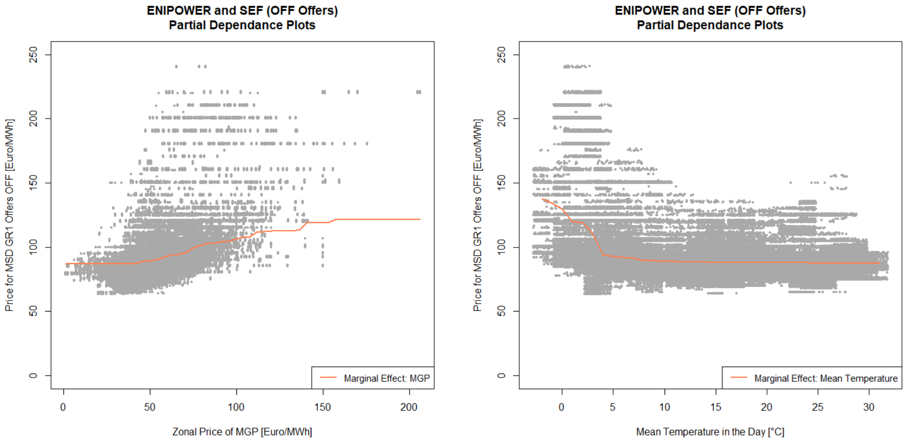


Figure 7.11: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., MSD **OFF bids**: visualization of marginal effects via Partial Dependence Plots of the variables **Zonal Price on MGP** (left) and **Mean Temperature** (right).

To conclude this section, in Figure 7.11 I have reported the visualization of the **marginal effects for the two most important variables** in the model.

From it, we can observe how the Mean Temperature influences the bids in quite a similar way than the previous case, since it is appears to be particularly relevant in the coldest days of the time span, with offers prices that sharply increase in those occasions. Moreover, the price of upward MSD GR1 bids tend to increase with MGP zonal price, in an analogue way to BID offers, meaning that the operator tends to keep its offers within a somewhat defined range from the price of energy on the Day Ahead Market.

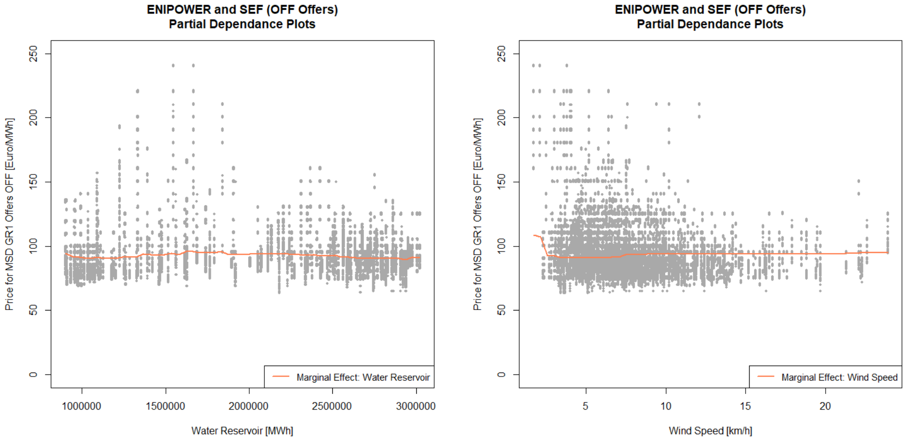


Figure 7.12: Cluster of Thermoelectric PUs from ENIPOWER and S.E.F., MSD **OFF bids**: visualization of marginal effects via Partial Dependence Plots of the variables **Water Reservoir** (left) and **Wind Speed** (right).

Finally, I have also reported the Partial Dependence Plots related to the variables Water Reservoir and Wind Speed, in Figure 7.12. These two graphs are definitely **more difficult to interpret** than the previous ones, since the marginal effects of the two variables do not follow a particular behaviour and seem to be constant on the range of observations. This, though, was partly expected, since those two variables may be important only in combination with other ones, such as the categorical variable related to the month of the offer.

7.4 Random Forest prediction on the cluster of Thermoelectric Power Plants active with BID Bids

In this last section, I am going to predict and interpret the bids from the cluster of **Thermoelectric Production Units active with BID offers on MSD GR1**. These units were introduced in Section 6.4.5 and then further analysed in Section 7.1.4, where I have highlighted how they present quite similar downward bids, but differ significantly for OFF type of bids. Eventually, I have decided to **only fit the model for the prediction of BID offers** for this cluster, focusing on the five Production Units identified by the codes: `UP_LEINI_1`, `UP_NCTLVRNFRR_1`, `UP_TAVAZZANO_5`, `UP_TORVISCOSA_1`, `UP_VOGHERA_1`. In fact, these five units were the ones that reported daily changes for the price of their bids, in contrast to the unit of *Vado Terme*, as I have previously explained on the basis of Figure 6 of **Appendix B**.

I proceeded to fit the Random Forest Random Input algorithm for this specific case, setting the number of trees in the forest to 70. In Figure 7.13 I have reported the **Out Of Bag error** as a function of the number of trees in the forest. From the graph, we can observe that the OOB error is decreasing in a smooth way (much more than the previous cases), eventually reaching the value of 16.80 $(€/\mathrm{MWh})^2$. Regarding the **percentage of variance explained by the model**, we find that it is equal to 90.5%, a value that is lower than the previous case, but still incredibly good if we consider that we are dealing with units from three different operators.

Moving on to the **study of Variable Importance** for the model under consideration, we find that the **highest ranked variables are similar to those of the cluster of Enipower**, with the addition of the one indicating the **Equivalent Hours of Production on the previous day** (`TOT_PROD_BEF`). Indeed, the Equivalent Hours of Production end up being the most important variable in the cluster, causing a worsening of the error of about 100%, when it is permuted within the observations in the OOB sample. Soon after we find the two variables that dominated the previous clusters, namely the Zonal Price of electricity on MGP and the Mean Temperature in the city of Milan.

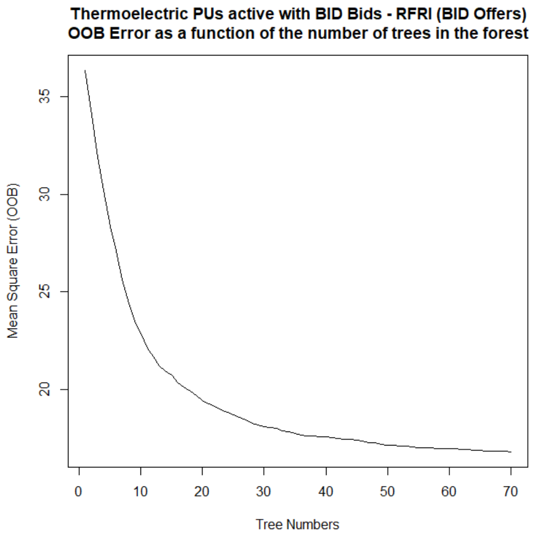


FIGURE 7.13: Cluster of Thermoelectric PUs active with BID bids, learning curve of the Random Forest algorithm for MSD GR1 **BID Offers** (OOB error as a function of trees in the forest).

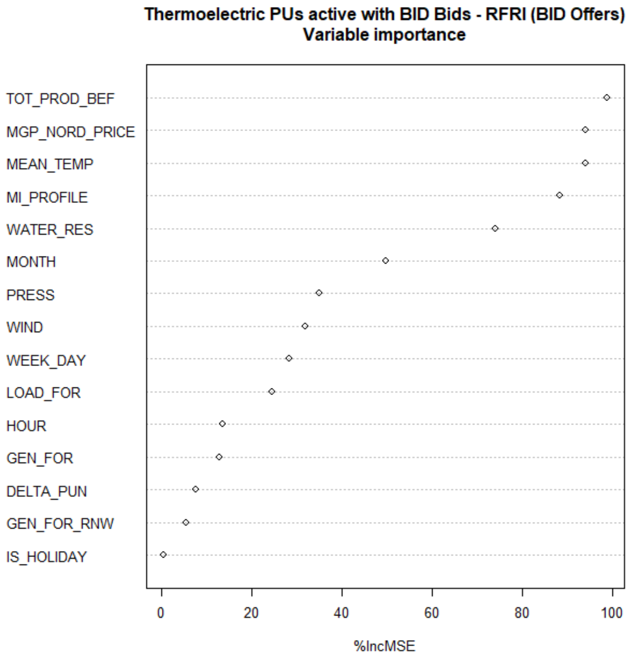


FIGURE 7.14: Cluster of Thermoelectric PUs active with BID bids, Random Forest algorithm for MSD GR1 **BID Offers**: Variable Importance.

Moreover, in this case it is quite important also the variable `MI_PROFILE`, related to the production percentage of the units as a result of the Intraday Market (MI).

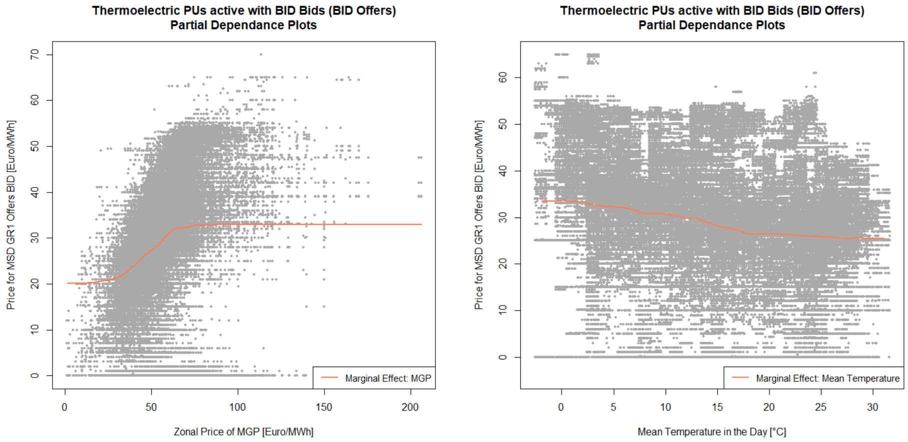


FIGURE 7.15: Cluster of Thermoelectric PUs active with BID bids, MSD **BID bids**: visualization of marginal effects via Partial Dependence Plots of the variables **Zonal Price on MGP** (left) and **Mean Temperature** (right).

To dive deeper in the interpretation of the variables, I have reported the **Partial Dependence Plots of the main variables under consideration**, in Figures 7.15 and 7.16. From the first two graphs, we can appreciate, respectively, the effects of the Zonal price of electricity on MGP and the Mean Temperatures, over the bids on MSD GR1.
In particular, we can notice that the bid price tend to increase with the zonal price on MGP, exactly as for the operator Enipower. Furthermore, even in this case, we can observe a prominent effect of the variable Mean Temperature, with offer prices on MSD GR1 - BID that decrease, on average, as the climate gets hotter.

Regarding Figure 7.16, we can start by noting that the Partial Dependence Plot highlights a great decrease in the offered price when the unit had almost no Equivalent Hours of Production in the previous day, effectively suggesting the initial hypothesis that the model is correctly using this variable as an **indicator for prolonged periods of inactivity**, usually characterized by null BID offers.
Regarding the graph reported in the top-right part of Figure 7.16, we can notice a similar behaviour to the previously presented one, also with respect to the variable `MI_PROFILE`. Indeed, **the price offered tend to decrease when the production unit is accepted for low quantities on MGP**, suggesting that the units tend to "protect" the amount of energy to be produced, especially when this quantity is low.
Lastly, I would like to talk about the variable Water Reservoir. As explained for the previous cluster, it is quite hard to interpret, since it is a proxy for seasonality and its values result in being "uninformative" if not seen on time. However, what we can observe from the graph reported in the bottom part of Figure 7.16 is that **there seems to be a notable increase** in the

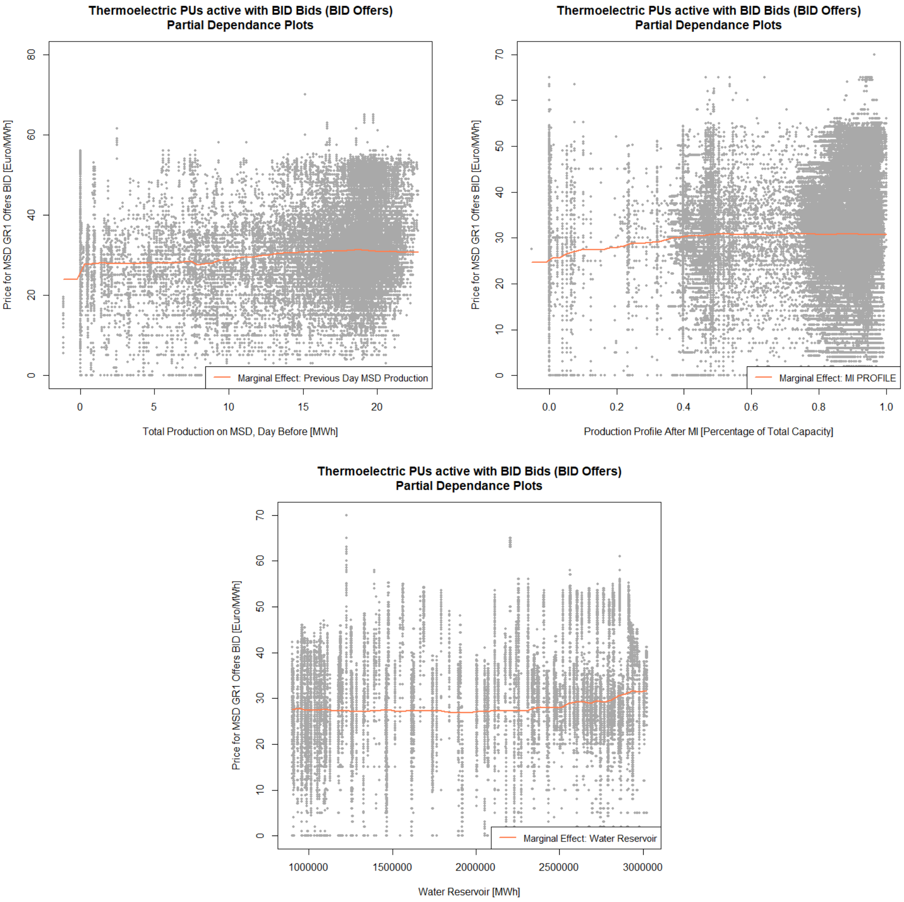


FIGURE 7.16: Cluster of Thermoelectric PUs active with BID bids, MSD **BID bids**: visualization of marginal effects via Partial Dependence Plots of the variables **Total Production in Previous day** (top left), and **Production Profile after MI** (top right), **Water Reservoir** (bottom).

price of the bids in the weeks characterized by great amounts of water in Northern Italian basins.

Lastly, it is interesting to notice from Figure 7.16 that the units in the cluster have extremely high values for Equivalent Operating Hours of Production, and tend to leave MI quite often at high capacity.

Future Analyses 8

The following chapter is dedicated to the introduction of few complementary topics that emerged during the course of my work and that I believe they may be **worth to consider in future analyses**.

I have decided to divide the topics presented in the following pages into two different sections. The first one is quite general, and has the aim to present a list of possible issues that could be tackled to **further refine the results obtained** in this work.

Eventually, Section 8.2 is dedicated to the introduction of the study of the **acceptance of bids on MSD by the TSO Terna**. This subject was already partly studied in [4], and I decided to dedicate a separate section to this topic, since I believe it has **crucial importance for future analyses**. For this reason, after a brief introduction, I will include two different preliminary analyses that I conducted during my work, in that direction.

8.1 An overview of suggested future analyses

The first major achievement of this work was the **construction of an extensive dataset** regarding the three main Italian electricity markets, which me and Sara Martucci built in the first months of internship at RSE. In this work, I mainly have used these datasets to analyse the market MSD in the Italian bidding zone `NORD`, however, the collected data comprehend also the bids from the markets MGP and MI, as well as all the other Italian zones.

For this reason, the first straightforward extension of this work should regard a **comparison of the bidding zone `NORD` with the rest of the *peninsula***. This analysis would be useful for mainly two reasons: on one hand, it should identify all the contrasts coming from the structural differences of the various zones[1] and, on the other hand, it would be useful to

---

[1] For instance, we know that in the southern part of Italy are located the major renewable power plants based on wind and solar technology, and these power plants

identify how each operator manages its power plants located in different bidding zones.
Just to give an example of the second goal, the **operator Enipower** (which I have analysed in Section 6.4.1) manages also one PU in the zone `CNOR` (`UP_NPWRLVORNO_7`) and three more units in the zone `SUD` (`UP_NPWRBRNDSI_8`, `UP_NPWRBRNDSI_9` and `UP_NPWRBRNDSI_10`). It would be extremely interesting to study the offer curves from these other PUs, highlighting any differences that may occur between these units and those in the zone `NORD`.

Keeping the focus on the market for ancillary services, it would be quite interesting to **expand the temporal horizon of the analyses** conducted in this work. Indeed, my work was focused on the years 2017 to 2019, therefore prior to the **Covid-19 pandemic**, which sharply decreased the demand of energy in Italy[2], at the same time increasing drastically the share of energy produced by renewable resources in the lockdown months. In this regard, it would be extremely interesting to assess how all the Production Units that placed constant bids in the months analysed in this work, **reacted to the sudden change in energy demand**, and to compare the results with all the units with more "complex" bids.

Another possible extension of my work on MSD, would be the **definition of more complex models, accounting for other ancillary services** rather than only the GR1. Indeed, in Chapter 7 I have decided to predict the "first step" in increase or decrease of energy production, but this (even if it accounts for the greatest portion of energy exchange on MSD) is only one of the services required by Terna. In fact, a more refined model would comprehend also the service denoted as "AS"[3] in our dataset.
Indeed, the units from the operator Enipower, which seem to place the best possible GR1 bids, have almost no accepted offers of the AS type. In contrast, the units from ENEL and A2A have almost the same amount of GR1 and AS bids accepted by Terna, and they often seem to begin their production only on MSD, thanks to the acceptation of AS+OFF bids. If confirmed, this hypothesis may indicate a profound difference in the approach of the units to the market MSD, that may be driven by the specific technological characteristics of each power plant.

Eventually, I also suggest to **extend the analysis of MSD to all the smaller units** that me and Sara Martucci have decided to exclude from this work. Indeed, up to now they only make up a negligible portion of the accepted bids on MSD, but in the future the expectations are that they will be substantially more relevant for ancillary services, as I have anticipated in

---

export massive amounts of energy to the North during the summer peaks of renewable energy production.

[2] I refer to the **International Energy Agency**'s website for an in-depth overview on the topic: https://www.iea.org/reports/covid-19-impact-on-electricity.

[3] i.e. Switching on or off the power plant, from the Italian *Accensione o Spegnimento*.

Chapter 5.

Other than refining the analyses on MSD, I believe that in future researches it would be interesting to **move the perspective** from the activity of many PUs on one fixed market, **to the study of the activity of a *single* Production Unit across all the different markets**.
By doing so, I suggest to begin with the **analysis of the clusters presented in Chapter 6**, as well as the results of the K-Means algorithm presented in **Appendix A**.
This new perspective would be particularly interesting for all the units that resulted quite active in other markets, such as the units of ENEL for the Intraday Market (MI).

Furthermore, it would also be interesting to study more in depth all the units presenting a similar situation to the cluster of **Hydroelectric PUs that switched management from ENEL to Alperia**, presented in Section 6.4.3. Indeed, a more profound analysis of those situations, extended on more years, would be crucial to understand the different management strategies adopted by two distinct operators.

8.2 Studying the acceptance process of bids on MSD by Terna

The Italian Market for Ancillary Services (MSD) is by far the most complex market among the three ones considered in this work. In Section 4.1 and Section 3.3.3 I have introduced MSD, explaining in depth the mechanics at the basis of its functioning and its rewarding mechanisms.

What I have not discussed in details yet, is the process of **acceptance of bids by the TSO Terna**. Indeed, up to now I only have talked about the fact that Terna acts as a central counterparty in MSD: it evaluates the needs of the power system and it accepts the *best* bids on the market, in order to cope with any potential hazards occurring to the power grid. As previously stated, accepted bids are then remunerated at the price offered, via a pay-as-bid procedure.

It is clear that the concept of "***best bid** to cope with potential hazards*" may be subject to different interpretations and, at the same time, only the TSO has the required knowledge to understand the real needs of the power system on a hourly basis.
Nevertheless, this issue may create many problems to the Production Units operating on MSD, which may find it difficult to place competitive bids on the market.

Indeed, the **difficulty to understand which bids may be accepted by Terna**, may be at the basis of the "counter-intuitive" behaviour shown by many operators, which decide to place constant bids on MSD for prolonged periods of time.

To show the complexity of this topic, I will present a preliminary analysis over the acceptance procedures by Terna for MSD bids of OFF type. Eventually, I will conclude the chapter by observing that, to further complicate this setting, it looks like that the TSO itself changes the rules underlying its acceptance procedures, at least in the time span of this work. To prove this claim, I will talk about the **acceptance rates of bids from hydroelectric power plants**, in late 2019.

8.2.1 The importance of *Grid Supply Points* in the acceptance of GR1-type bids on MSD

In this work I have focused my attention to only the **GR1-type bids on MSD**. This approximation was made in order to ease the study of an already complex market, but it was still a reasonable one, since most of the accepted bids on MSD are of the above-mentioned type.

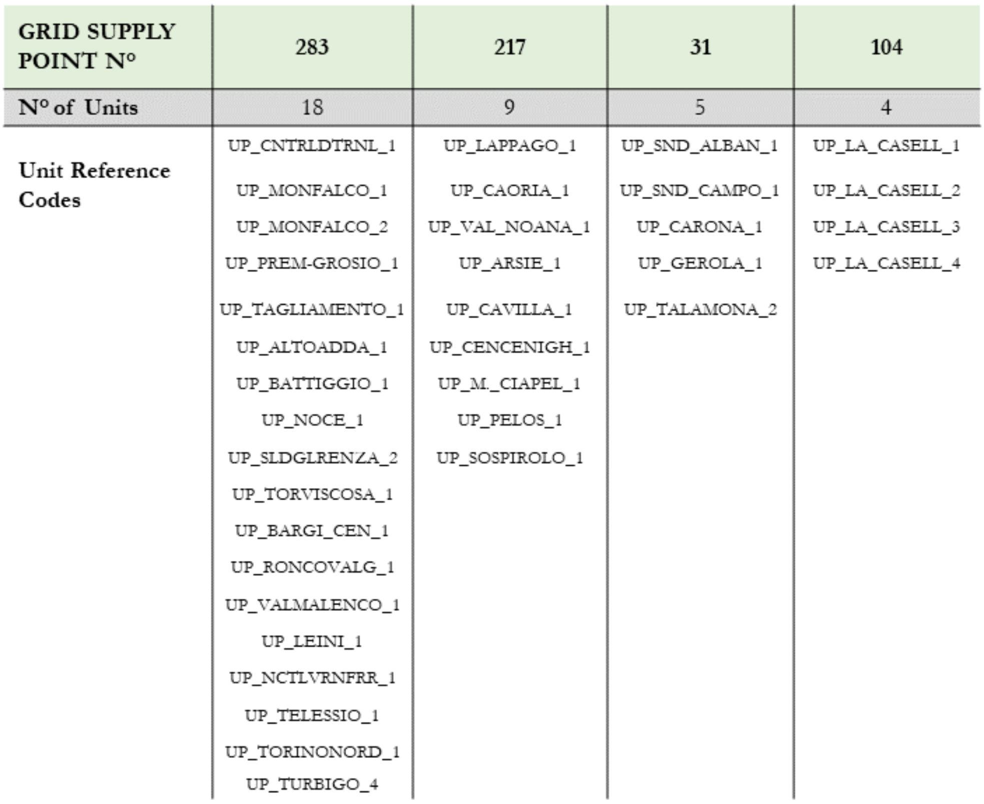
Consequently, during my work on MSD, I wanted to reach a deeper understanding of how GR1 bids were accepted by Terna. To do so, I have decided to perform a *transversal* preliminary analysis of MSD, done by filtering the dataset for all the GR1 bids at various hours of the selected time span.

Indeed, even a brief analysis confirmed the fact that **MSD is far from having a merit order basis**, since at every hour we can observe accepted GR1 bids that have very different prices, for both upward (OFF) and downward (BID) types of bids. Furthermore, it was interesting to notice that the highest possible BID bid, as well as the lowest possible OFF bid, which are **the most competitive bids, were almost never accepted**, confirming the initial hypothesis.

This said, I found an interesting pattern in the acceptance process of bids on MSD by Terna, related to one of the variables in the electricity markets' datasets. The variable in question is the one identified as **`GRID_SUPPLY_POINT_NO`** in Chapter 5 and looks to be the only significant one in explaining the acceptance of a portion of bids on MSD.

This variable, however, is **quite obscure in its interpretation**. On its website, the Italian *Gestore dei Mercati Energetici (GME)* specifies how the grid supply point is "a numerical code that indicates the relevant exchange point associated to the unit" [10], suggesting that this variable is used to identify some sort of *electrical proximity* within the units sharing the same code. Unfortunately, it also states that the definition of GSPs is done by Terna and, even after an in-depth research on Terna's website, I **could not find more specifics at the base of the divisions**.

In Figure 8.1 I have reported the four major "relevant exchange points" related to Northern Italy and for each of them I have listed the Unit Reference Codes of the PUs in it. From the picture, we can observe that **only 4 GSPs**



GRID SUPPLY POINT N°	283	217	31	104
N° of Units	18	9	5	4
Unit Reference Codes	UP_CNTRLDTRNL_1	UP_LAPPAGO_1	UP_SND_ALBAN_1	UP_LA_CASELL_1
	UP_MONFALCO_1	UP_CAORIA_1	UP_SND_CAMPO_1	UP_LA_CASELL_2
	UP_MONFALCO_2	UP_VAL_NOANA_1	UP_CARONA_1	UP_LA_CASELL_3
	UP_PREM-GROSIO_1	UP_ARSIE_1	UP_GEROLA_1	UP_LA_CASELL_4
	UP_TAGLIAMENTO_1	UP_CAVILLA_1	UP_TALAMONA_2	
	UP_ALTOADDA_1	UP_CENCENIGH_1		
	UP_BATTIGGIO_1	UP_M._CIAPEL_1		
	UP_NOCE_1	UP_PELOS_1		
	UP_SLDGLRENZA_2	UP_SOSPIROLO_1		
	UP_TORVISCOSA_1			
	UP_BARGI_CEN_1			
	UP_RONCOVALG_1			
	UP_VALMALENCO_1			
	UP_LEINI_1			
	UP_NCTLVRNFRR_1			
	UP_TELESSIO_1			
	UP_TORINONORD_1			
	UP_TURBIGO_4			

FIGURE 8.1: Specifics of the major Grid Supply Points in the `NORD`.

have at least four units in them, with the number 283 that has by far the greatest amount of units inside.

Focusing on the Grid Supply Point 283, I found that the units it comprehends are **located quite far one from the other geographically**[4], which further complicates the comprehension of this division[5]. Furthermore, the **GSP 283** comprehends quite a variegate set of units, since it includes:

- 8 thermoelectric PUs
- 4 pumped-storage hydroelectric PUs
- 4 dispatchable hydroelectric units of relative big dimensions (>51MW of installed capacity)
- 4 smaller hydroelectric units (≤51MW of installed capacity)

Conversely, the **Grid Supply Point numbered 217** comprehends mainly dispatchable production units of hydroelectric technologies, with

---

[4] Indeed, we have units from the provinces of *Torino*, *Vercelli*, *Udine*, *Milano* and much more, ranging in almost all the regions of the zone `NORD`.

[5] I have actually tried to understand further the reasons behind the division in Grid Supply Points by Terna, but did not find any valid conclusion. Indeed, the technology of the units inside a GSP may vary, as shown by the GSP 283. Some GSPs include all the units from a single power plant, like the one numbered 104 for *La Casella*. On the opposite, though, the power plant of *Sesto San Giovanni* has two Production Units that are divided into two different GSPs, which seems quite counter intuitive.

dimensions between 20 and 55 MW and belonging mainly to the operator ENEL.

Quite surprisingly, when I analysed the bids within the Grid Supply Point number 283, I found that **almost all the accepted OFF bids followed a merit-order based on the price of the offer**.
Indeed, it is almost straightforward to predict which MSD OFF bids will be accepted inside that Grid Supply Point, since Terna's procedure works more or less as follows:

1. At first, Terna evaluates if there is the **need to call some units from this GSP upward** on MSD. This, of course, might not always be the case, and I have to stress that from solely our data we have no capabilities to predict when this need will be percieved by the TSO.
2. Terna **orders the OFF bids** from the GSP 283 based on their price, from the lowest one to the highest one.
3. Terna **discards all the units that are ineligible to produce** for the hour in consideration. This step excludes all the units that may already be at their maximum capacity, the units that are currently not producing as a result of prior markets and those undergoing maintenance periods.
4. Eventually, the TSO **accepts the most convenient bids** among the feasible ones, in such a quantity to fulfill the original needs.
   Sometimes, if the cheapest unit can not produce the amount of energy sufficient to cope with the demand, Terna accepts directly (and only) the bid made by the second cheapest unit.

During the final stages of my work, I was able to study in detail the upward bids from solely the GSP 283, but I managed anyway to verify a **similar behaviour also on the GSP 217**.

Overall, from this preliminary analysis it is impossible to reach meaningful conclusions, but the above consideration looks to be quite insightful to partially understands the acceptance dynamics on MSD.
Indeed, if this hypothesis was confirmed, it would imply the **existence of segments of units that are effectively in close competition** to produce on the **highly remunerative market MSD**.

To conclude this section, I believe that much more research is needed on the topic, which, however, I have to leave to future analysis. Indeed, in upcoming researches it will be necessary to understand better the meaning of the variable Grid Supply Point and the mechanics at the basis of the segmentation. I also suggest a study on the bids of BID type, to see if they present a similar behaviour to upward bids. Eventually, it would also be interesting to search if there exist **other segments of units competing on MSD on the basis of a merit order**, which may not be included in a single Grid Supply Point.

8.2.2 Acceptance of MSD bids by Hydroelectric Units

I have stressed many times in this work how it is quite hard to study MSD, due to the lack of real-time information behind the specific needs of the Transmission System Operator[6] at a given hour of the day.

This, however, is not the only layer of difficulty when studying the problem of acceptance of bids on the market, since my analysis has also revealed that sometimes the TSO **Terna varies the criteria** according to which it accepts offers in the Ancillary Services Market.
In this section, I would like to introduce a practical example of this consideration, talking about the **sudden change in the acceptance of MSD bids from Hydroelectric Units**, that happened in the second half of 2019.

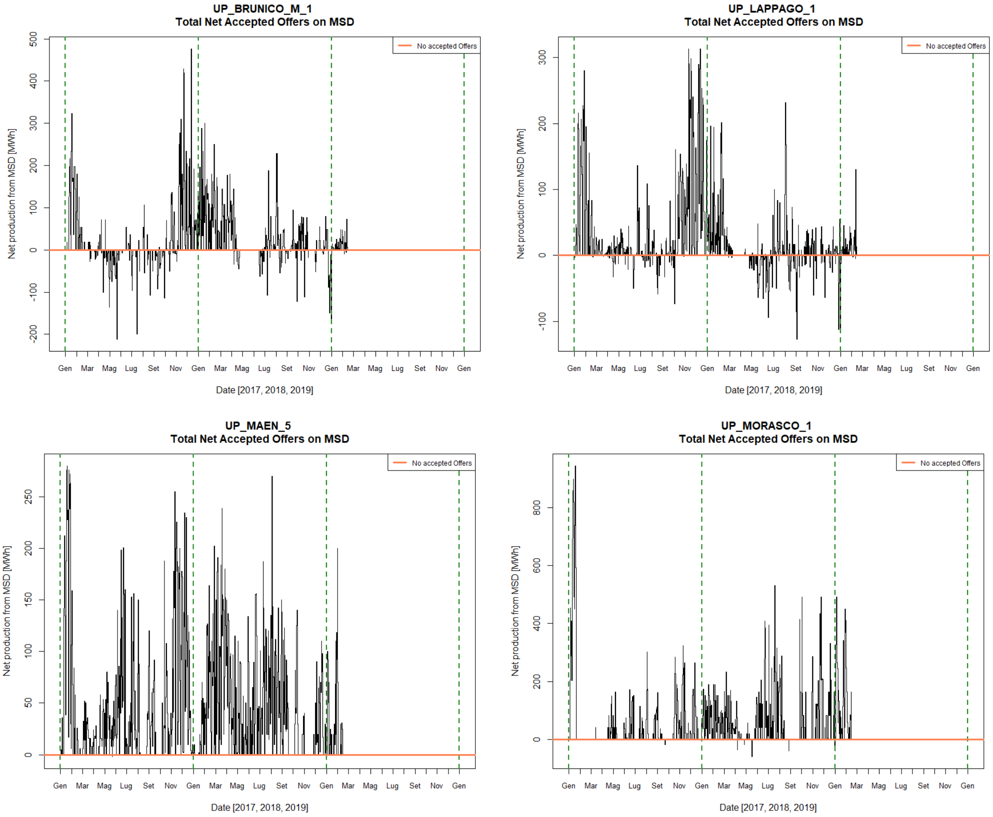


FIGURE 8.2: Daily net production quantity from accepted bids in MSD, for the PUs of *Brunico* [Alperia], *Lappago* [Alperia], *Maen* [CVA Trading] and *Morasco* [ENEL] (left to right, top to bottom). The dashed green lines separate the offers of the years 2017, 2018 and 2019.

The four pictures in Figure 8.2 report the **net daily energy production (in MWh) coming from accepted bids on MSD**, for a selected sample

---

[6] To be more precise, the potential hazards related to the Power System, which the TSO has to prevent and manage in advance by accepting offers in MSD.

of hydroelectric units belonging to different operators. The target variable, plotted on the y-axis, is obtained as the sum of the energy quantity from all the accepted OFF offers on MSD, to which I have subtracted the total quantity of energy from all the accepted BID offers on MSD.

From these graphs we can see how there is a sudden decrease in the activity on MSD by the reported units, since their bids stopped being accepted after February 13^{th}, 2019. This is not an isolated behaviour, since a further analysis highlighted how 25 out of 31 dispatchable hydroelectric units in the North zone do not receive acceptance in MSD after that date, even if **their offers' prices remain effectively unchanged**, as we can see from Figure 8.3. Furthermore, I managed to identify this behaviour also in other zones, such as Center-North and Center-South, with very similar considerations and starting from the same date.

I had the occasion to discuss this topic with experts from RSE, but we all struggled to find an explanation for it, since we could not find any official note from Terna related to this evident change.
The topic of acceptance of bids by Terna is quite hard to study, and it should require **more and more profound analyses in the future**, starting with a verification of the persistence of this phenomenon also in the year 2020. Nevertheless, I found it quite interesting to report a practical example of an evident **change in behaviour** from Terna, which effectively confirms the difficulty to operate in MSD also for the Production Units, which may experience sudden changes in the amount of accepted bids.

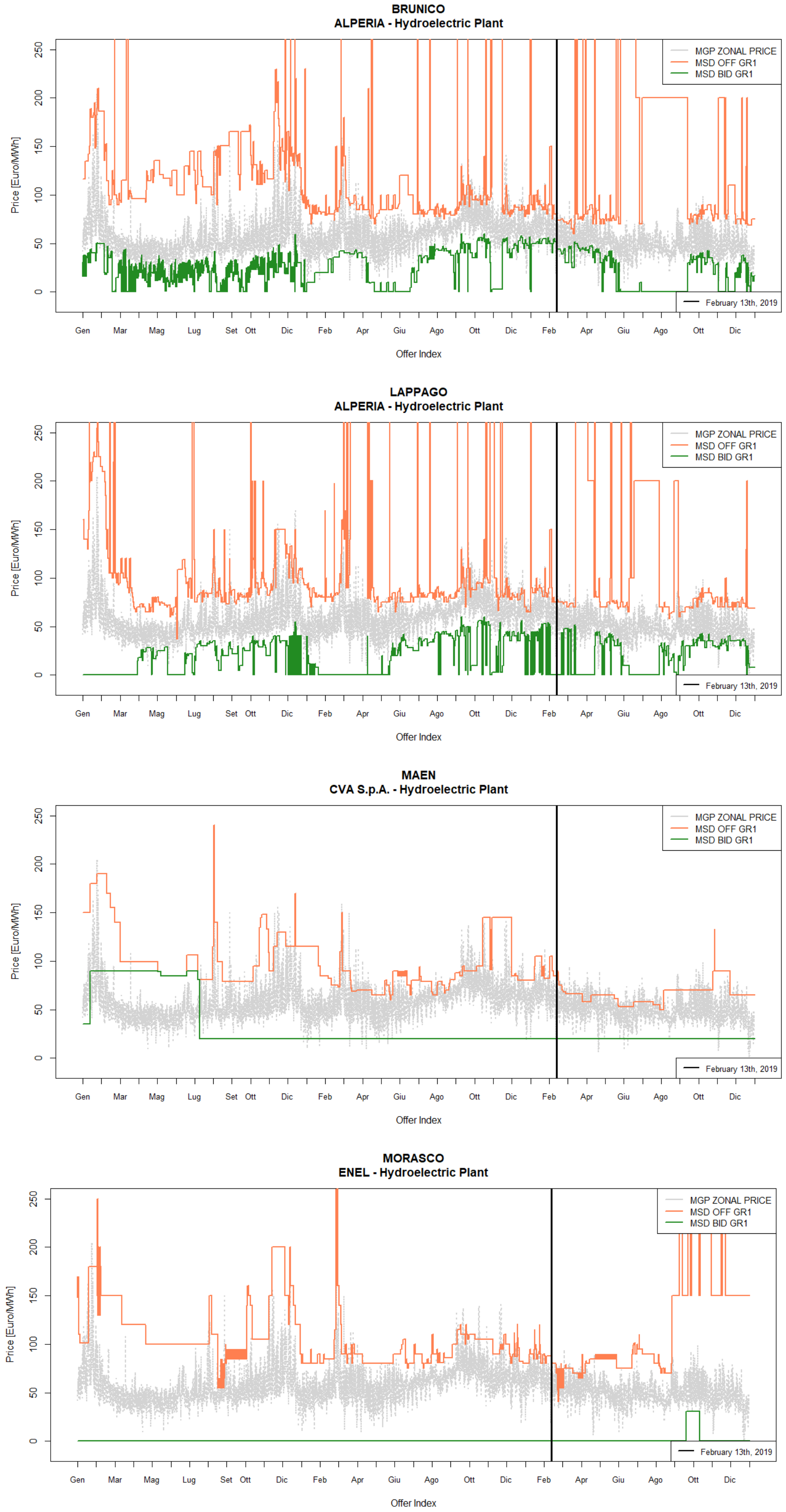


FIGURE 8.3: MSD GR1 offers in the years 2017 to 2019, from the hydroelectric units of *Brunico*, *Lappago*, *Maen* and *Morasco*

Conclusions 9

The liberalization of energy markets gave birth to extremely competitive environments, where operators can bid in strategic ways, possibly even at non-cost-reflective prices, in order to maximise their revenues from *all* the units they control. In the past years, we have also observed profound changes in the dynamics within the electricity markets, especially driven by the rise of Renewable Energy Sources, with units operating in these environments that need to be extremely flexible, in order to adapt to ever-changing conditions. Furthermore, a lot of attention was drawn on electricity markets lately, principally due to the sharp rise in the price of electrical energy, that we have experienced in the month of October 2021, with prices on the Day-Ahead Market (i.e. the so-called PUN) that more than doubled in value from those considered in the period of this analysis, which ranged over the years 2017 to 2019.

For all the aforementioned reasons, it becomes extremely important to have a profound understanding of all the dynamics within electricity markets. In this work, I focused on the Ancillary Services Market (MSD), which is one of the most complex ones, due to the fact that it is not based on a merit order and it is definitely not trivial to understand the specific needs of the Transmission System Operator, at every hour.

The analysis of MSD highlighted major differences between the strategies adopted by different operators, that tend to approach the market in completely different ways.

In general, though, I found out that the great majority of Production Units powered by Renewable Energy Sources, especially those with limited installed capacity, tend to submit non-competitive bids on MSD, since they generally offer at constant prices for many weeks, or even months. On the opposite, thermoelectric power plants seem to be more interested in taking part in the market, which they probably see as a potential way to cope with their limited revenues from the markets that are based on a merit order, where the price of electrical energy is lower.

In Chapter 6, I have proposed a classification of the units operating in the bidding zone `NORD`, and I have highlighted how the majority of the considered power plants update more frequently their OFF bids than the BID bids, with the extreme examples being the clusters of hydroelectric units of the operator ENEL, together with the thermoelectric units from A2A. This fact was interpreted with the idea that upward bids on MSD have a much greater strategic importance for Production Units, since they imply greater revenues than the net balance that could be earned through downward BID bids, considering also prior markets.

In general, the units eligible for MSD place their bids only considering few variables, that are generally limited to the price of electrical energy on MGP, as well as some unit-specific conditions, such as the production profile after the market MI and, for some hydroelectric Production Units, the amount of water in their basin. Bids are also corrected in maintenance periods, where we observe BID offers at a null price and OFF offers at quite high prices, usually equal to the old maximum value of 500 €/MWh.

Focusing on thermoelectric operators, I have analysed more in-depth the bids from four clusters with this technology, in Chapter 7. In that occasion, I confirmed the presence of common strategies for all the units of a given operator, even though these strategies appear to be quite different from one company to another.
Indeed, the company Enipower places the same bids across all the units it controls, with a high degree of variability and bids that change on a hourly basis, but without taking too much care of the working range after the Intraday Market of its Production Units.
In contrast, the operators A2A and ENEL seem to differentiate more their bids across the units they control, but they update them less often than the cluster from Enipower. Specifically, the operator ENEL places its bids on a daily basis (i.e. constant prices for 24 hours straight), thus disregarding completely the forecasted load and the consequent fluctuations of the MGP and MI prices.

For two of the identified clusters, I was able to predict with incredible accuracy the submitted bids, thanks to the algorithm Random Forest. This analysis confirmed the fact that most of the considered variables do not influence the offered price, as is the case for the forecasted load and generation from Renewable Energy Sources. Nevertheless, the bids on MSD almost always follow the seasonal patterns of the energy price on MGP, as confirmed by the importance acquired by the variable "Mean Temperature", in the Random Forest models.

To conclude, most of the offer curves for the service GR1 on MSD appear to be somewhat perfectible, with many plants that do not care to bid at competitive prices at every hour. This is probably due to the difficulty of the operators to predict when there will be an effective need for their service on MSD. Thereby, future research should focus on the prediction of the need for each ancillary service at every hour, also explaining the acceptation dynamics on MSD from the Transmission System Operator Terna.

Appendix A

K-Means clustering of the PUs active on MSD for the `NORD` zone

In Chapter 6 I have introduced the clustering on the macro-behaviour of production units in the Italian bidding zone `NORD`, discussing the motivation behind this approach and the final clusters I have obtained after a subsequent refining process. In this first Appendix chapter, I will go back to the **K-Means clustering analysis**, explaining in depth its **application, tuning process and unrefined results**.

I decided to work with the **K-Means algorithm**, since it is one of the most established clustering algorithms among the available ones, with its original idea that dates back to the mid 1950s. Furthermore, it can also be very quick to assess the quality of the results provided by this methodology, mainly by looking at the position of the **centroids** of the identified clusters. To perform the analysis I adopted the programming language **R** which has already built in the function *`kmeans()`*, inside the package "stats" [18].

Variables used for the clustering

I report here a brief summary of the variables adopted in the final version of the algorithm, referring to Section 6.2 for a complete description of their meaning and their selection process.

- **`TOT_PROD`** := It is the variable related to Equivalent Operating Hours of production, which is obtain by dividing the total amount of energy produced by the Installed Capacity of the power plant;
- **`TOT_INC`** := Total net revenues from MGP, MI and MSD combined, divided by the Installed Capacity;
- **`MGP_OFF`** := Total revenues from MSD bids of OFF type, divided by the Installed Capacity;
- **`MI_BID`** := Total expenses from MI bids of BID type, divided by the Installed Capacity;
- **`MI_OFF`** := Total revenues from MI bids of OFF type, divided by the Installed Capacity;

- `MSD_BID` := Total expenses from MSD bids of BID type, divided by the Installed Capacity;
- `MSD_OFF` := Total revenues from MSD bids of OFF type, divided by the Installed Capacity;

Tuning the algorithm K-Means

The following procedures were necessary, in order to fine-tune the K-means clustering algorithm:

1. **Scaling the variables**: This is usually common practice when dealing with equally-important variables. This case is borderline, since I have already weighted the variables with a division by the variable Installed Capacity, however, I still have 6 variables related to the economic sphere (the income in the various markets), while the last one is related to production quantities and it has a different scale.
   In any case, having tried both the two approaches, I realized that the best results were obtained with scaled variables, option I adopted in the final clustering analysis.
2. **Choosing the number of centroids K**: This is a crucial step in the clustering algorithm K-means and it consists in the choice of the final number of clusters to be retrieved by the algorithm. In our case, we do not know *a priori* this number, therefore this choice is quite delicate and I dedicate to it the following section.
3. **Choosing the parameter *nstarts***, which corresponds to the number of random configuration initially chosen by the algorithm [18]. I found quite problematic the choice of this parameter, therefore I decided to choose it fairly large and equal to 100000. This was enough to make sure that the algorithm explored a great variety of initial configurations, so that it could eventually choose the globally best one.
4. **Setting the seed** to a known value, in order to make the analysis reproducible in a second instance.

Choosing the parameter K

As anticipated, the choice of the parameter K is crucial for the optimal tuning of the algorithm. In this case **we do not have prior knowledge on the number of clusters** behind the phenomenon, since the preliminary analyses done with Sara Martucci only showed the existence of a couple of well-defined clusters, but left unclassified the greatest portion of PUs.

One possible way to **define the optimal value for the parameter K**, is based on a sort of "greedy" approach, that fits many iterations of the algorithm K-Means, trying each time to look for a different number of clusters (i.e. ranging over possible values for K).

At each iteration it is studied how the data are located with respect to the centroids of the final clusters, assessing how the **variability within each cluster** changes each time.

I decided to follow the same approach for this problem, and I have reported the results in Figure 1. In our specific case, I tried each value of K from 1 to 30, reducing momentarily the parameter *nstart* in order to obtain faster results[1].

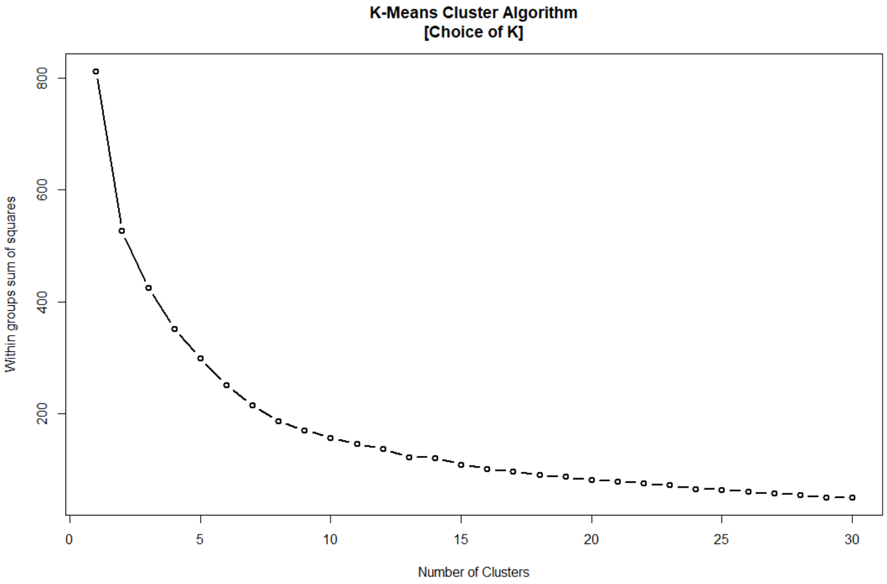


FIGURE 1: Tuning the algorithm K-Means: Within group sum of squares vs. Number of Clusters

The graph in Figure 1 reports the trend of the **within groups sum of squares** (evaluated as the sum of the squared distance between every point and its centroid) against the number of clusters K, and we can see that their relationship is clearly an inverse one. This is expected, since I have evaluated the within groups sum of squares as the sum of the overall variability of every cluster, and this variable tend to decrease naturally when the number of centroids increase, since the data will become closer on average to the closest centroid. Unfortunately, **we do not see from the graph any clear elbow**, that usually indicates the number of effective groups behind the data[2].

In our case, the choice of K results being quite a hard one, since from Figure 1 the optimal value would be between 8 and 14 groups, but every cluster fitted with these value ended up being too messy and presented sub-optimal behaviours. For this reason, **I decided to increase the number of clusters**, first to 15 and **eventually to 20**, observing a clear improvement of the results and more defined clusters.

---

[1] Indeed, these will not be the final models, rather part of the tuning process of the algorithm.

[2] In case, of course, that this number is known and fixed a priori.

In the following section, I will introduce and comment each of the 20 final clusters identified by the algorithm.

Analysing the results

I will now comment each of the 20 identified groups, listing the units in each cluster and adding a personal interpretation of the reasons why the algorithm decided to group those units together. By doing so, I will often **refer to the final clusters presented in Section 6.4**.

Moreover, I would like to stress the fact that **I did not feed to the clustering algorithm the variables technology and operator**, which I have decided to use afterwards, in order to interpret the given clusters. However, as I will show, many of the obtained clusters will be characterized by the presence of units from a single technology or operator, which confirms the goodness of this approach.

For the sake of clarity, I also specify that I will present the clusters in the original order given by the algorithm, knowing that any permutation of the labels will not change the underlying meaning. Furthermore, I have reported at the end of **Appendix A**, in Table 1, a **summary of the position of the centroid** of each cluster of units, with respect to the seven variables used by the algorithm K-Means.

1. **CLUSTER 1**

   REFERENCE CODES:
   `UP_EDOLO_1, UP_ETQCHIOTAS_1, UP_GARGNANO_1, UP_RONCOVALG_1, UP_S.FIORANO_1, UP_SOVERZENE_2, UP_TELESSIO_1`

   The first cluster groups a set of 7 Hydroelectric power stations, 5 of which are **Pumped-Storage hydroelectric power plants from the operator ENEL**. This cluster, together with cluster 12, inspired the definition of the analogue cluster of Pumped-Storage Hydroelectric power plants from ENEL Produzione, presented in Section 6.4.7.
   Indeed, by the position of the centroids of this group and that of cluster 12, we can affirm that the K-Means algorithm grouped these units together since they present the lowest value of Equivalent Operating Hours, as well as total revenues per unit of installed capacity. This consideration makes sense if we consider the specific technology of these production units, since **Pumped-Storage hydroelectric power plants only produce energy in very specific hours of the day**, when electrical energy is more demanded and has higher prices.

2. **CLUSTER 2**

   REFERENCE CODES:
   `UP_LEINI_1, UP_NCTLVRNFRR_1, UP_TORVISCOSA_1, UP_VADOTERM_5, UP_VOGHERA_1`

The second cluster groups a total of **5 thermoelectric power stations of CCGT technology**, belonging to various operators. As we can see from Table 1, these units are **extremely active with BID bids on MSD** and therefore compose a very interesting group for our final purposes. For this reason, I have kept them together in the cluster I presented in Section 6.4.5.

3. **CLUSTER 3**

   REFERENCE CODE:
   `UP_TAVAZZANO_5`

   The third cluster consists of a single thermoelectric unit belonging to the operator **EP Produzione**. Looking at Table 1, we can see how it is quite close to the units of cluster 2. For this reason, in the analyses of Chapter 6, I have united this unit with the thermoelectric units active with type BID bids on MSD.

4. **CLUSTER 4**

   REFERENCE CODES:
   `UP_NPWRFRRRRB_8, UP_SCTNPWPFRR_3`

   This cluster comprehends two thermoelectric units from the operators of *ENIPOWER* and *SEF* respectively. Those two units present **very similar behaviour to those of cluster 19**, for this reason I have grouped the two clusters together in the set of units from *ENIPOWER* and *SEF*, which I have presented in depth in Section 6.4.1.

5. **CLUSTER 5**

   REFERENCE CODES:
   `UP_CNTRLNTRNO_11, UP_NOCE_1, UP_S.FLORI.A_1, UP_S.PANCRAZ_1, UP_SLDGLRENZA_2, UP_TALAMONA_2, UP_VILLA_1`

   This cluster comprehends various units of Dispatchable Hydroelectric technology. Among these units, we find **two units from the operator Alperia and two units from ENEL**. The algorithm groups them together, due to the fact that they seem to be quite active with OFF bids on the Intraday Market. Other than that, they appear to have an average behaviour on the electricity markets and, for this reason, I did not include most of them in the clusters presented in Chapter 6.

6. **CLUSTER 6**

   REFERENCE CODES:
   `UP_LEVANTE_3, UP_MOLINE_1, UP_MONCALIERI_3, UP_MONCALRPW_2, UP_MONFALCO_1, UP_MONFALCO_2, UP_NOVEL_1, UP_TORINONORD_1,`

This is a cluster of mainly thermoelectric power plants from various operators. Judging from the position of the centroid of this cluster, it looks like that these units are more active than the average on MGP (like most of the units with similar technology), but do not operate much on the Ancillary Services Market (MSD). For this reason, after a visual comparison of their offer curves, I have decided not to analyse them further.

7. **CLUSTER 7**

   REFERENCE CODE:
   `UP_VAL_NOANA_1`

   This cluster comprehends a single hydroelectric power plant from the operator **Dolomiti Energia**. This unit has quite a peculiar behaviour, since it reports a value way above than average for the variable `MI_BID`, therefore the K-Means algorithm treats it as an outlier, placing this unit in a single separate cluster. Since MI is out of the scope of this work, I have decided not to consider the unit of *Val Noana* in further analyses.

8. **CLUSTER 8**

   REFERENCE CODES:
   `UP_BATTIGGIO_1, UP_BORGO_TRE_1, UP_CAORIA_1, UP_CHIVASSO_1, UP_CLHRCSLGNO_1, UP_CTE_DEL_M_2, UP_LASA_ME_1, UP_ROSONE_1`

   This group comprehends a set of PUs from various operators and with different production technologies, therefore is not trivial to interpret. Looking at Table 1, we can infer that the centroid of this cluster is slightly above average for the variables `TOT_PROD` and `TOT_INC`, while below average for the two variables related to the Intraday Market. Eventually, only the unit of *Borgo Trento* ended up being classifed in a cluster of major importance, that of hydroelectric power plants from ENEL.

9. **CLUSTER 9**

   REFERENCE CODES:
   `UP_ALTOADDA_1, UP_CARONA_1, UP_FADALTO_1, UP_FONTANA_B_1, UP_PNTVENTOUX_3, UP_S.VALBURG_1, UP_VALMALENCO_1`

   This cluster comprehends a set of 7 hydroelectric units, mainly from the operator ENEL, and all the coordinates of its centroid are close to the average. Most interestingly, **this cluster includes two of the units that switched management from ENEL to Alperia** in the reference period of this analysis. Indeed, it is due to the graphical visualization of the bids from units in this cluster that I was able to identify the major cluster of units that switched management from ENEL to Alperia, which I have presented in Section 6.4.3.

10. **CLUSTER 10**

    REFERENCE CODES:
    `UP_LA_CASELL_1, UP_LA_CASELL_2, UP_LA_CASELL_3, UP_OSTIGLIA_3, UP_PORTO_COR_3, UP_PORTO_COR_4, UP_TAVAZZANO_C_6`

    The tenth cluster groups a total of **7 thermoelectric power stations of CCGT technology**, belonging to two different operators: ENEL and EP Produzione. The algorithms suggests that these PUs are active way above than average for MI offers of BID type, and that they also are quite active on MSD OFF.
    Simply by looking at the reference codes, we can appreciate how this cluster comprehends the Production Units from the two power plants of *La Casella* and *Porto Corsini*, which I have included in the cluster of **thermoelectric plants from ENEL**, in Section 6.4.4.

11. **CLUSTER 11**

    REFERENCE CODES:
    `UP_MAEN_5, UP_PERRERES_1, UP_RIVADEL_3`

    This group comprehends two units from **CVA Trading** and one from **Dolomiti Energia** (the one of *Riva del Garda*). The algorithm seems to suggest that these units are below average for what concerns the Equivalent Operating Hours of production and the variable `TOT_INC`, while they are way above average for activity with MSD offers of OFF type. Eventually, I decided to **introduce more in depth the cluster of units from CVA Trading**, in Section 6.4.8, and I spent few words on the operator *Dolomiti Energia* in Section 6.4.9.

12. **CLUSTER 12**

    REFERENCE CODES:
    `UP_BARGI_CEN_1, UP_ETQ_ROVINA_1, UP_LEVANTE_4, UP_M._CIAPEL_1, UP_PRACOMUNE_1`

    This cluster comprehends a set of units with similar behaviour to those of cluster 1. Eventually, I have decided to add the first two units of the cluster, those of *Bargi Centrale Entracque Rovina*, to the cluster of **Pumped-Storage Hydroelectric power plants from ENEL Produzione**, presented in Section 6.4.7.

13. **CLUSTER 13**

    REFERENCE CODES:
    `UP_FUSINA_T_1, UP_FUSINA_T_2, UP_SPEZIA_CE_3, UP_VALPELLIN_1`

    This cluster comprehends the two Production Units that compose the power plant of *Fusina* (ENEL), another PU from ENEL and the unit from *Valpelline*, belonging to CVA Trading. As it is often the case for units belonging to the operator ENEL, these units are active more than the average in MI, presenting the second-highest value for the variable `MI_BID`.

14. **CLUSTER 14**

    REFERENCE CODES:
    `UP_CASSANO_2, UP_GRESSONEY_1, UP_LA_CASELL_4, UP_MASOCORON_1, UP_MORASCO_1, UP_PIACENZA_4, UP_SERMIDE_4, UP_SFRNGNRZNE_2, UP_SSTSNGVNN2_1`

    This cluster is one of the most difficult to interpret, yet it is of major importance, since the position of its centroid shows how it groups together units very active with MSD bids of OFF type, having, however, values below the average for the remaining 6 variables. Indeed, this group comprehends a set of PUs of various technology and operators, that seem to be characterized by **relatively low values of EOH, but end up being accepted in great quantities for upwards MSD offers**.

    Looking in more details at the units in this cluster, it looks like that we can identify three different sub-groups inside it: the first one is composed by **three thermoelectric units from the operator A2A** (those of *Cassano*, *Piacenza* and *Sermide*), the second one by four hydroelectric units (mainly from the operator ENEL) and eventually we have two more thermoelectric Production Units, one from ENEL, the other from EDISON.

    Eventually, I have decided to move the three units from A2A in the **cluster of thermoelectric power plants active with OFF bids**, which I have presented in Section 6.4.6. Furthermore, since the installed capacity of the PU of *Sesto San Giovanni* (`UP_SSTSNGVNN2_1`) is very limited, I have decided to discard it from further analyses.

15. **CLUSTER 15**

REFERENCE CODES:
`UP_BOAZZO_1, UP_CASTELDEL_1, UP_CHIEVOLIS_2,`
`UP_CIMEGO_1, UP_CURON_ME_1, UP_GEROLA_1,`
`UP_GOGLIO_2, UP_LIRO_1, UP_NPWRRVENNA_9,`
`UP_PANTANO_D_1, UP_PONTE_1, UP_PREM-GROSIO_1,`
`UP_ROVESCA_1, UP_S.MASS.CL_1, UP_SANGIACOMO_1,`
`UP_SND_ALBAN_1, UP_SND_CAMPO_1, UP_SOSPIROLO_1,`
`UP_SSTSNGVNNI_1, UP_TAGLIAMENTO_1,`
`UP_VALCAMONICA_1`

This is by far the biggest cluster of the twenty and comprehends a total of **21 PUs mainly of hydroelectric technology**, as well as some thermoelectric units with limited installed capacity, such as the one of *Eni Ravenna 9* (`UP_NPWRRVENNA_9`). Indeed, looking at the position of the centroid of this cluster, it looks like that it comprehends all the units that do not seem to have any peculiarities for the selected variables. For this reason, it is one of the most difficult cluster to interpret. Indeed, it is exactly due to this tendency to group together many units that I had to increase considerably the parameter K of the K-Means clustering algorithm.

16. **CLUSTER 16**

REFERENCE CODES:
`UP_CAVILLA_1, UP_CENCENIGH_1, UP_FUSINA_T_3,`
`UP_FUSINA_T_4, UP_GRAVEDONA_1`

This cluster comprehends 3 hydroelectric units of limited dimensions and two of the PUs from the power plant of *Fusina*, by ENEL. The algorithm highlights how these units are above the average for EOH and total revenues from the electricity markets, while they present lower activity on MSD. Eventually, the units of *Fusina* were added to the cluster of thermoelectric power plants of ENEL, while the units of *Cavilla* and *Cencenighe* to the cluster of hydroelectric units presented in Section 6.4.2.

17. **CLUSTER 17**

REFERENCE CODES:
`UP_ARSIE_1, UP_PELOS_1, UP_SOVERZENE_1`

This is a very small cluster composed by three hydroelectric PUs from the operator ENEL. The algorithm suggests that these units have a number of EOH that is above the average in the three considered years, and are characterized by a very high activity on the Intraday Market. Eventually, this cluster was taken as the baseline to analyse the behaviour of **hydroelectric units from ENEL**, which I have reported in Section 6.4.2.

18. **CLUSTER 18**

    REFERENCE CODES:
    `UP_BRUNICO_M_1, UP_CNTRLDTRNL_1, UP_LAPPAGO_1,`
    `UP_SFLORIANO_2, UP_SLDGLRENZA_1`

    This cluster comprehends four hydroelectric units from the operator Alperia, and one thermoelectric unit from A2A (that of Turano Lodigiano - `UP_CNTRLDTRNL_1`). This cluster is quite similar to cluster 14, therefore I have included the PU from A2A inside the cluster of thermoelectric units active with OFF bids and treated as a separate class all the units from Alperia.

19. **CLUSTER 19**

    REFERENCE CODES:
    `UP_NPWRFRRRRB_9, UP_NPWRMNTOVA_2,`
    `UP_NPWRMNTOVA_3, UP_NPWRRVENNA_10,`
    `UP_NPWRRVENNA_11, UP_SCTNPWPFRR_2`

    This is by far one of the most important cluster in the dataset, comprehending the second set of units from the **operators ENIPOWER and SEF**. For this reason, it is quite similar to cluster 4, and in the analyses conducted in Chapters 6 and 7 I have treated the two clusters as a single one.
    All the units in it have thermoelectric technology of CCGT type, and present a **very high value for the Equivalent Operating Hours of production**. Looking at the offers of these units on the market MSD, it is quite clear that the company ENI adopts **similar strategies** among all the power plants it controls, and in Section 6.4.1 I have discussed in depth this topic, reporting a graphical visualization of their offers curves.

20. **CLUSTER 20**

    REFERENCE CODES:
    `UP_AZOTATI_5, UP_TURBIGO_4, UP_VENAUS_1`

    This cluster comprehends the thermoelectric Production Units of *Azotati* and *Turbigo*, as well as the *dispatchable* hydroelectric unit of *Venaus*, by ENEL. The peculiarity of this cluster is definitely that its units are **the most active ones for MSD bids of OFF type** and this cluster was at the basis of the definition of the cluster of Thermoelectric units active with OFF offers, explained in Section 6.4.6.
    Regarding the unit of *Venaus*, it is a "basin" hydroelectric power

plants with a total installed capacity of 230 MW, which is a value way above the average hydroelectric power plant, which could be the reason why it is accepted many times on MSD.

CLUSTER NUMBER	TOT_PROD	TOT_INC	MGP_OFF	MI_BID	MI_OFF	MSD_BID	MSD_OFF
1	-1.43	-1.4	-1.37	-0.05	0.46	-0.28	-0.2
2	1.65	1.82	1.97	-0.39	-0.4	4.05	-0.34
3	0.73	0.86	0.81	0.83	0.34	2.28	0.7
4	1.68	1.7	1.34	1.56	2.66	0.58	0.14
5	-0.02	-0.04	-0.19	-0.14	1.59	-0.46	-0.7
6	1.56	1.53	1.58	-0.52	-0.8	0.24	-0.36
7	-0.92	-0.94	-0.45	4.91	-0.63	-0.46	0.31
8	0.36	0.35	0.42	-0.8	-0.76	-0.17	-0.32
9	-0.65	-0.71	-0.7	-0.08	0.47	-0.4	-0.39
10	-0.42	-0.38	-0.35	1.55	0.28	-0.14	0.44
11	-0.96	-0.71	-1.07	0.24	0	-0.35	2.37
12	-1.5	-1.6	-1.36	-0.72	-0.97	-0.36	-0.45
13	0.12	0.23	0.48	1.96	-0.08	-0.43	-0.58
14	-0.87	-0.79	-0.92	-0.67	-0.79	-0.31	1.31
15	-0.29	-0.38	-0.22	-0.46	-0.68	-0.42	-0.56
16	0.94	0.83	0.93	0.91	0.11	-0.37	-0.64
17	0.63	0.5	0.32	1.77	2.4	-0.46	-0.48
18	-0.03	0.07	-0.23	-0.26	0.94	-0.18	0.77
19	1.77	1.75	1.56	-0.08	1.14	1.01	-0.28
20	-0.81	-0.58	-1.14	-0.87	-0.98	-0.15	3.95

Table 1: K-Means clustering of the PUs in the `NORD` zone, position of the **centroids** with respect to the original (scaled) variables

Appendix B

Visualization of MSD bids by Clusters of Thermoelectric PUs

The following pages are dedicated to a **visualization of the offer curves on MSD** of four major clusters of thermoelectric units, among those that were identified in Chapter 6 and analysed in greater details in Chapter 7.

In particular, I will focus on the following clusters of Production Units:

1. Cluster of Thermoelectric Power Plants from **ENIPOWER and S.E.F.**
2. Cluster of Thermoelectric Power Plants from **ENEL Produzione**, focusing on the power plants of ***La Casella*** and ***Fusina***
3. Cluster of **Thermoelectric Power Plants active with bids of BID type**
4. Sub-Cluster of Thermoelectric Power Plants from **A2A**, extracted from the group of units active with MSD-OFF bids

In Section 6.4 I have already reported a visualization of the bids by these Production Units, relative to the whole time span of my analysis. In this Appendix chapter, however, I would like to report a **zoomed in visualization of the same curves**, depicting four separate weeks of bids[3], in order to appreciate in greater detail the peculiarities of each cluster.
This choice was made to better appreciate the **differences in behaviour** between the selected clusters, which are hard to perceive on the plot related to the complete time span.

All the **major comments related to the presented graphs are reported in Section 7.1**. In that occasion, I explained how the cluster from **ENIPOWER and S.E.F.** was the only one presenting very similar bids for all its units, in contrast to the cluster of PUs from **A2A**, which

---

[3] Ideally, I would have liked to zoom in every week of bids in our dataset, in order not to loose to much information from the plots reported in Chapter 6. This, though, was clearly unfeasible, and I had to restrict the analysis to only four separate weeks of bids. In any case, I have tried to choose weeks that come from different months and years, ranging over almost all the chosen time span.

presents major differences for OFF bids. I also pointed out how the operator **ENEL Produzione** differentiates quite a lot from its competitors, opting for a daily definition of the price for its offers, which remains constant from midnight to 11:00 p.m. of almost every day. Eventually, I identified two different behaviours inside the cluster of **Thermoelectric - MSD BID** PUs, with the unit of *Vado Terme* presenting offers at a higher price, and constant on a daily basis.

The four chosen weeks are the following (from top to bottom, for each reported figure):

1. First week of January 2017 (2 January 2017 to 8 January 2017)
2. First week of October 2017 (2 October 2017 to 8 October 2017)
3. First week of April 2018 (2 April 2018 to 8 April 2018)
4. First week of July 2019 (1 July 2019 to 7 July 2019)

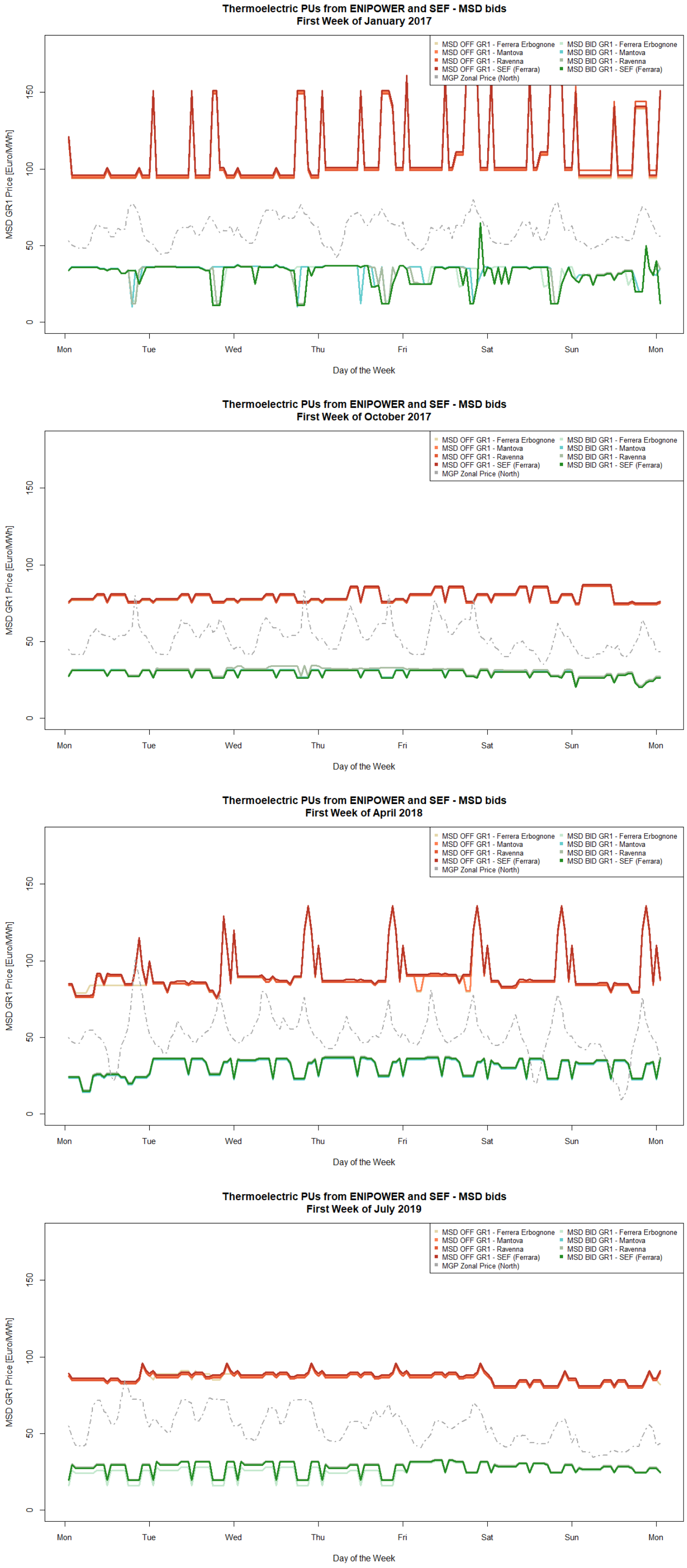


FIGURE 2: Cluster of **ENIPOWER and S.E.F.**: Visualization of MSD GR1 bids in four different weeks of our time span.

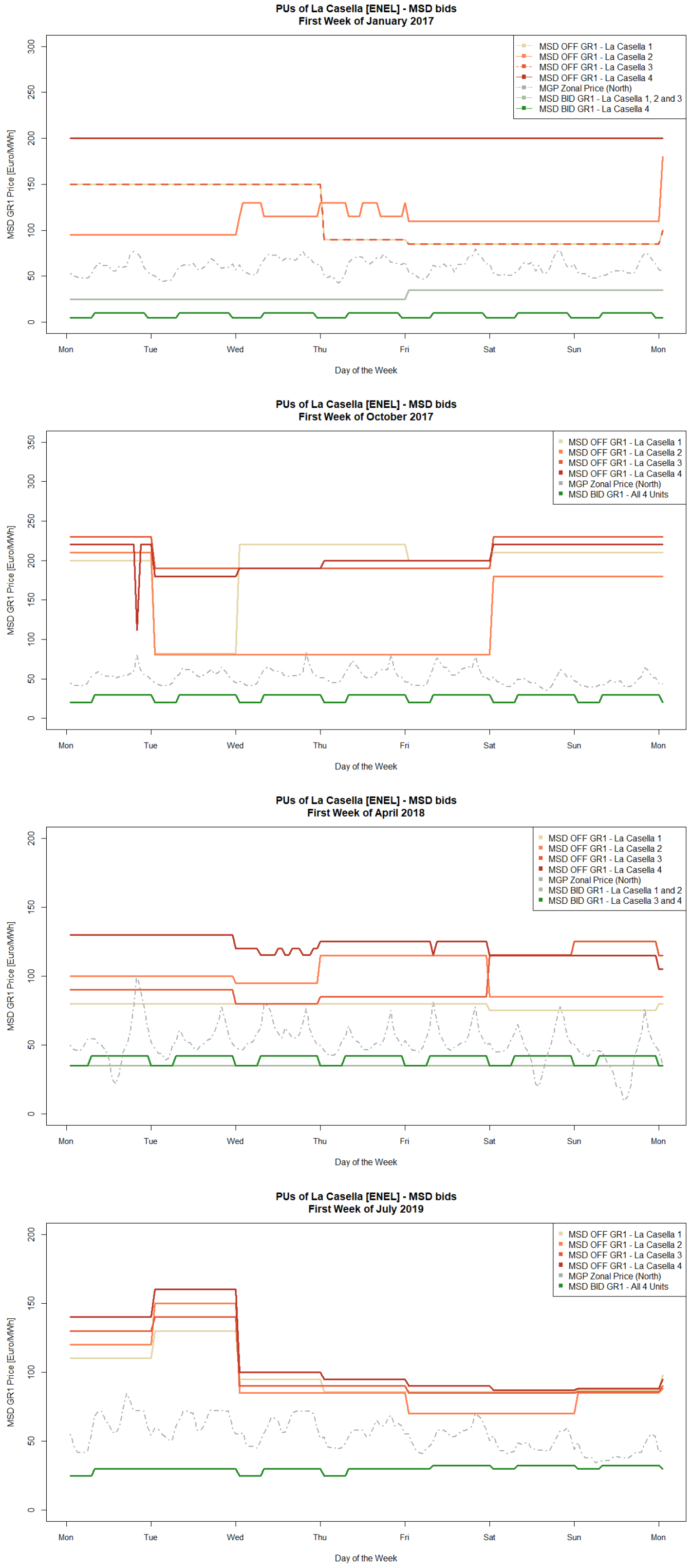


Figure 3: Cluster of **ENEL - *La Casella***: Visualization of MSD GR1 bids in four different weeks of our time span.

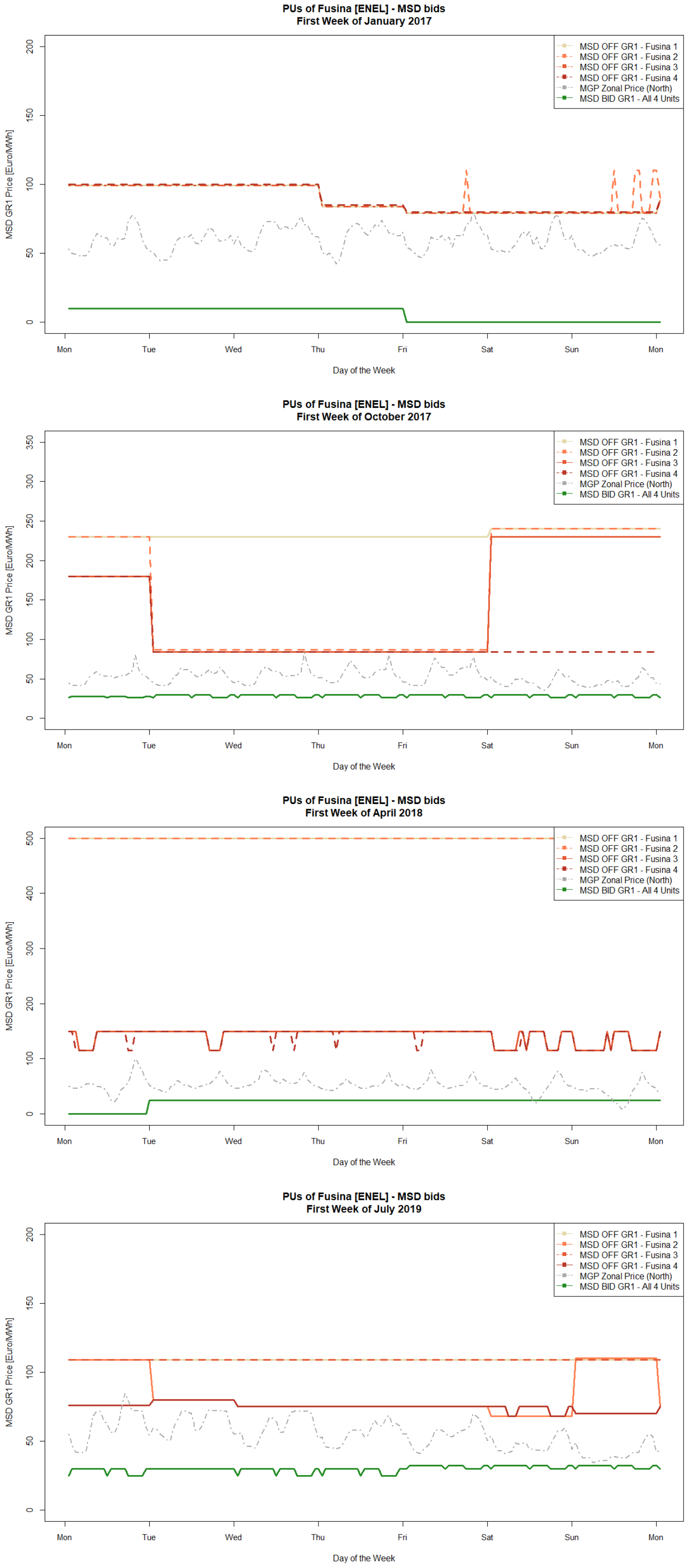


FIGURE 4: Cluster of **ENEL - *Fusina***: Visualization of MSD GR1 bids in four different weeks of our time span.

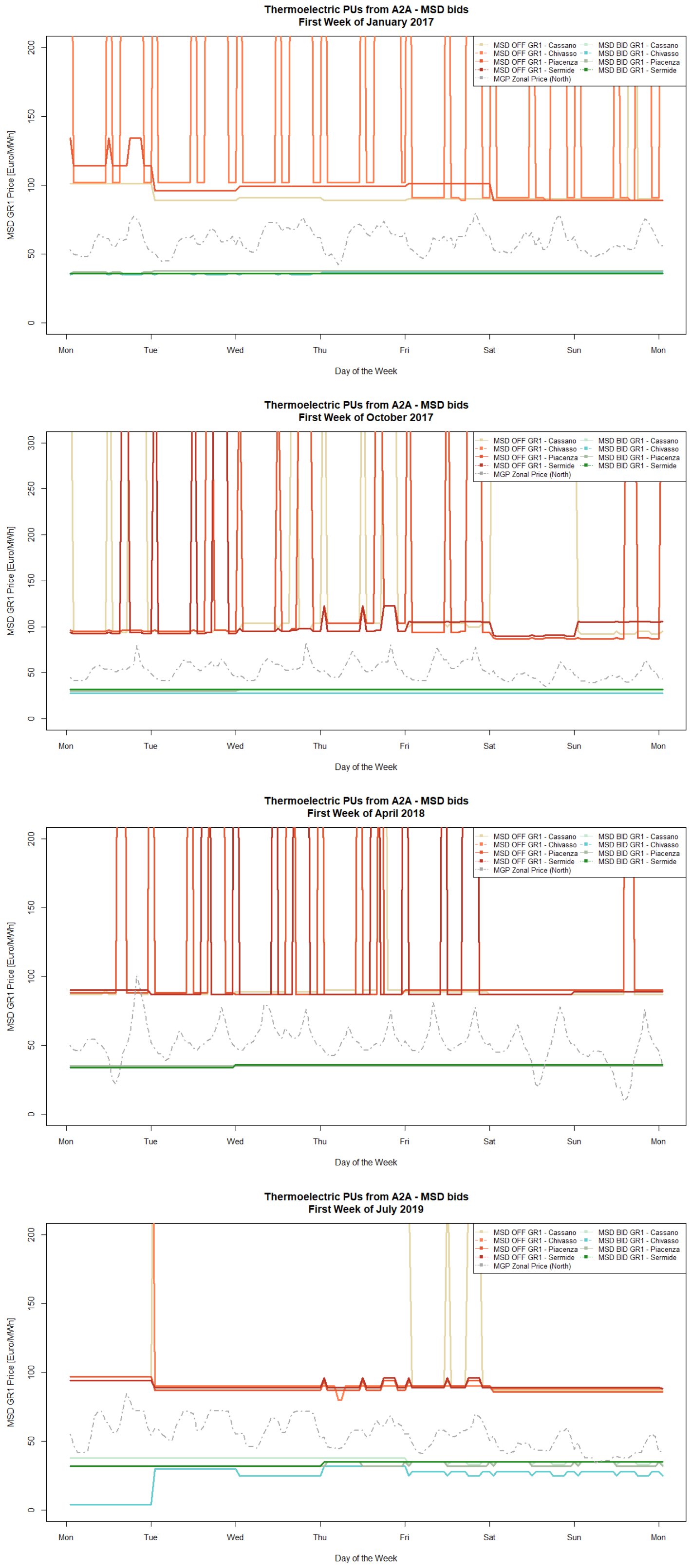


FIGURE 5: Cluster of **A2A**: Visualization of MSD GR1 bids in four different weeks of our time span.

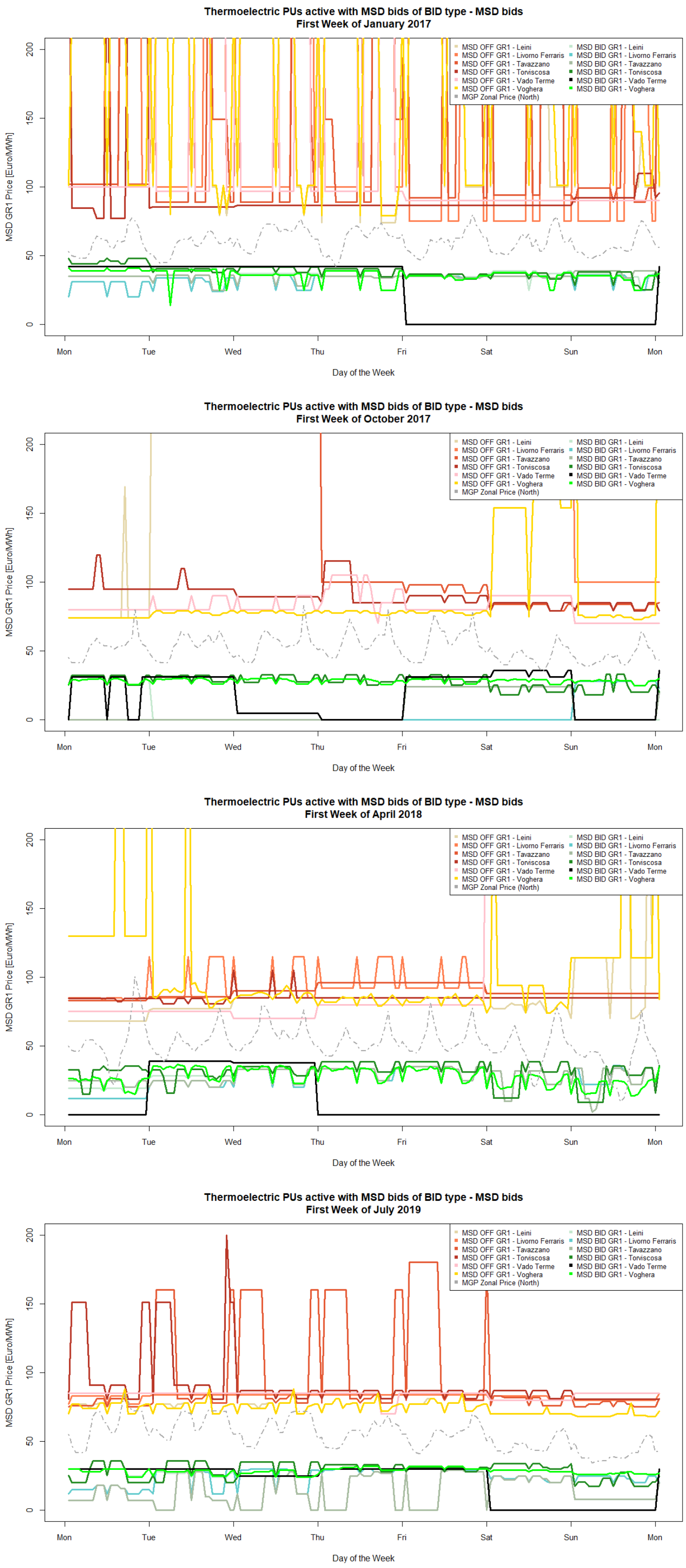


Figure 6: Cluster of **Thermoelectric, MSD - BID** units: Visualization of MSD GR1 bids in four different weeks of our time span.

References

[1] Sara Martucci - Politecnico di Milano [T.B.D.]
Master's thesis on the work conducted during the internship at R.S.E. - *Ricerca sul Sistema Energetico*

[2] Pietro Innocenti - Politecnico di Milano [AY 2019-2020]
Master's thesis: *Statistical Models and Algorithms for the analysis of the Electricity Market and the Economic Dispatch Service*

[3] Virginie Marchionni - Politecnico di Milano [AY 2019-2020]
Master's thesis: *TAJIVE: a new tool for integrative analysis of multidimensional data. Application to the Italian Energy market.*

[4] F. Bovera, A. Blaco, G. Rancilio, M. Delfanti [2019]
Assessing the Accuracy of Different Machine Learning Classification Algorithms in Forecasting Results of Italian Ancillary Services Market. - In: 2019 16th International Conference on the European Energy Market (EEM), pp. 1–5.

[5] Robin Genuer, Jean-Michel Poggi
Random Forest with R - Springer, "Use R!" series

[6] Hossein Haghighat, Hossein Seifi and Ashkan Rahimi Kian [2012]
International Journal of Electrical Power & Energy Systems: Volume 42 n. 1, pages 350-358. *Pay-as-bid versus marginal pricing: The role of suppliers strategic behavior*

[7] RSE - Seminario sui Mercati Elettrici [10^{th} November 2020]
La Borsa Elettrica Italiana - Il funzionamento del mercato elettrico e la relativa evoluzione

[8] ENTSOe - European Network of Transmission System Operators
Electricity Market Transparency Platform

[9] Terna S.p.A. - Italian Transmission System Operator
Electric System - Terna Website

[10] GME - Gestore dei Mercati Energetici SpA
Mercatoelettrico.org - Download Section

[11] GME - Gestore dei Mercati Energetici SpA [2009]
Vademecum della Borsa Elettrica Italiana

[12] GME - Gestore dei Mercati Energetici SpA [2018]
Annual Report 2018 - Relazione annuale Anno 2018
Available from GME's Website - *Rapporti e Relazioni Annuali 2018*

[13] GME - Gestore dei Mercati Energetici SpA
Testo Integrato della Disciplina del Mercato Elettrico [Approvato con D.M. del 19 dicembre 2003 come successivamente modificato e integrato]
Available from GME's Website - *TIDME*

[14] GSE - Gestore Servizi Energetici [2019]
Diffusion and Development of Renewable Resources in Italy and Europe
Available from GSE Website - *Scenari e Statistiche*

[15] Terna S.p.A. - Codice di Rete
Capitolo 4, Regole per il dispacciamento

[16] ISTAT - Istituto Nazionale di statistica
Archivio elenco comuni e denominazioni - Anni 2017-2021

[17] Meteored Italia - Archivio Storico Meteo Milano
www.ilmeteo.net

[18] RDocumentation - The `stats` package in R
K-Means Clustering: function kmeans()

[19] Leo Breiman - University of California [2001]
Machine Learning: Volume 45, pages 5–32. *Random Forests*

[20] Andy Liaw & Matthew Wiener (2018) - The `randomForest` package in R, version 4.6–14
Breiman and Cutler's Random Forests for Classification and Regression.