UNIVERSIDAD PONTIFICIA COMILLAS

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

OFFICIAL MASTER'S DEGREE IN THE ELECTRIC POWER INDUSTRY

Master’s Thesis

SYSTEM IMBALANCE FORECASTING AND SHORT-TERM BIDDING STRATEGY TO MINIMIZE IMBALANCE COSTS OF TRANSACTING IN THE SPANISH ELECTRICITY MARKET

Author: /ł·ŀl2ťyl· /2ÿúŀŀŜŀŀ·ă

Supervisor: tł·ōť2 ŘŜť {ł·Îπhŀŀ2Îŏ2

Madrid, ŀŀdzťè нлмс

Master's Thesis Presentation Authorization

THE STUDENT:

NAME OF THE STUDENT

Carolina Contreras

THE SUPERVISOR

Pablo Del Saz-Orozco

Signed: .... Date: 28/ 06/ 2016

Authorization of the Master's Thesis Coordinator

Dr. Luis Olmos Camacho

Signed.: .... Date: ....../ ....../ ......

UNIVERSIDAD PONTIFICIA COMILLAS

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

OFFICIAL MASTER'S DEGREE IN THE ELECTRIC POWER INDUSTRY

Master’s Thesis

SYSTEM IMBALANCE FORECASTING AND SHORT-TERM BIDDING STRATEGY TO MINIMIZE IMBALANCE COSTS OF TRANSACTING IN THE SPANISH ELECTRICITY MARKET

Author: /l·ǌŽƚyl· /ŽyůǌŜǌl·ă

Supervisor: tl·ōƚŽ ŘŜƚ {l·ĵƚhǌŽĵŏŽ

Madrid, ǁdzƚè нлмс

Abstract

Energy imbalances can represent a significant cost for agents transacting in markets that penalize participants' imbalances. In markets with increasing penetration of intermittent renewable sources of energy (RES-E), system imbalances can not only be costly, but also increase, as is the case for the Spanish power market. Market participants, especially those trading non-dispatchable energy, are therefore interested in minimizing this cost while simultaneously maximizing their profits.

A lot of work has been developed around the forecast accuracy and uncertainty of RES-E production to determine bidding strategies that minimize imbalance costs, especially for wind power trading. Challenges inherent to agents specialized in power trading and/or retailing activities, especially wind power trading of energy produced by third parties or retailing to small consumers means that applying strategies that rely on production forecasts may not be sufficient.

In this master thesis we considers those challenges by developing an optimized bidding strategy that reduces the expected imbalance cost for a real case-study of a Spanish energy trader/retailer based on a forecast of the system´s imbalance volume and past imbalance costs, while using new information available after the day-ahead market gate closure for participation in the intra-day market to influence the imbalance volume of the agent's portfolio towards the direction that reduces their potential imbalance cost. This strategy does not replace accurate forecasting but considers the practical aspects of energy traders/retailers with numerous small clients who cannot operate production units. The strategy can be applied from the perspective of both a trader and retailer.

We have developed an advanced model based on random forest technique to forecast the system imbalance and used a genetic algorithm to apply the bidding strategy that minimizes the imbalance costs based on system imbalance forecasts and past imbalance costs. The proposed strategy application using new information available after the day-ahead gate closure outperforms its application in the pre-day-ahead market.

Acknowledgements

I would like to express my gratitude first and foremost to Pablo Del Saz-Orozco for supervising this thesis project and whose expertise, understanding, and patience are an integral part of this work; to Ignacio Ortiz Lopez for the warm and supportive welcome into the Energy Management group at AXPO Iberia, and to the rest of the team whose experience greatly enhanced the learning process while providing good laughs.

A special thanks to the EMIN coordinators for their support and guidance as I navigated the challenges of undertaking this graduate program in two different countries, and to my fellow EMIN students whom experienced this process from Paris to Madrid together with me, sharing in both the struggles and joys it has showered us with.

Finally and most importantly my deepest gratitude to my husband and best friend for his unrentlentless support, love, and confidence in me, and to my family for encouraging this journey from beginning to end, not just now but always.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION

The inability to efficiently store electricity for later consumption requires continuous balancing of power supply and demand. With liberalization of the electricity sector came the decentralization of activities, including the responsibility to balance the power system. Within this decentralized context, electricity market designs have incorporated mechanisms to ensure that the balance of the system is maintained at all times. The so-called "Balancing Responsible Parties" (BRPs) are market agents responsible for ensuring that the generation and/or consumption within their portfolio is balanced. The physical balancing, however, is done by the Transmission System Operator (TSO). In European-type markets, power balance management is achieved through market-based mechanisms also referred to as "balancing markets" (van der Veen, 2012). The TSO can procure balancing services and a range of other functions intended to guarantee system security at the least cost (ENTSO-E, 2016). TSOs penalize BRPs for causing imbalance.

The high penetration of RES-E (renewable energy sources for electricity) has exacerbated the challenge of balancing the electricity system. This is due to their intermittent nature and uncertainty of production, in particular for wind energy production. As a result, the participation of wind energy sources in the electricity market may imply large deviations from the initial schedule. In the Spanish electricity system, as in certain other European markets, these deviations are economically penalized, leading to a cost that has to be borne by the market agent causing the imbalance. For wind energy producers and traders[1] this is particularly challenging as their energy source is uncertain and their production non-dispatchable, and although forecasting techniques have greatly improved, these are still not perfectly accurate.

[1] Energy traders may engage in both procurement and selling of energy.

Considering that Spain has the second largest wind energy market in Europe, and the fourth world wide[2], and that these costs have also increased in the Spanish electricity system since 2009 (Batalla-Bejarano, et al., 2015), managing imbalance costs is critical for agents transacting in the Spanish electricity market.

Minimizing imbalance costs, and consequently maximizing their profits, is an objective that all market agents strive to achieve.

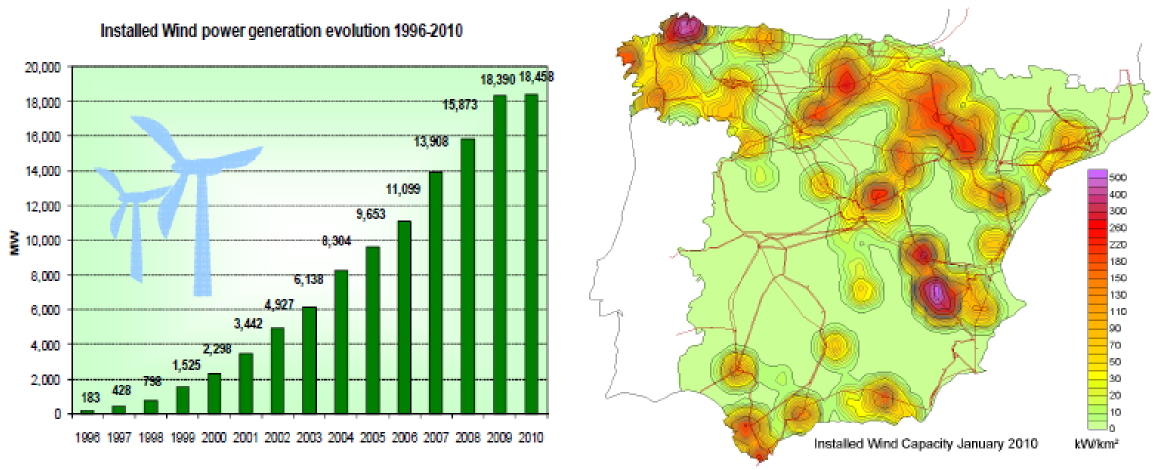


Figure 1.1: Evolution of installed wind power generation evolution 1996-2010. ***Source:*** **(de la Fuente, 2009)**

Figure 1.2: Evolution of net production from Renewable Energy in the Spanish Peninsula. ***Source: REE***

---

[2] REN21, Renewables 2015 Global status report. Available at http://www.ren21.net.

1.1.1 Bidding Strategies to Minimize Cost

Agents trading wind power in the market seek to maximize profit and minimize the imbalance costs, but uncertainties in the hourly available wind and forecasting errors make the bidding risky. The BRP may choose from two different options to trade in the market (Matevosyan & Söder, 2006):

1- assume wind power forecasts are certain and bid that amount in the market; or
2- bid the amount that minimizes expected costs for imbalances (considering uncertainty, and possibly imbalance cost) .

Option 1: Wind Power Forecasting

Generally, the first option relies on estimating production using short-term term wind power production tools, which usually provide the forecasted power level and the associated uncertainty (Bueno-Lorenzo, et al., 2013). Commonly such tools provide the future production of a wind farm for a period ranging from the next hours to the next days, and are based on meteorological predictions, on onsite measurements and on wind farm characteristics. The increase in accuracy of these predictions has been widely documented to increase the closer the prediction horizon is to delivery time.

Despite significant progress in the accuracy of wind power forecasting with these tools in recent years, they are still not perfect; deviations from forecasted and committed power produce imbalances which have an economic impact on the traders/producers of such RES-E.

Option 2: Develop an Optimized bidding strategy

In Pinson, et al., the authors show that option 2, when accompanied with *optimal* bidding strategies in conjunction with accurate forecasts developed in option 1, outperform the application of option 1 alone in reducing imbalance costs (Pinson, et al., 2007). Several studies have been published proposing such optimal bidding strategies when trading wind power to minimize the imbalance cost, as option 2 above suggests. As indicated in Bueno-Lorenzo, et al., (Bueno-Lorenzo, et al., 2013), most of those optimal bidding strategies have focused on considering the uncertainty of the forecast, and in some instances, also the imbalance energy cost/price to minimize the imbalance cost as in (Matevosyan & Söder, 2006). See Bueno-Lorenzo, et al (Bueno-Lorenzo, et al., 2013) for an over review of certain approaches.

In addition to the optimization strategy, some previous studies considered risk management techniques to reduce the threat of high imbalance economic losses (e.g., applying VaR or CVaR risk measures) or to reduce the imbalances volumes. These methods consider the variability of imbalance prices and/or production and address them to reduce the risk of incurring excessive costs.

As reviewed in (Cháves-Ávila, 2012), the strategies discussed in the literature differ in the markets they considered (i.e. Dutch or Spanish market, day-ahead or intra-day Market), the methodology (stochastic linear programming, mixed integer formulation, time series, among others), and the assumptions on market behavior. Sse (Cháves-Ávila, 2012) for a brief overview

Agents specialized in trading and/or retailing activities.

The strategies discussed above, however, don't consider certain logistical challenges specific to energy traders and retailers who 1- do not own or operate the generation units in their portfolio (especially RES-E) in the case of the former, or 2- who retail to numerous small consumption points (small customers) in the case of the latter. These cases present unique challenges because their accessibility to real-time data and knowledge of specific operational details vital to forecasting production may be limited, meaning that in reality their portfolios will always have an imbalance. The literature review did not yield a strategy based on considering a forecast of the system imbalance itself. In this project we aim to develop a strategy based on forecasting the system´s imbalance volume which is useful for any market agent who may not have reliable or can readily access site specific wind production data. Energy traders or retailers who do not own and operate the units whose energy they are trading may be faced with this situation. By forecasting the direction and level of the system imbalance, the agent can inform their bid adjustment decisions to influence their own imbalance volume towards the direction that reduces their potential imbalance cost.

BRPs are able to influence their imbalance volume by means of over and under contracting of energy before final gate closure of the last adjustment market (in Spain, this ranges between 3.25 and 6.25 hours before real-time, through the intra-day market), and by means of internal balancing in real-time. With both activities, a BRP can create an 'intentional imbalance', in order to hedge against the financial risks of imbalance settlement, i.e.to limit imbalance costs (van der Veen, 2012).

To determine this volume, an optimization tool that considers the historical balance costs will be applied.

1.1.2 Advanced tools

With increased complexity of competitive markets, traditional forecasting tools may no longer be sufficient. Advanced modeling techniques, such as artificial–intelligence based tools, have evolved enormously in recent years finding widespread application in many industries, including the electricity sector.

1.1.3 Prediction and Strategy Horizons

The day-ahead market is usually liquid and provides many trading possibilities for market agents (Chaves-Avila, et al., 2013). However, the increase in accuracy of wind production forecasts has been widely documented to increase the closer the prediction horizon is to delivery time. Additionally, from the day-ahead market gate-closure until real-time delivery, there is new information available, such as previous market results and meteorological forecasts that can be used to increase the accuracy of forecasts to further optimize an agent´s bidding strategy. Consequently intraday markets provide agents an opportunity to incorporate that information.

The forecast horizon, the timing of intraday markets, and availability of market variables are critical to developing any strategy.

1.2 Project Objectives

1.2.1 General Intention

In this project we aim to develop a strategy that considers the challenges faced by agents specializing in either power trading and/or retailing activities. The strategy is especially applicable to RES-E trading and retailing to small customers as their portfolio imbalances are unavoidable and will always exist. Yet their costs can be reduced by influencing the direction of the imbalance. This strategy is not intended to replace those based on production/load forecast accuracy, but instead intended to be a complementary tool for small traders and/or retailers.

We seek develop a strategy based on forecasting the hourly system imbalance volume: a variable that is is useful to all market agents who may or may not have reliable or easy access to real-tieme data such as site specific wind production data. Energy traders or retailers who do not own and operate the units whose energy they are trading may be faced with this situation. By forecasting the direction and level of the system imbalance, the agent can inform their bid adjustment decisions to influence their own imbalance volume towards the direction that reduces their potential imbalance cost.

1.2.2 SPECIFIC OBJECTIVE

The specific objective of this project is to:

> ***Develop an optimized bidding strategy that reduces the expected imbalance cost for a real case-study of a Spanish energy trader/retailer[3], considering a forecast of the system´s actual imbalance volume and past imbalance costs, while using new information available after the day-ahead market gate closure for participation in the intra-day market.***

The strategy to reduce imbalance costs is developed from the point of view of an RES-E trader transacting in the short-term Spanish electricity market, while assuming the agent will bear all imbalance costs of its portfolio. However, the strategy can also be applied to a retailer. The trader´s objective is to reduce its imbalance costs through possible participation in the intra-day market in order to use new and updated information that may increase the accuracy of the forecast and effectiveness of strategy. The strategy will be based on influencing its imbalance volume by over and under contracting of energy before final gate closure.

A model to optimize the bidding strategy will be developed, and it is based on developing two main components using AI-based tools:

A. **Forecasting model** to predict the hourly net system imbalance volume considering information available post-gate closure of day-ahead market;

B. **Optimized strategy application to determine the hourly energy level to bid in the intra-day market** that will minimize the imbalance costs considering the forecasted imbalance volume and past system imbalance costs.

[3] In the Spanish electricity market, RES-E producers designated as "special-regime" have the option of selling their energy (portfolio or as a unit) in the wholesale market directly or to an energy retailer who will submit sale bids to the day-ahead and intra-day markets, and execute bilateral contracts. They may also do so through an energy trader who, by means of a contract, acts as representative and is responsible for submitting sale bids to the day-ahead and intra-day markets, and executing bilateral contracts.

Other elements that are considered in this include:

- Optimization of forecasting model parameters.
- Comparing of the strategy´s performance based on its implementation pre and post day-ahead market gate closure.
- Evaluation of the effect of day-ahead and intra-day market price spread on strategy.

1.3 Document Structure

The first chapter serves as an introduction to the issues at hand motivating the project, and lays out the objectives and scope of the project.

Chapter 2 describes the Spanish electricity market, concentrating on the short-term power market and the mechanisms applied for balancing of the system. A brief overview of the regulatory framework under which the market operates is included to provide further context on the market's evolution and environment.

Chapter 3 discusses the system imbalance and source of the imbalance, including the role played by intermittent RES-Es. Previous work on forecasting the system imbalance is discussed, along with a description of the mathematical tools applied to develop the bidding strategy, and the data and programming tools used to support the strategy.

An analysis of potential predictor variables followed by description of the development of the model are contained in Chapters 4 and 5, respectively. Chapter 5 includes a detailed description of the two main components of the strategy: 1- the forecasting model and 2 – the bidding strategy application. Details on the methods used to validate the model are discussed at the end of that Chapter.

In Chapter 6 we present and review the results of the model's performance. The final chapter (Chapter 7) brings together the general conclusions derived from the different analyses and suggests new directions.

CHAPTER 2:

THE IBERIAN ELECTRICITY MARKET

This chapter provides introduction to the Iberian Electricity Market (MIBEL)[4] with a brief overview of the regulatory framework governing its operation, followed by descriptions of the Spanish wholesale market focused on short-term energy procurement market and adjustment services.

2.1 REGULATORY FRAMEWORK OVERVIEW

The electricity sector in Spain has been governed by its Electricity Sector Act 54/1997 (1997 Directive), which liberalized the electricity market and incorporated the European Commission´s 96/92/EC Directive. Later, Directives 2003/54/EC and 2009/72/EC (2003 and 2009 Directives, respectively) concerning common rules for the internal electricity market, were also incorporated into Spanish law as amendments to the original 1997 Directive. In 2013, a new Electricity Sector Act, 24/2013 was approved. It contains the main electricity sector regulation in Spain aimed at providing regulatory certainty, ensuring effective competition in the electricity sector and the economic and financial sustainability of the electricity system (International Energy Agency, 2015).

As a result, today:

- A new management and regulatory system has been established.
- All consumers have free choice of electricity supplier.
- Development decisions for new generation plants are decentralized under the context of a competitive market model.
- International electricity trade has been liberalized.
- Transmission and distribution are regulated activities and generation and retailing are fully liberalized activities.

[4] Mercado Ibérico de la Electricidad.

2.1.1 HISTORICAL OVERVIEW

Following the worldwide liberalization trend of the 1990s, Spain began its own reform process aimed at increasing competition and competitiveness. In terms of the electricity sector, this process began in 1998[5] after adoption of the 1997 Directive, and 2007 integrated the Portuguese electricity market to makeup what is MIBEL today.

Although progress is still under way, following are some of the key aspects of this process:

Privatization efforts of the previous state-owned electricity entities which dominated the electricity sector. System assets were placed under private ownership (generation and distribution) or under entities with minority state-ownership (i.e. transmission).

Spain decided to privatize the publicly owned generator Endesa – the system´s main generator - but only after allowing it to absorb two other companies. The transmission grid has been controlled by a separate partly state-owned entity, Red Eléctrica de España (REE).

Unbundling of activities[6] based on the principle of separating regulated activities (distribution and transmission) from other segments of the electricity value chain (i.e., generation, and retailing activities). Prior to reform most Spanish electricity companies were vertically integrated[7], so a legally independent[8] transmission company – REE - was created to separate generation from transmission activities. REE was created in 1985 (prior to the 1997 Directive) as the first company in the world exclusively involved in electricity system operation and transmission (Red Electrica de España, 2016).

Unbundling of distribution activities was more lenient. Distribution activities required legal and accounting separation (1997 Directive) followed by functional separation in 2007 (based on 2003 Directive).

---

[5] Following approval of the Electricity Sector Act 54/1997.

[6] There are essentially four separate economic activities in electricity markets: generation, transmission, distribution and retailing, and varying degrees of unbundling: legal, functional, accounting, and ownership separation.

[7] In the mid-1980´s the Spanish electricity industry was made up of eleven vertically integrated companies operating in generation, transmission, and distribution, in addition to state-owned Endesa in generation.

[8] The Spanish government continues to hold a minority ownership-stake in REE (approximately 20%).

Companies performing regulated activities (transmission, system operation, or distribution) cannot develop production, trading or energy recharge activities (as of 2010) or make investments in companies engaged in these activities. However, and subject to certain requirements, different companies engaged in any of aforementioned regulated or unregulated activities can operate under the umbrella of the same holding company or corporate group.

For retailing activities, the most significant unbundling occurred in 2010 when legal unbundling of distribution system operators (DSOs)[9] from retailing activities was required[10]. DSOs are no longer able to supply electricity to their customers. Prior to that date, consumers were able to choose between supply from distribution companies – through end-user regulated prices – or from retailers under free market conditions. End-user regulated electricity prices disappeared along with the DSO's role as suppliers.

The unbundling process in Spain is an on-going effort.

Deregulation of generation and retailing[11] activities intended to achieve economic efficiency through free market competition. To foster competition at the generation level, a wholesale electricity market was established for sellers (generators/traders) and buyers (retailers/traders) to transact according to the principles of objectiveness, transparency, and free competition. An integrated wholesale market for Spain and Portugal was proposed, resulting with the MIBEL which began operating on July 2007.

A market operator – Operador del Mercado Eléctrico (OMEL) – was created to manage economic transactions in the wholesale market. In an effort to promote and ensure fair competition and transparent markets, an independent regulatory authority was to be established. As of 2013, the National Markets and Competition Commission (CNMC) fulfills that regulatory role after absorbing the duties of the National Energy Commission. The CNMC maintains regulatory oversight of energy and other industry markets to prevent and sanction abusive practices, among other duties. The MIBEL will be discussed at greater length in the following sections of this Chapter.

[9] These DSOs are the owners of the networks they operate.
[10] DSOs are permitted to belong to a group that undertakes other activities including: power generation, electricity recharging services (for electric vehicles) and selling electricity provided that a separate company performs the regulated activities

[11] Electricity retailing refers to the act of purchasing energy for sale to end users.

Deregulation of retailing activities has been more gradual. Following the 1997 Directive only very large consumers were allowed to choose a supplier, which changed in 2003 after which all customers are now eligible and can freely choose their supplier.

Transmission and distribution activities, characterized as natural monopolies, remained under regulation, but remuneration based on cost-of-service regulation was replaced by incentive-based remuneration. The Ministry of Industry, Energy and Tourism, who has the lead responsibility for formulating and implementing energy policy, approves the electricity network access tariffs, regulated components of electricity prices and level of access tariffs.

Regulated third party access (TPA) to ensure non-discriminatory access to the transmission and distribution network. To this end, the roles of market operator and system operator were created to manage the economic and technical activities at the interface of competitive (generation) and regulated (transmission) businesses. As mentioned earlier, OMEL was established as the market operator responsible for economic operation of the system. The role of Transmission System Operator (TSO), who is responsible for technical operation of the grid was absorbed by the owner of the transmission grid (REE) under a TSO ownership unbundling model[12].

The CNMC sets out the methodology for calculating the network access tariffs, and the Ministry of Industry, Energy and Tourism, approves the electricity network access tariffs and the level of access tariffs.

2.2 THE SPANISH WHOLESALE MARKET

The Spanish wholesale market is part of the MIBEL and made up of 1- an organized market (with day-ahead and intra-day activities managed by OMIE, the electricity market operator) and 2- a non-organized market for bilateral trade (OMIP in Portugal manages the futures market). The non-organized part consists of physical bilateral contracts, whose economic terms and conditions are agreed between the signing parties. According to the CNMC bilateral contracts represented 29% of the energy sold in the daily program in 2014. This section will focus on the short-term electricity market.

[12] Functional and accounting unbundling also separate transmission from system operation activities.

Both OMIP and OMIE belong to the Iberian Market Operator´s [13] (OMI´s) corporate group, where the Spanish and Portuguese wholesale market operators (OMEL and OMIP, respectively) share equal ownership. Figure 2.1 below depicts the Spanish electricity markets.

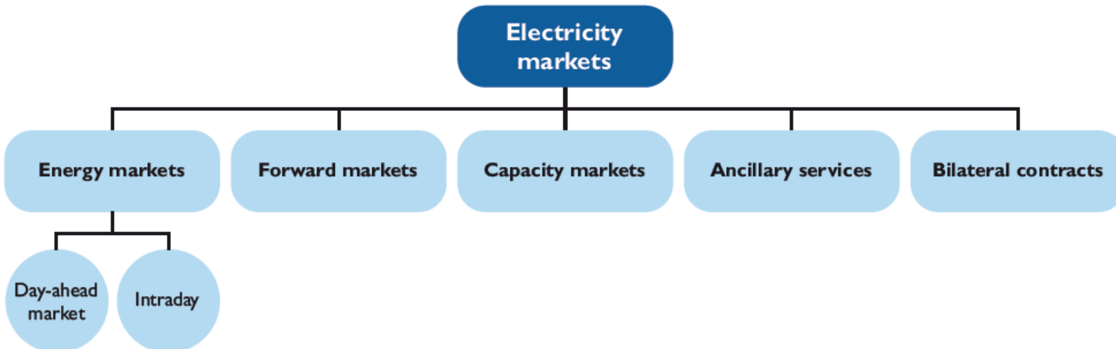


Figure 2.1 Overview of electricity markets in Spain. ***Source: IEA, 2015***

REE, the Spanish system operator, is responsible for the technical management, including system security and balancing of the Spanish transmission grid. Balancing is a market-based activity through a market of ancillary services which are discussed in greater detail under Section **Error! Reference source not found.** of this Chapter .

In 2014, OMIE managed transactions amounting to almost 11 billion euros, accounting for more than 80% of the electricity supplied in Spain and Portugal (OMIE, s.f.).

For 90 to 95% of the time MIBEL observes a single price for 90% to 95%, with market splitting (into a Spanish and Portuguese price) the rest of the time when interconnections are congested. According to the Spanish regulator, the spot market is very liquid; it gathers 214 buyers and 110 sellers (CNMC, 2015). In its most recent country report the International Energy Agency (IEA) assessed the Spanish wholesale market as "fairly competitive", with five main players each having a 15% to 24% market share (International Energy Agency, 2015).

See Figure 2.2: Function separation between the market operator and the transmission system operator.

[13] Operador del Mercado Ibérico

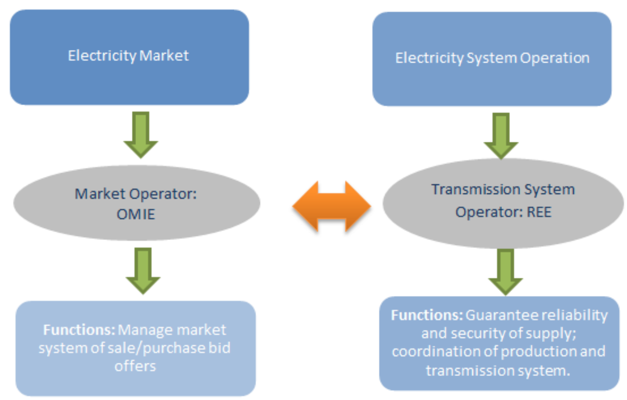


Figure 2.2: Function separation between the market operator and the transmission system operator.

Spain´s short-term electricity market is further described below.

2.2.1 Spanish Short-Term Electricity Market

Short-term procurement of electricity in Spain is done through a spot market organized as a sequence of markets:

- A *day-ahead market*, where most of the physical production is traded, and *intra-day market* consisting of six discrete auctions which operate closer to real-time for agents to adjust their portfolio (the balancing mechanism). These energy markets are managed and run by the market operator, OMIE, and described in the following sections. These markets are fully integrated between Portugal and Spain, thus implicitly allocate cross-border capacity.
- The so-called *adjustment services* to resolve *technical constraints* (i.e. network constraints and/or reserve requirements) and the market-based procurement of ancillary services for balancing active power, namely *secondary reserves*, *tertiary reserves*, and *deviation management*. Balancing of reactive power is another ancillary service, but is not the focus of this paper. These market-based services are managed by the system operator, REE, and discussed in Section 2.3 .

As agents can also trade electricity through bilateral contracts with physical delivery, those parties holding bilateral contracts have to inform the system operator of the electricity contracted before the day-ahead market is held. These contracts will not the focus of this paper. A high-level sequencing of events for the short-term market in Spain is further detailed in Figure 2.3.

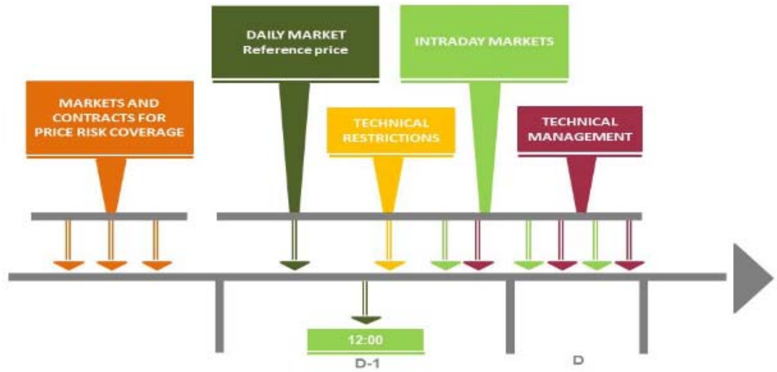


Figure 2.3: Sequence of markets and processes in MIBEL. ***Source: OMIE***

2.2.2 DAY-AHEAD MARKET

The day-ahead market is the main electricity trading market in Spain. It applies the marginal pricing principle in which the price and trading volume in each hour are set according to the point of equilibrium between supply and demand (Figure 2.4).

Every day, up until gate-closure at 12:00 pm, agents submit bids to purchase or sell electricity for delivery every hour of the next day. These bids are aggregated and the market cleared using a European algorithm called EUPHEMIA, based on economic merit order. OMIE then publishes the market results of energy to be delivered each hour the following day and the price clearing the market. This day-ahead price becomes a reference price in other subsequent markets. In 2013, the daily market in 2013 traded an average of approximately 71% of the energy consumed in the Iberian market (OMIE, s.f.). As most energy is traded in this market, it considered very liquid.

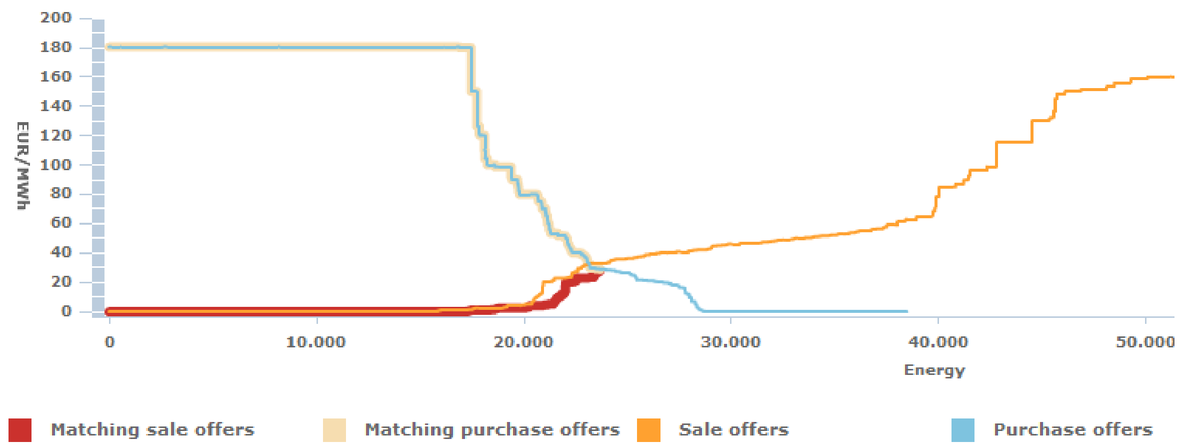


Figure 2.4: MIBEL aggregate demand and supply curves, 15/06/2016. ***Source: OMIE***

Participating Agents

All available generation units, if not bound by physical bilateral contracts, are required to submit bids to the day-ahead. Thermal generators must submit offer on a unit per unit basis. From the day-ahead market on, all agents have to submit schedules on a per unit basis, and not based on a portfolio aggregation.

Purchasing units, made up of retailers, re-sellers and final consumers of a certain size, may procure electricity in the day-ahead market. According to OMIE:

- Reference retailers participate in the market to procure electricity for their portfolio of consumers.
- Resellers participate in the market to purchase electricity to sell to direct consumers.
- Final consumers may purchase electricity directly on the organized market, through a reseller by signing a physical bilateral agreement with a producer.

Bid Formats

Multipart bidding is accepted in the day-ahead market, were participants can submit bids containing so-called semi-complex conditions. These conditions can be technical or economical to account for 1- minimum income conditions, 2- "indivisibility", load gradients, and/or scheduled stop. The introduction of these conditions was introduced as a tool for generating units to mitigate risks faced by a simple auction with increasing participation of variable RES-E. If a minimum average price is for a bid is not met for the time interval, the bid is removed from the matching process, creating a gap between the accepted offers and the offers submitted, as shown in Figure 2.5.

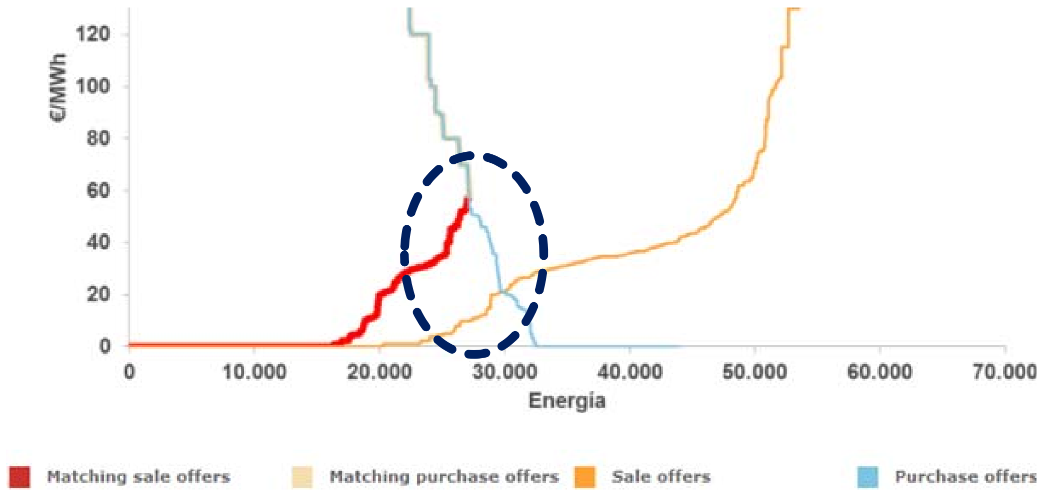


Figure 2.5: Effect of complex bid conditions on the aggregate demand supply curve of the MIBEL market. ***Source:*** **(OMIE, s.f.)**

2.2.3 Intraday Market

Once the day-ahead market has cleared, six adjustment market sessions are held. These intraday markets allow buyers and sellers to adjust their generation and consumption schedules to their best forecasts for real-time needs by submitting bids for the purchase and/or sale of electricity to modify their schedules. In Spain, these session are the means that generators can adjust

The intraday markets end at 12:45 p.m. the following day, so adjustments are possible only up until that time. On average, the intraday markets in 2013 traded 16.67% of the total energy managed on the daily market (OMIE, s.f.).

	SESSION 1°	SESSION 2ª	SESSION 3ª	SESSION 4ª	SESSION 5ª	SESSION 6ª
Session Opening	17:00	21:00	01:00	04:00	08:00	12:00
Session Closing	18:45	21:45	01:45	04:45	08:45	12:45
Matching Results	19:30	22:30	02:30	05:30	09:30	13:30
Reception of Breakdowns	19:50	22:50	02:50	05:50	09:50	13:50
Publication PHF	20:45	23:45	03:45	06:45	10:45	14:45
Schedule Horizon (Hourly periods)	27 horas (22-24)	24 horas (1-24)	20 horas (5-24)	17 horas (8-24)	13 horas (12-24)	9 horas (16-24)

Figure 2.6: Timing and structure of the Spanish intra-day session. ***Source:*** **(Red Electrica de España, 2016)**

2.2.4 OVERVIEW OF GENERATION SCHEDULING

The actual scheduling of generation is the result of a combined effort between the market and system operator following a sequence of events that lead the process from economic merit-order dispatch determined by the market operator, through to a technically feasible dispatch and real-time balancing of the system conducted by the system operator. This process is described below.

The market operator runs the day-ahead markets and intra-day markets without consideration for technical network constraints, which are later handled through the system adjustment service markets managed by each Iberian system operator (REE and REN for Portugal). In Spain, if the production schedule resulting from the day-ahead market, referred to as "PBF", does not comply with technical restrictions, so if network constraints exist and/or reserve requirements are insufficient, the system operator re-dispatches generation based on a specific procedure detailing the management of technical constraints [14]. Re-dispatching results in a revised and technically feasible schedule (or "PVP") for each hour of the following day.

Based on the aforementioned procedure, the need to procure reserves that guarantee the availability of sufficient on-line reserve margins to supply demand in real-time has increased significantly. The need for more reserves is thought to be caused by the growing penetration of RES-E generation which has increasingly displaced thermal generators from the day-ahead market**Invalid source specified.**. Thus, on May 2012, the Spanish system operator began procuring additional reserves through a so-called market of "Additional Upward Reserves" (RPAS)[15].

After obtaining a technically feasible schedule by resolving the technical network constraints, the system operator holds several ancillary markets intended to assure that sufficient regulating reserves (frequency) are available to balance the demand and generation in real-time. These markets include the secondary regulation market, which is a reserve market, as well as the tertiary regulation and the so-called deviation market, both energy markets. Voltage control services are also ensured by the system operator yet these will not be the focus of this discussion.

---

[14] Red Eléctrica de España REE, Operation Procedure 3.2: Technical constraints.

[15] Reserva de Potencia a Subir

Once the day-ahead market is cleared and network/supply constraints are resolved, agents can adjust their schedules in the intraday market to compensate for equipment failures and energy forecast errors, or to apply strategic modifications. Traditionally, only conventional generators could provide balancing services and intermittent RES-E could participate in the balancing services markets[16], thus the intraday market is the last option for these producers to adjust their production schedules according to updated generation profiles. This opportunity is critical for non-dispatchable RES-E producers in Spain as energy imbalances are strongly penalized and responsibility is allocated to all market participants contributing to imbalance.

The day-ahead and intraday components of the spot market are further discussed below, followed by Section **Error! Reference source not found.** with a description of the system adjustment and ancillary services run by system operator.

Figure 2.7 below illustrates the sequence of processes and events in the short-term market leading up to the real-time delivery of scheduled generation and load/generation balancing. The timing of these processes is critical to agent´s bidding strategies.

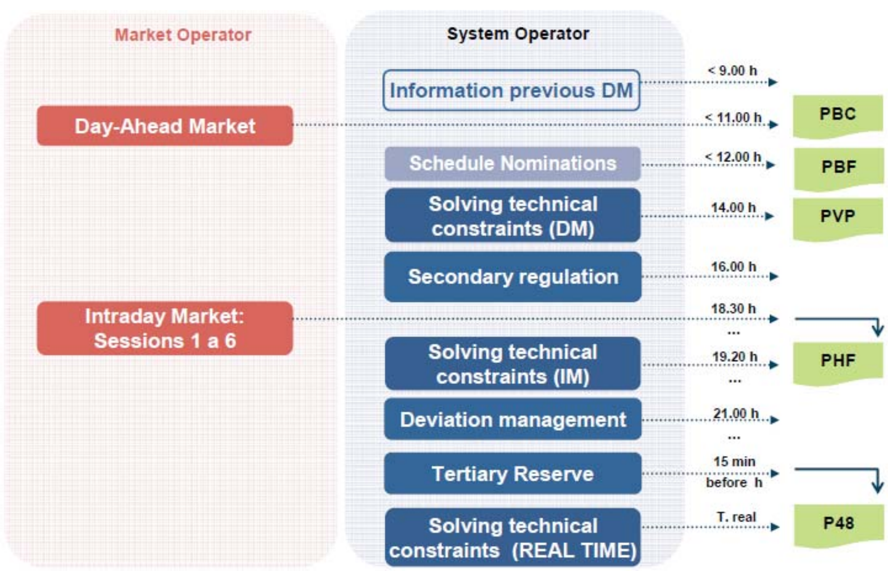


Figure 2.7: Sequence of events in the Spanish short-term electricity market. ***Source:*** **(de la Fuente, 2009)**

[16] It is worth noting that by a Resolution dated December, 18 2015 adopted by the Spanish State Secretary of Energy, certain criteria and processes were established for RES-E to participate in the provision of balancing services (BOE-A-2015-13875, www.boe.es/diario_boe/txt.php?id=BOE-A-2015-13875).

2.3 SYSTEM ADJUSTMENT SERVICES

The system adjustment services[17] managed by the Spanish system operator and address the following activities through market-based mechanism for active power:

- ➔ Resolution of technical constraints (Technical Constraints Resolution Market)
- ➔ Procurement of balancing services[18] (deviations between scheduled and measured energy are addressed through these markets):
  - i. Secondary (regulation) Reserves (capacity and energy)[19]
  - ii. Tertiary (regulation) Reserves (energy)[20].
  - iii. Management of large load deviations (energy, real-time).
  - iv. Additional "upwards" reserve (market for capacity).

There are other markets managed by the system operator, such as balancing of reactive power, which are not discussed herein. Figure 2.8 contains figures relevant to the the management of these services in 2013-2014.

	2013		2014		Δ % 2014/2013	
	Upwards	Downwards	Upwards	Downwards	Upwards	Downwards
Supply guarantee constraints [2]	4.085	-	3.260	-	-20,2	-
Technical constraints [3]	7.240	193	9.571	110	32,2	-42,9
Additional Upward Power Reserve [4](GW)	3.010	-	4.279	-	42,2	-
Secondary reserve availability [5] (MW)	691	512	677	502	-2,1	-1,9
Secondary reserve usage	1.806	1.070	1.746	995	-3,3	-7,1
Tertiary reserve	3.330	1.812	3.066	1.765	-7,9	-2,6
Deviation management service	2.347	905	1.865	571	-20,5	-36,9
Real time constraints [6]	558	1.701	556	1.274	-0,5	-25,1

Figure 2.8: System Adjustment Services in the Spanish peninsular electricity system (GWh), for years 2013-2014. Source: (Red Electrica de España, 2016)

The cost recovery of the balancing services procured by the system operator is designed to provide appropriate incentives for market participants to balance their scheduled generation and loads. This has led to the introduction of a dual imbalance charge to determine the *settlement of imbalances*. Although schedules have to be submitted on a unit-

[17] Official State Bulleting No. 303 from 19/12/2015 governing the regulation of system adjustment services (ancillary services) and their related operational procedures.

[18] Also referred to as ancillary services in some of the literature.

[19] Referred to as Frequency Restoration Reserves under ENTSO-E´s Network Code's definition.

[20] Referred to as Restoration Reserves under ENTSO-E´s Network Code's definition.

by-unit basis, to enable system operator to manage system security, the settlement of imbalances is done on an aggregate portfolio level.

As shown earlier, the processes managed by the market operator and system operator are not co-optimized, thus close collaboration between them is required for final scheduling of generation and real-time balancing and delivery. The timing of these events indicate the availability of new information, which is critical for market agents' bidding strategies and decisions, and as a result also of particular interest to this project. A more detailed representation of the timing and sequence of events is shown by Figure 2.9: Timing of the market and System operator markets in Spain. *Source:* Figure 2.9 below.

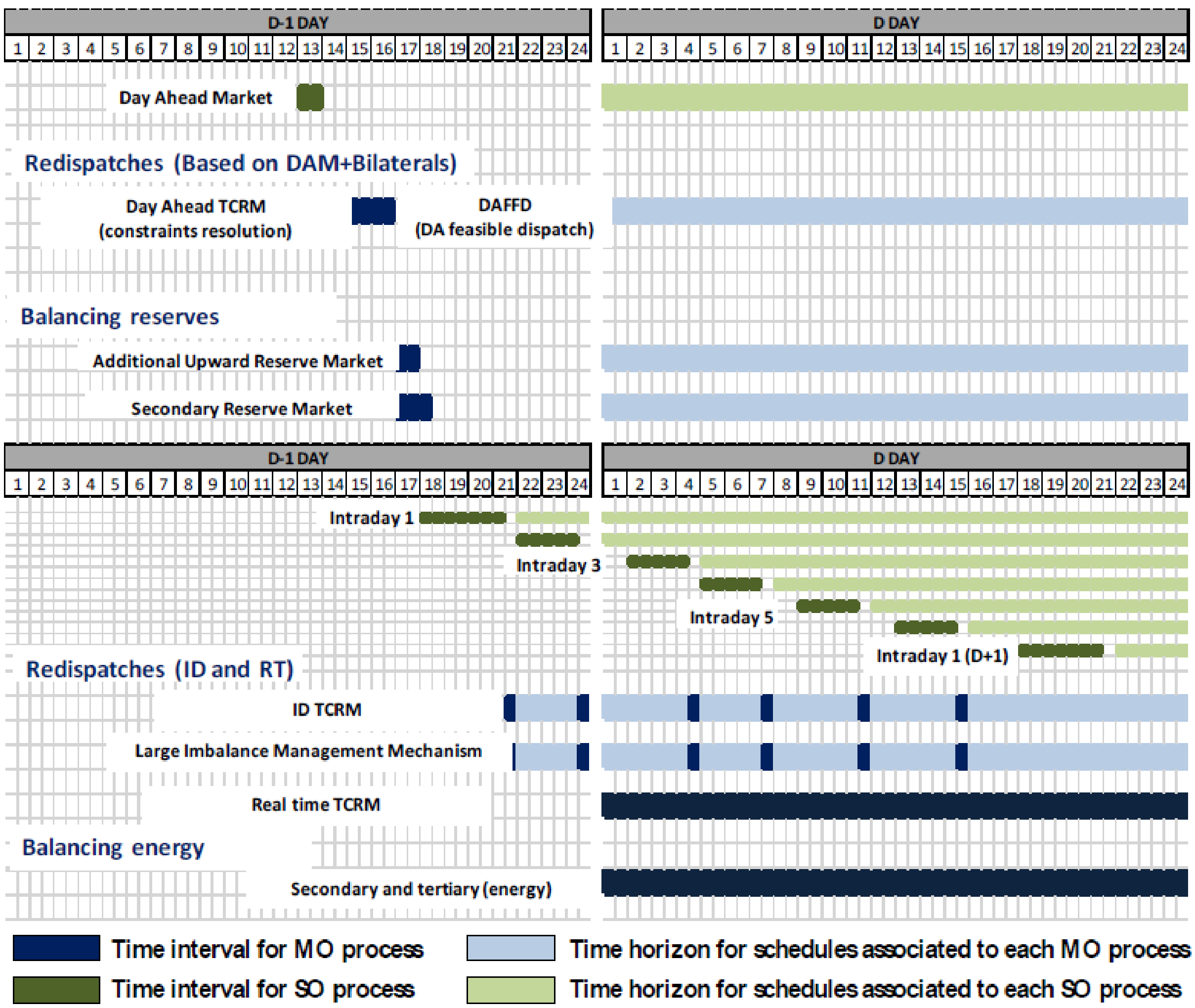


Figure 2.9: Timing of the market and System operator markets in Spain. ***Source:*** *Invalid source specified.*

These market mechanisms are further discussed below.

2.3.1 Technical Constraints Resolution Markets

Once the day-ahead market is cleared, the system operator must verify that those economic transactions are actually physically. Upon receipt of the PBF schedule published by the market operator, the system operator will determine its technical feasibility by considering the network's physical properties which may limit its delivery. Such constraints represent any circumstance were delivery of the base schedule negatively impacts the safety, security, and/or reliability of the system, according to pre-established limits and criteria; for example: line overloads, low transmission grid voltage, congestion of interconnections, or insufficient reserve power margins in the PBF schedule. Resolution of the latter condition is discussed under the section corresponding to additional upwards power reserve.

The system operator will modify the base schedule by increasing and/or decreasing the scheduled generation, as necessary, to obtain a technically feasible delivery schedule which maintains the demand-generation balance critical for real-time delivery. To do so, a market mechanism is employed through a two-phase process were the technical constraints are resolved using a least-cost approach. In this market - Technical Constraints Market - generators offer to reduce their base production PBF schedule by re-buying the energy in this market, while the remaining generators offer their additional un-matched capacity. Due to the priority dispatch for RES-E, reductions in production mainly affect thermal power plants.

The two-phase analysis of the technical constraints is as follows:

Phase 1

Resolves the technical constraints identified after the system operator performs a power flow model of the system using the PBF. Energy is increased and/or decreased to resolve the constraints. A least-cost economic criteria is, in principle, applied to resolving the constraints. This means that if, for example, two units can resolve the constraint, the unit which can do so imposing the least-cost to the system will be selected. Prices in this market are based on a pay-as-bid system so the price settled for each awarded participant is their bid price. In the case that units are withdrawn from the PBF schedule; the day-ahead price is used.

Phase 2

This phase is used to re-balance the generation and demand imbalances caused on the PBF by the process in phase 1. As is done in Phase 1, a least-cost economic criteria approach is applied to increase and/or decrease energy as necessary. In order to avoid new technical constraints from arising in this re-balancing process, the system operator imposes limitations on certain units' ability to modify their production program as part of Phase 1. The price for increasing or decreasing electricity is again pay-as-bid.
Figure 2.10 graphically represents both phases of the process.

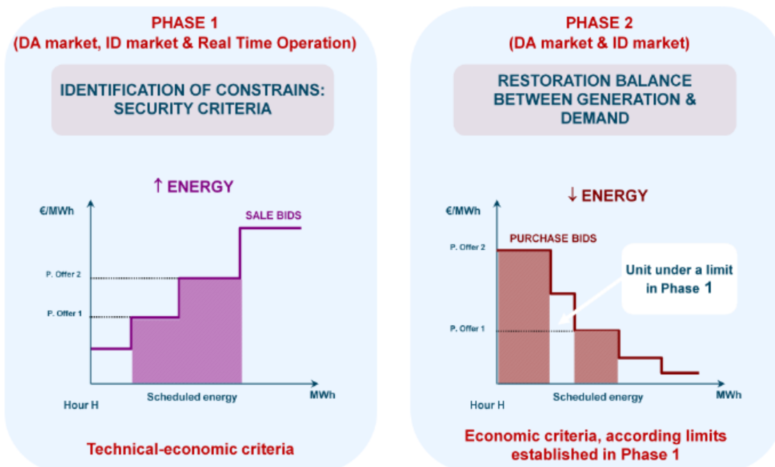


Figure 2.10 Phase 1 and 2 process. Invalid source specified.

It is important to emphasize that generators which have their day-ahead generation schedule reduced by this procedure receive 15% of the day-ahead market price for this energy reduction (or the bid price to participate in this procedure). In this situation, the generators could offer energy to the intraday market at prices lower than their marginal costs. Generators which reduce their scheduled production due to security of supply constraints management are not financially compensated for that reduction.

Although not a *technical* constraint, the management of security of supply constraints is another process that is considered by the system operator. It specifically pertains to the priority dispatch guaranteed to select generation units powered by Spanish coal[21]. It does not affect RES-E as it can only replace production from other thermal power plants.

[21] It was established by the Spanish government through the Royal Decree 134/ 2010 of February 2010 and published as an operational procedure of the Spanish system operator on October 2010.

Generators which reduce their scheduled production due to security of supply constraints management are not financially compensated for that reduction.

Figure 2.11 below shows the sequence of decisions faced by system operator to determine the feasible production schedule (PVP).

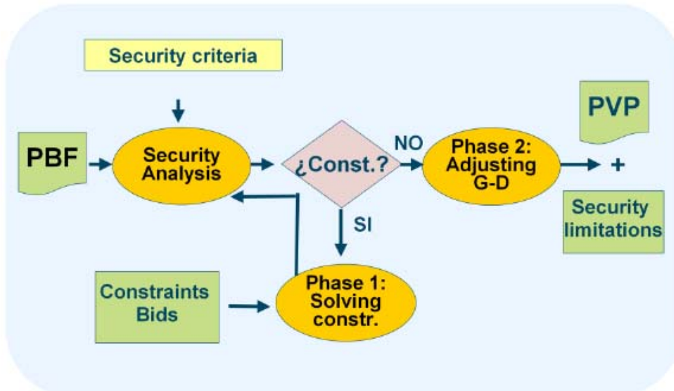


Figure 2.11: Resolution of Technical Constraints. Invalid source specified.

Intraday - Units affected by dispatching may still participate in subsequent intraday sessions as long as their participation does not create new constraints. Minimum and maximum zonal production constraints determined from day ahead market resolution process are attached to the operation of the intraday market. After each intraday session, however, another short-term security analysis is performed to consider the program changes resulting from that session.

The costs resulting from application of this process is entirely born by the demand.

2.3.2 Balancing Services

The secure operation of power systems requires that supply and demand are balanced continuously. In order to guarantee this balance, the system operator procures balancing services.

Four types of main reserves are utilized in Spain for balancing active power: primary, secondary, tertiary and slow reserved (addressed through the *deviation management* mechanism) Figure 2.12 provides a general definition of these regulation services. It is worth noting that primary regulation is a mandatory, unpaid service in Spain. It is important to note that in Spain, the provision of primary regulation is a mandatory and non-remunerated service, and equivalent to 1.5% of the unit´s nominal power.

Type	Definition
Primary Regulation	Action of speed regulators from generator units responding to changes in system frequency (<30 s to 15 minutes)
Secondary Regulation	Automatic and hierarchical control that faces changes in system frequency and power deviations with respect to France-Spain exchange program. (≤100 s to 15 minutes)
Tertiary Regulation	Manual power variation with respect to a previous program in less than 15 minutes. (<15 min to 2 hours)
Slow reserve	Running reserves of connected thermal units (30 min. to 4-5 hours)

Figure 2.12: Definition of System balancing services. ***Source*** **(de la Fuente, 2009)**

2.3.2.1 *Additional Upward Reserve*

The "additional upward reserve" market or RPAS[22], as it is commonly referred to in Spain, was established in 2012 by the system operator to address situations of low running reserve margins[23] on an as-needed basis. Running or spinning reserve is the on-line reserve that is synchronized to the grid system and ready to meet electric demand.

Through the technical constraints resolution process described earlier, the system operator determines whether there are sufficient tertiary reserves will be available for real-time operation.

If, after incorporating the re-dispatches necessary to resolve the technical network constraints, the system operator identifies that tertiary reserve margins resulting from the PBF are insufficient for real-time operation (since tertiary reserves are procured in real-time), additional thermal unit groups will be brought online to satisfy the reserve margins necessary. To do so, and since 2012, the system operator will call up the "additional

[22] Red Electrica de España Operation procedure 3.9 for the procurement of additional upward reserve (in Spanish), Boletin Oficial del Estado 190. Ministry of Industry, Energy and Tourism; August 2013. www.ree.es/sites/default/files/01_ACTIVIDADES/Documentos/ProcedimientosOperacion/RES%20PROOPE%2020130801%20PO_3.1_2_8_9.10_Modificacion_cambio_hora_cierre_MIBEL.pdf

[23] Red Electrica de Espana REE, Operation Procedure 3.2: Technical constraints.

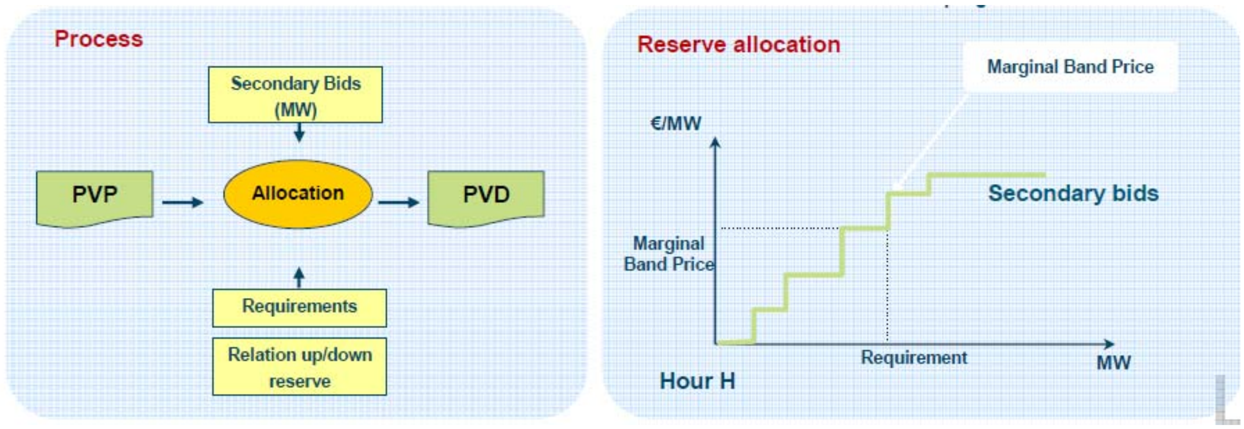
upward reserve" market to bring additional generators online and provide the necessary level of running reserves for real-time operation. All available thermal units not previously committed in the day-ahead market, and those with a program that declines during the first 3 hours of the delivery day are eligible to participate in this market.

Those units who are assigned to provide the service are scheduled at their minimum technical requirement because their cold start-up can often take longer and, should the need arise, their production into the grid must be immediate and safe. Therefore they must offer at least their minimum technical requirement in the intra-day market, resulting in their submittal of offers as "price takers", to ensure their availability to provide the service, should it be needed. If they are not be committed in the intraday market, the system operator must be made aware immediately of the situation so it can carry out a re-dispatch**Invalid source specified.**. In this case their minimum technical requirement not be committed in the market, the unit will be economically responsible for the re-dispatching costs.

2.3.2.2 Secondary Reserves

Eligible agents are able to submit offers for secondary regulation consisting of a band of power (MW) and a price (€/MW) for each settlement period (24 hours) of the following day. The secondary regulation band links upward and downward reserve. The relationship between these bands should be equal to the ratio between the total upward and the downward reserve required by system operator for the whole system. The offer can contain different price-quantity blocks, with the possibility of defining one of these blocks as "indivisible".

Secondary reserve offers are selected based exclusively on the capacity price (no energy price is taken into account). Units cleared in the auction receive the marginal price for capacity reserve, which is the same for upward and downward reserves.



2.3.2.3 TERTIARY REGULATION

This is complimentary optional service but with a mandatory bid, managed and remunerated through market mechanisms. When dispatched, the energy must be sustainable for two hours. Tertiary regulation energy is dispatched in real time based on submitted energy price bids (upward and downward).

The tertiary regulation bid consists of a price-quantity bid, where two additional constraints can be included: a ramp-up/ramp-down limit and an indivisibility condition. Bids are submitted right after the day-ahead market is closed. As mentioned earlier, tertiary reserves bids are mandatory for all generators with tertiary energy is. Bids can be updated up to 1 hour ahead of real time. The figure below shows the process and allocation of this service.

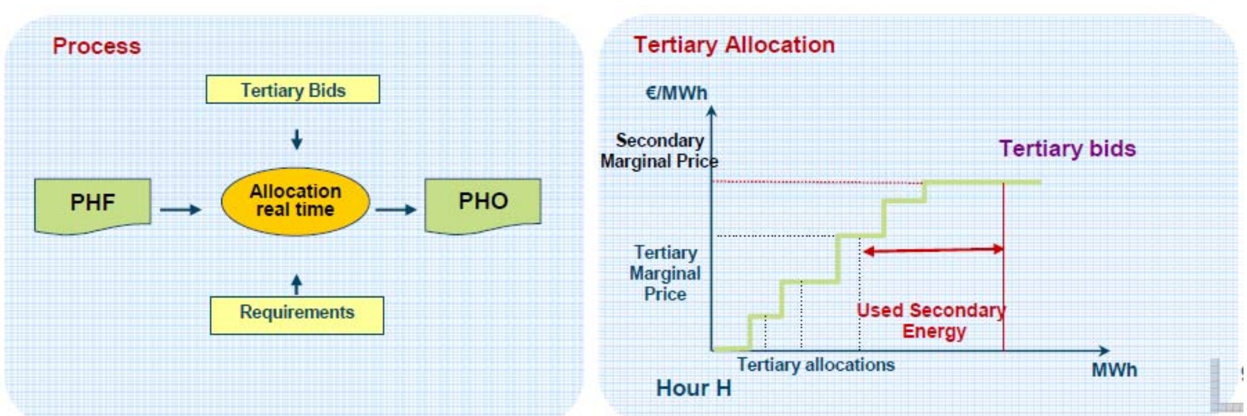


Figure 2.13: Tertiary Reserves procurement process and allocation. ***Source:*** **(de la Fuente, 2009)**

Generally speaking, demand cannot participate in balancing markets. That said, demand service interruption can be used by the system operator to provide tertiary (upwards) reserves when certain conditions are met in real-time operation (both technical and economical).

The settlement period is one hour. This is a long time step for the dynamics of electricity systems in this time scope, because, for example the direction of the system imbalance/energy needs can change within that time frame. This is represented in the left part of the figure below **Invalid source specified.**.

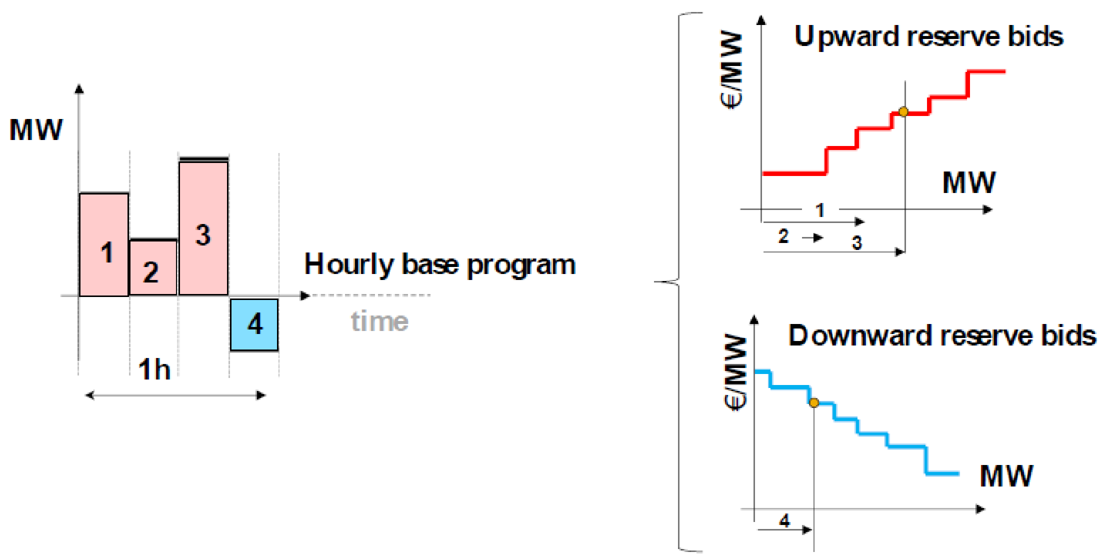


Figure 2.14: Determination of the tertiary regulation energy price. ***Source:****Invalid source specified.*

The hourly marginal prices (upwards and downward) correspond to the maximum upwards and downward energy usage during the settlement period. Bids are selected by merit order and receive the marginal price (upward or downward). The right part of the figure illustrates the determination of the marginal price in each settlement period.

2.3.2.4 DEVIATION MANAGEMENT

In addition to primary, secondary and tertiary regulation, an additional reserve of active power called deviation reserves can be used. Deviation reserve helps to balance large differences (greater than 300 MWh) between scheduled generation and forecasted demand. It covers the period between intraday market sessions, participation is optional, and allocation is assigned on an economic merit order basis. An hourly marginal price is offered to remunerate the service whose cost is paid by market players who deviate from their schedule. The market's procurement process is depicted in Figure 2.15.

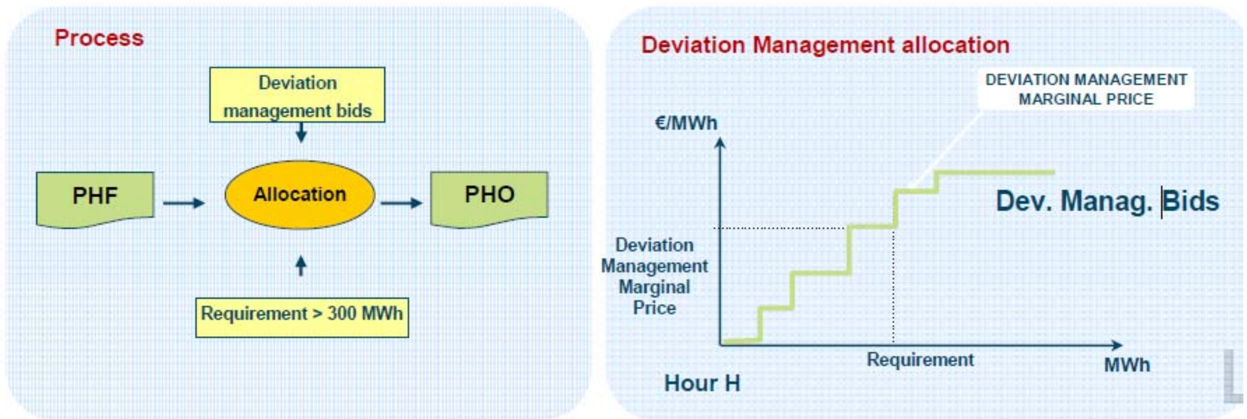


Figure 2.15: Deviation management market procurement process. ***Source: de la Fuente, 2009.***

2.3.3 IMBALANCE PRICING MECHANISM

All generators, including wind and solar, are responsible for the costs of any schedule deviations and for paying for the costs of the balancing energy necessary. A penalty is applied if the individual unit´s scheduled deviations are opposite/against the system´s needs. The Spanish electricity market applies a dual-pricing mechanism, where a higher price is charged for market participants that are short of power in real time than is offered to market participants that are long in the same instance.

The cost of procuring balancing services is allocated to the imbalanced market parties (i.e. parties that deviate from their schedule) through the imbalance settlement. Balance responsibility defines the obligation of market participants (generators, consumers and traders) to send schedules (for both consumption and production) to the system operator and the financial responsibility for deviating from those schedules. In this regard, market participants are BRPs.

Therefore, if:

i. Unit produces less than what it is scheduled for
   a. If deviation supports the system - it pays the daily market price (DMP) for balancing generation if the deviation supports the grid (i.e. system´s needs where less production).
   b. If the deviation is opposing the system's needs- it pays the maximum between average "upward" price of energy used to meet system needs or DMP.

ii. Unit produces more than what it is scheduled for.

a. If deviation supports the system's needs – it receives the DMP for the excess energy produced;
b. If the deviation is opposing the system's needs- it receives the average of the "downward" price paid to generators not to produce or the DMP, whichever is lower, for the excess energy produced.

For clarity, a graphic representation of these conditions is contained in Figure 2.16 and Figure 2.17. Figure 2.16 can be used to interpret the sign convention used in the data published by REE: 1- A positive *system* imbalance means that the system experienced an energy deficit were (there was more consumption than production) and negative means there was an excess of energy produced (more production than consumption); and 2- Upwards *unit* imbalance means the generator produced more than its schedule, and conversely, downwards *unit* imbalance means the produced less than its schedule. After determining the *unit* imbalance condition (opposing or supporting) with respect the system, Figure 2.17 can be used to determine the type of price (penalty or DMP) that would be applied. As is shown in this latter figure, the balancing cost or "penalty" assigned to the BRP who deviates from its schedule opposing the system will consider the cost for energies used to provide secondary regulation (SR), tertiary regulation (TR), and under the deviation management (DM) mechanism.

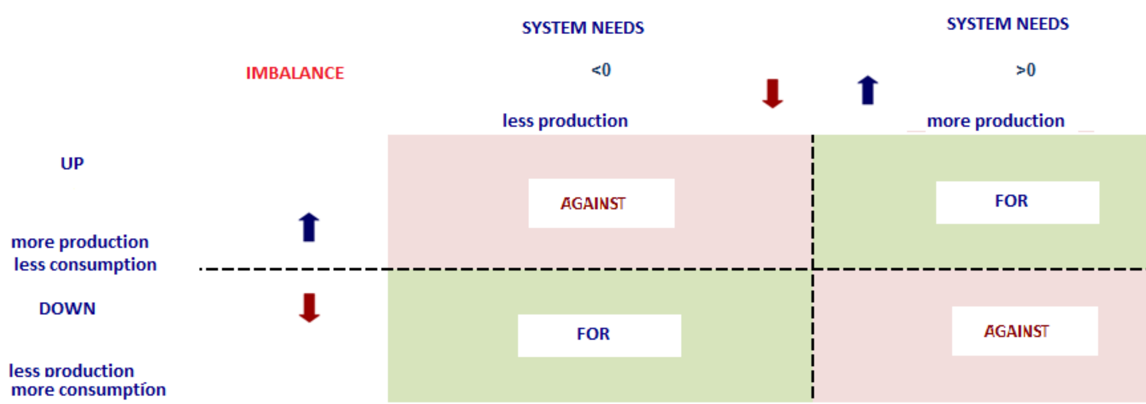


Figure 2.16: System Needs and imbalance direction.

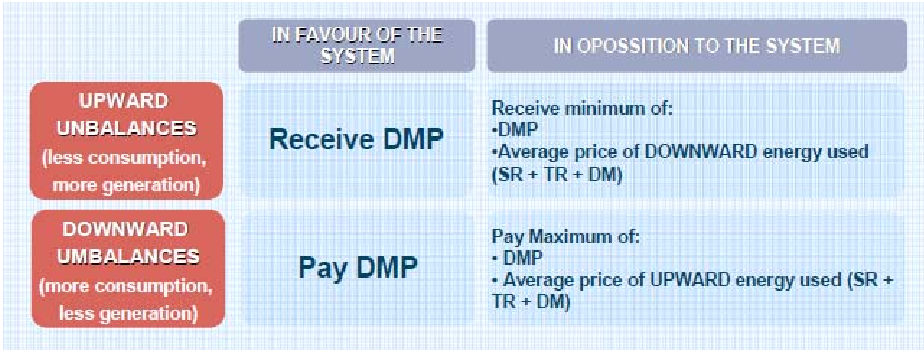


Figure 2.17 : Imbalance price settlement for units which deviate from their scheduled. ***Source:*** **(de la Fuente, 2009)**

CHAPTER 3: SYSTEM IMBALANCE AND MATHEMATICAL TOOLS

In this Chapter, the system imbalance variable is introduced, along with the sources of the system imbalance, and a brief description the influence that RES-Es may have on the variable. An overview of previous efforts to forecast the system imbalance is provided, along with a description of the mathematical tools based on machine learning techniques (random forest and genetic algorithms) applied to develop the bidding strategy model, ending with a description of the computing tools and data used to support the model.

3.1 INTRODUCTION

The system imbalance is caused by both supply and demand side factors. In Spain, intermittent renewables are considered to be the source of much the system's imbalance.

As a result of liberalization and, to another extent, the increasing integration of intermittent RES-E, forecasting has not only become a more critical and larger topic for the power industry, but also a more complex one. The system imbalance volume is a highly non-linear, noisy, and unstructured variable. Given its complexity, traditional tools may not be sufficient to accurately forecast the imbalance volume. Machine learning techniques, such as random decision forests – an ensemble learning method and the technique of choice in the forecasting component of this project – encompass predictive modelling approaches which can improve on forecasts to further optimize a market participant´s bidding strategy. Random forest is one of the most popular learning methods and has many ideal properties, namely its robustness, stability, and competitive accuracy relative to other machine learning algorithms, which are of interest to our project.

Another tool are genetic algorithms (GAs) as an alternative to traditional optimization methods. It is possible to use GA techniques to consider problems which may not be modelled as accurately using other approaches, making it a promising approach for our project.

3.2 System Imbalance

The schedule deviation, or imbalance volume of a BRP is the difference between the planned net electrical energy exchange with the power grid over its entire energy portfolio (as specified in the energy schedule) and the actual net electrical energy exchange, which is measured in real-time.

Balancing power is used to stabilize the active power balance of integrated power systems on short time scales from seconds to hours. In AC power systems, the demand-supply balance has to hold at every instant of time to ensure frequency stability at, usually, 50 Hz or 60 Hz. Frequency deviations have a number of problematic consequences, one being that they can mechanically destroy rotating machines such as generators. Technical procedures and economic institutions have evolved to prevent frequency instability, and the most important of these is “balancing power”.

3.2.1 Sources of Imbalance

There are several factors which cause imbalances in electricity systems, as classified in Table 3.1. These variables are driven by either the supply or demand side of the electricity value chain, and based on either forecast errors, as in the case of RES-E generation and the load, unplanned outages from the supply side, and “schedule leaps” from both sides (Batalla-Bejarano, et al., 2015). “Schedule leaps” refer to the deviations of *actual* load and production from the *scheduled* load and production. One more source are deviations from standard losses, as not all energy produced by generators arrives to consumers: the energy metered at distribution entry points does not match energy at metered distribution exit points due to network losses (Batalla-Bejarano, et al., 2015) .

	Variable	Imbalance source
Supply	Conventional generation	- Unplanned plant outages - Schedule leaps
	VRES generation	- Forecast errors - Schedule leaps
	Interconnectors	- Unplanned line outages - Schedule leaps
Demand	Load	- Forecast errors **- Deviations from standard losses** - Schedule leaps

Figure 3.1: Sources of system imbalance. ***Source*****: Batalla-Bejarano, et al., 2015.**

In Spain, the BRPs will slightly over contract because it is less risky to have overcapacity than under capacity (as confirmed by our analysis of the imbalance cost in the next chapter).

Since the 2009 disappearance of the DSO´s role as supplier in Spain, the grid energy losses are estimated based on a regulated coefficient of losses which retailers use to estimate the amount of energy necessary to supply their load (in Spain, losses are allocated to each consumer taking into consideration their consumption characteristics). However, that estimation does not necessarily coincide with the actual energy that is dispatched, and an inherent imbalance exists. Batalla-Bejarano, et al., 2015 explain that, from a regulatory perspective, "the electricity imbalances resulting from the differences between the average transport and distribution losses and the standard losses used in balancing the system as a whole are considered additional system deviations".

3.2.2 IMBALANCE AND RES-E

Electricity generation from variable RES-E, such as wind and solar power, has grown rapidly during recent years and is expected to continue to grow. The fact that these generators are distributed, non-synchronous, and weather-dependent causes specific challenges when integrating them into power systems (Grubb 1991, Holttinen et al. 2011, IEA 2014a). With increasing amounts of variable RES-E, in many countries, system integration has become a major public policy debate with a particular emphasis on the stress that forecast errors put on balancing systems.

Some power systems with higher penetration levels of RES-E have observed an increase in system imbalance volumes. In the event that market rules assign economic responsibility to those market agents who contribute to the imbalance, bidding strategies should consider this variable and its associated costs. In fact, in the Spanish power system, wind generators are the source of much of system's imbalance, as indicated by the TSO (Bueno-Lorenzo, et al., 2013) (Gonzales-Aparicio & Zucker, 2015) Figure 3.2 below shows an example of the wind power forecasting errors, with the green circles highlighting areas of difference between the actual and forecasted energy.

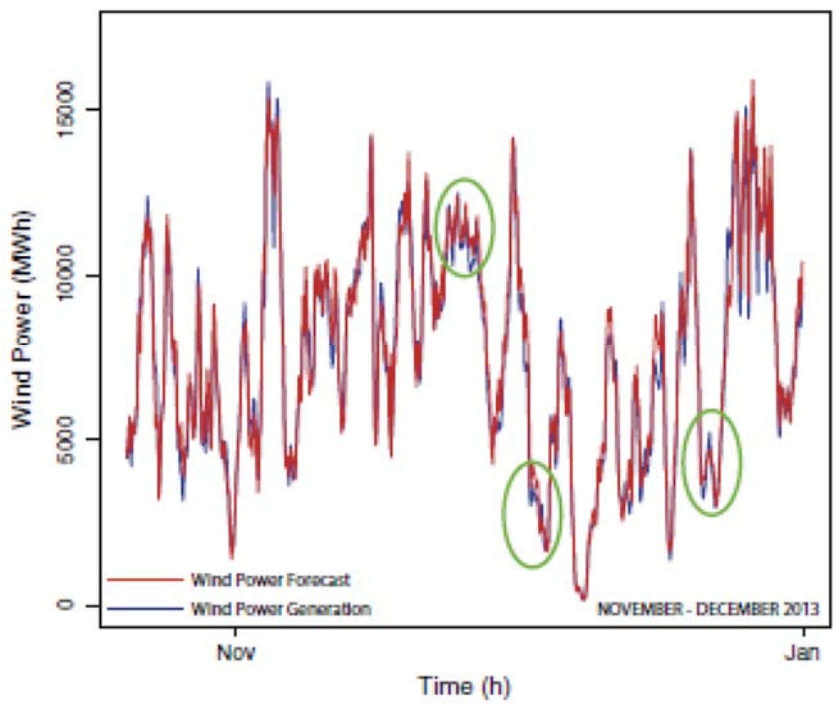


Figure 3.2: Time-series of forecasted and actual wind power generation in Spain for December 2013.Wind forecast by SIPREOLICO tool, developed by REE. ***Source:*** **(Gonzales-Aparicio & Zucker, 2015)**

3.3 Forecasting System Imbalance Volumes

Much of the literature and research efforts have focused on wind power forecasting and its associated uncertainty; very few articles were found on forecasting of the system imbalance itself. This is not to say that, in practice, market participants are not using such forecasting modes, but the academic literature just has little documentation of work with forecasting that vatiable.

Traditional forecasting methods, such as autoregressive integrated moving average (ARIMA) and exponential smoothing, are limited to predicting values for one variable based on its previous values. The system imbalance, however, does not satisfy the assumptions to predict with ARIMA technique (e.g. stationarity condition). As demonstrated in their time series analysis of the system imbalance volume, Garcia et al (Garcia & Kirschen, 2006) showed how this variable had no seasonality, no constant mean, and a constant noisy structure, concluding that past imbalance volume data by itself was not a good predictor of future values. Another disadvantage of traditional methods is that they may not represent the nonlinear characteristics of complex variables (Cheng, et al., 2012) such as the system imbalance volume. Traditional multivariate techniques, such as least square linear regression models, are limited in their ability to detect non-linear relationships of the predictor variables.

Addressing the above noted complexities, Garcia et al (Garcia & Kirschen, 2006) use artificial neural network techniques to forecast the system imbalance volume. The forecasts using neural networks, an advanced modeling technique based on artificial intelligence, yielded better results than conventional forecasting techniques. Artificial neural networks increase the forecasting accuracy because the models represent nonlinear relations between the variable to be predicted and its influencing factors.

However, there are still some problems for artificial neural networks, such as slow convergence in training and the need for manually determining the structure and parameters (Cheng, et al., 2012). To overcome these and other challenges, Cheng et al propose using random forest technique for forecasting of another time-series market variable (the short-term load).

Garcia et al (Garcia & Kirschen, 2006) noted the system imbalance volume as being "noisy, unstructured, changing, and normally distributed". Random Forest

Learning approach methods are often referred to as artificial intelligence methods. These advanced modelling techniques are called learning approaches because they learn from the relationship between the observed values target variable (the variable we want to predict) and the predictor variables (variables influencing the target variable). Random forests are an increasingly popular machine learning method for classification and regression problems. In this project, the technique has been selected for its robustness, stability, and competitive accuracy relative to other machine learning algorithms

3.3.1 Characteristics of Decision Trees

For comparison purposes, Table 3.1 depicts some characteristics of learning methods, including neural nets, super vector machines (SVM), decision trees (Trees), multiple additive regression tree (MART), and k-NN Kernels: the green symbol indicates good performance, the yellow is fair performance and the red represents poor performance.

Characteristic	Neural Nets	SVM	Trees	MARS	k-NN, Kernels
Natural handling of data of "mixed" type	▼	▼	▲	▲	▼
Handling of missing values	▼	▼	▲	▲	▲
Robustness to outliers in input space	▼	▼	▲	▼	▲
Insensitive to monotone transformations of inputs	▼	▼	▲	▼	▼
Computational scalability (large N)	▼	▼	▲	▲	▼
Ability to deal with irrelevant inputs	▼	▼	▲	▲	▼
Ability to extract linear combinations of features	▲	▲	▼	▼	◆
Interpretability	▼	▼	◆	▲	▼
Predictive power	▲	▲	▼	◆	▲

Table 3.1: Some characteristics of different learning methods. ***Source:*** **(Hastie, et al., 2008).**

Decision trees come closest to meeting most of the criteria listed in the Table above. By itself, a single decision tree is very easy to interpret, it handles mixed data types well (i.e. mix of binary, quantitative, and/or categorical variables), it can handle missing data which is often the case with large data sets, it is immune to the effects of predictor outliers, and it is invariant under transformations of individual predictors (e.g. scaling is not an issue) (Hastie, et al., 2008). However, its accuracy is noted as lower than the other methods and it is relatively unstable.

As an ensemble method, a random forest is a collection of *individual* trees built based on random samples of training data, whose output using new/test data is averaged to obtain a prediction. The idea behind this was that many trees together have better generalization capabilities than one tree. This approach satisfactorily improves on the accuracy and stability of a single decision tree, while compromising only on its interpretability (not easy to interpret). Its accuracy is competitive versus other state of the art machine learning algorithms, and if the data changes, an individual tree may also change, but the forest as a whole is relatively stable as the output is the combination of the different trees. This has made random forest one of the most successful ensemble methods: it is fast, robust to noise, and is less prone to overfitting[1] - an issue that has to be carefully considered with most learning methods.

3.3.2 Random Forest Process

An ensemble method means a model with several "sub-models" within. Developed by Leo Breiman at UC Berkley in 2001, the Random Forest technique is based on the principle of decision (classification and regression) trees (Breiman, 2001). The idea is to average many noisy but approximately unbiased models (each tree being a model, many trees being a forest) and hence reduce the variance. The random forest algorithm grows individual decision trees through randomization using a training data set.

This method was selected based on its *robustness, stability and competitive accuracy* compared to other machine learning algorithms. It is today considered state-of –the –art in forecasting.

Supervised vs Unsupervised Learning

In machine learning, supervised and unsupervised learning are two different approaches selected based on the availability of real observed data for the element we are trying to predict. In this project, supervised learning was used. The model learns from identifying complex relationships between predictor variables and the actual observed values of the variable we want to predict (referred to as the *target variable* herein). A separate set of data containig predictor variables only is used to obtain a prediction of the target variable.

Randomeness

Randomness is a very important feature of these forests. Randomness is introduced in the algorithm through its proccess of:

- data selection - random sampling of the training data; and
- node optimization – at each split a subset of the predictors is chosen and only from this subset the best split variable is taken at each node of each individual tree.

[1] Random Forest is less prone to overfitting because it uses a separate training set for each tree.

Growing Decision Trees

The random forest technique is based on growing an ensemble of decision trees which are completely independent from each other, so there is no interaction between them (see Figure 3.3). This project uses regression trees (other objectives may use classification trees – for binary predictions as in the case of image recognition, clustering, semi-supervised learning, and others). In regression trees, each tree fits a linear model, and each node optimizes a continuous function. Each tree can be a piecewise linear function.

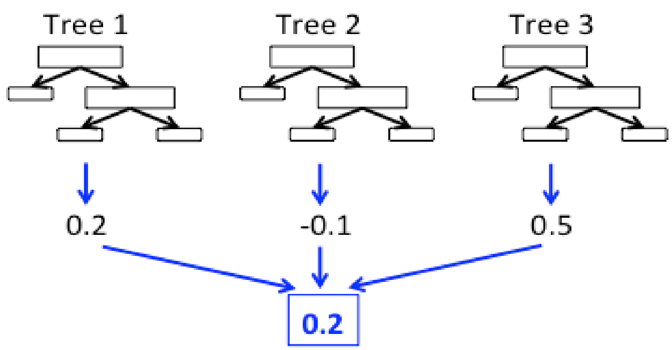


Figure 3.3: Example for regression trees in ensemble model

The process is as follows:

- Each tree is grown individually (no link with other trees).
- Each tree will only see a subset of the data (selected randomly, hence "random" forest).
    - X = matrix with no. of observations.
    - Features are selected randomly.
    - *m* variables of the *p* variables (or features) are selected at random
    - The best variable/split point among the variables is selected for each tree.
        - For best information gain: node optimization is performed.
    - The node is then split into two children nodes.
- The endpoints/leafs of the tree have a probability.
- Output of algorithm is the trees

Making a Prediction

Using a second subset of data known as the test data set, which includes only the predictor variables, a prediction is made by:

- The same random process explained above, each vector in the matrix is run through the previously built trees until the terminal leaf. See Figure 3.4 for example of a regression tree.
- This prediction is then averaged (sum of trees divided by the number of trees): a mean prediction (see Figure 3.3);
    - This means high variance, low bias trees that are very different from each other (uncorrelated to one another), thus averaging them should provide a function closer to the true value. See Figure 3.5 for a graphical example plot of the outcome of this process.

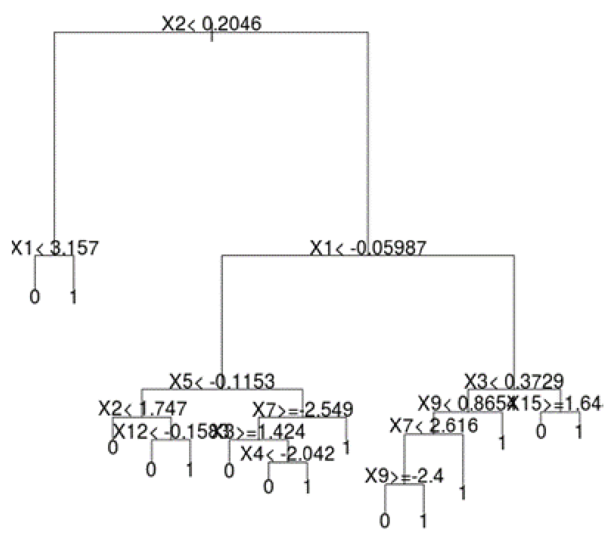


Figure 3.4: Single regression tree example

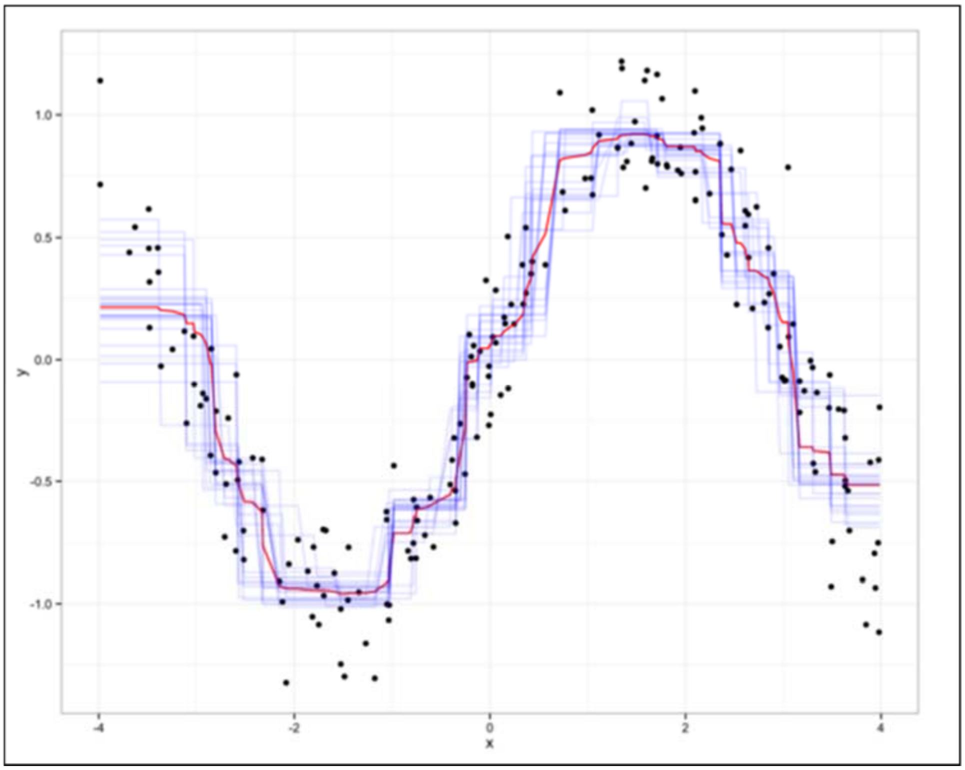


Figure 3.5: Example of 25 randomly selected decision trees in a random forest, prediction shown in red.

3.4 Genetic Algorithm Optimization

In the field of artificial intelligence Genetic algorithms (GAs) are stochastic search algorithms inspired by the basic principles of biological evolution and natural selection. GA's simulate the evolution of living organisms, where the fittest individuals dominate over the weaker ones, by mimicking the biological mechanisms of evolution, such as selection, crossover and mutation. GAs have been successfully applied to solve optimization problems, both for continuous (whether differentiable or not) and discrete functions (Scrucca, 2013).

At a certain stage of evolution a population is composed of a number of individuals, also called strings or chromosomes. These are made of units (genes, features, characters) which control the inheritance of one or several characters. Genes of certain characters are located along the chromosome, and the corresponding string positions are called loci. Each genotype would represent a potential solution to a problem.

The decision variables or phenotypes in a GA are obtained by applying some mapping from the chromosome representation into the decision variable space, which represent potential solutions to an optimization problem. A suitable decoding function may be required for mapping chromosomes onto phenotypes.

The fitness of each individual is evaluated and only the fittest individuals reproduce, passing their genetic information to their offspring. Thus, with the *selection operator*, GAs mimic the behavior of natural organisms in a competitive environment, in which only the most qualified and their offspring survive. Two important issues in the evolution process of GAs search are *exploration* and *exploitation*. *Exploration* is the creation of population diversity by exploring the search space, and is obtained by genetic operators, such as *mutation* and *crossover*. *Crossover* forms new offsprings from two parent chromosomes by combining part of the genetic information from each. On the contrary, *mutation* is a genetic operator that randomly alters the values of genes in a parent chromosome. *Exploitation* aims at reducing the diversity in the population by selecting at each stage the individuals with higher fitness.

Often an elitist strategy is also employed, by allowing the best fitted individuals to persist in the next generation in case they did not survive.

The evolution process is terminated on the basis of some convergence criteria. Usually a maximum number of generations is determined. Alternatively, a GA is stopped when a sufficiently large number of generations have passed without any improvement in the best fitness value, or when a population statistic achieves a pre-determined bound.

Figure 3.6 shows the flow chart of a typical genetic algorithm, as done with the GA package for R programming language. A user must first define the type of variables and their encoding for the problem at hand. Then the fitness function is defined, which is often simply the objective function to be optimized. More generally, it can be any function which assigns a value of relative merit to an individual.

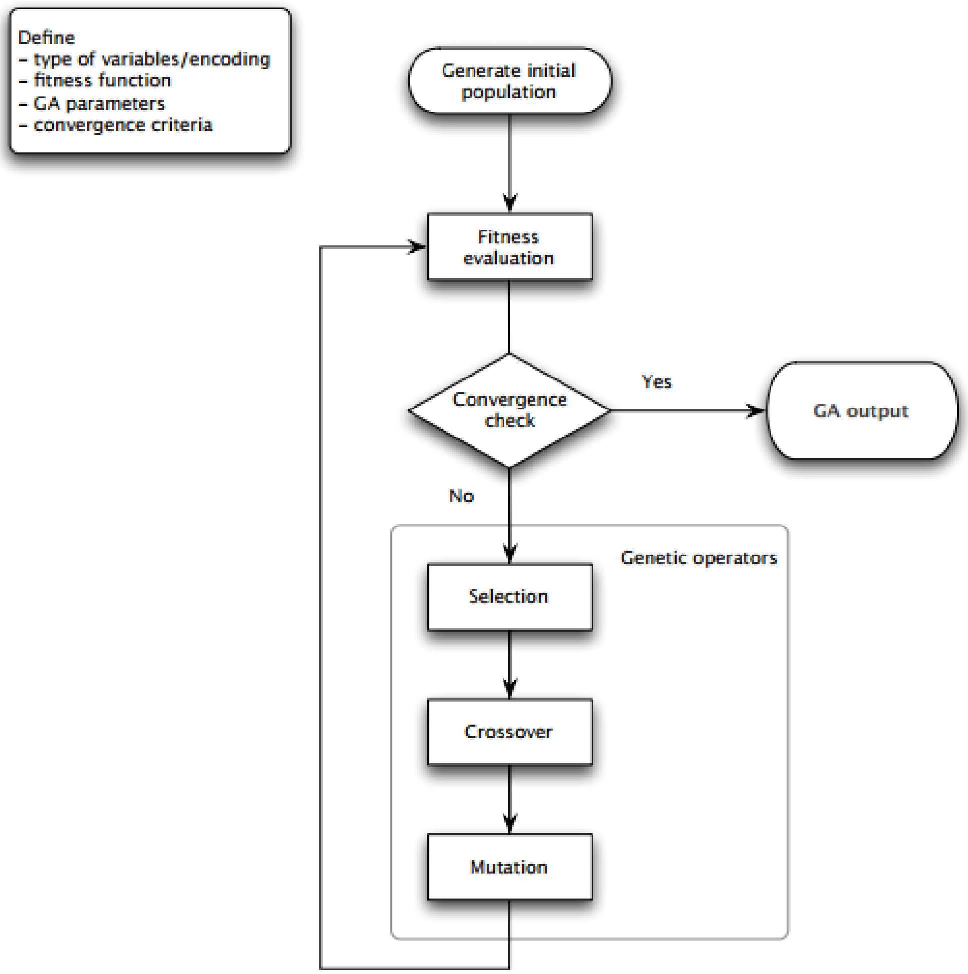


Figure 3.6: Flowchart of Genetic Algorithm. ***Source*** **(Scrucca, 2013)**

Genetic operators, such as crossover and mutation, are applied stochastically at each step of the evolution process, so their probabilities of occurrence must be set. Finally, convergence criteria must be supplied.

The evolution process starts with the generation of an initial random population of size *n*. The fitness of each member of the population at any step *k*, is computed and probabilities are assigned to each individual in the population, usually proportional to their fitness. The reproducing population is formed (*selection*) by drawing with replacement a sample where each individual has probability of surviving equal to its probability. A new population is formed from the reproducing population using *crossover* and *mutation* operators. Then, $k = k + 1$ is set and the algorithm returns to the fitness evaluation step. When convergence criteria are met the evolution stops, and the algorithm delivers the optimum. Figure 3.7 shows the fitness landscape with different evolutionary paths.

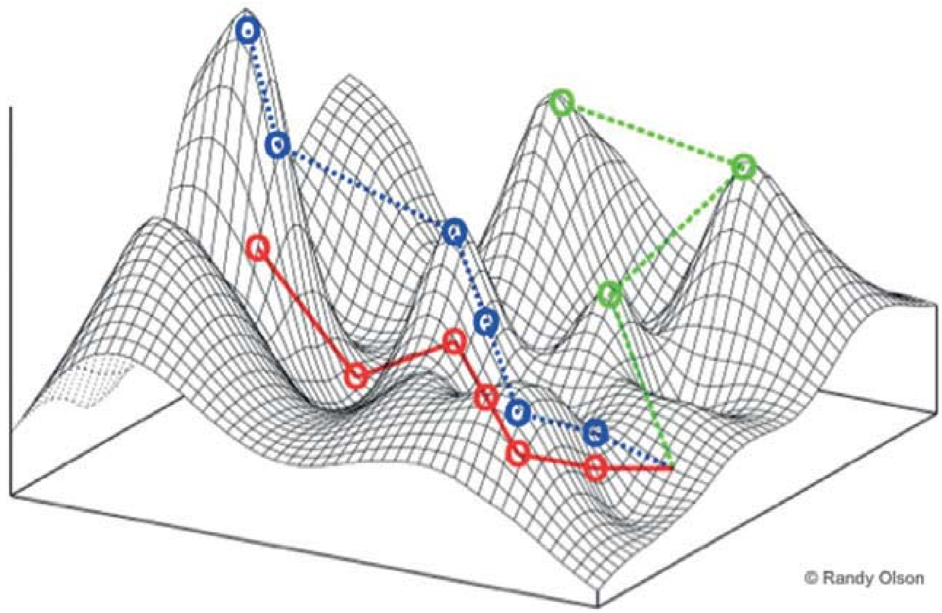


Figure 3.7: Fitness landscape with different evolutionary paths. ***Source: Randy S. Olson*[2].**

3.5 COMPUTING TOOLS AND DATA

R programming language and a range of library packages to perform specific functions were used to develop the model. The main R library packages used include:

- Lubridate – to handle time series date and time data.
- RDOBC – to access SQL sever database
- RandomForest[3] – to generate the random forest.
- GA (Genetic Algorithm) – for optimization.

Publicly available data from published by REE, OMIE, and the Spanish metereological agency (AEMET) were used, in addition to real-energy portfolio data specific to the Trader in the case study.

The data was centrally stored in an SQL server database at the Trader's facilities, and accessed therefrom.

[2] http://adamilab.msu.edu/research/

[3] https://cran.r-project.org/web/packages/randomForest/index.html

CHAPTER 4:

VARIABLE ANALYSIS

This chapter presents the variables to be used and analysis if those variables as they relate to the system imbalance and cost.

4.1 INTRODUCTION

The main components to developing the optimized bidding strategy can be broken down as follows:

1. *Exploratory data analysis of variables;*
2. Development of forecasting model;
3. Development of optimized bidding strategy application tool;
4. Validation;
5. Assessment.

The objective of analyzing the system´s imbalance volume is to conduct an *exploratory analysis* focused on understanding the behavior of variables (market or otherwise) influencing or related to the power system´s balancing mechanism and understanding, their behavior as they relate to the system´s imbalance volume and, in some instances, the cost.

The above analysis may also be used as a discriminatory step to identify the input for the *forecasting model*.

4.2 ANALYSIS OF SYSTEM IMBALANCE

4.2.1 SYSTEM IMBALANCE VOLUME

The system imbalance for the England and Wales pool was shown to have no seasonality, no constant mean, a constant noisy structure, and normally distributed (Garcia & Kirschen, 2006). Similar assumptions could be inferred from the time-series of the Spanish system imbalance volume depicted in Figure 4.1 and Figure 4.2, which show fourteen months of the hourly net imbalance volume data published by the Spanisg TSO; the time-series exhibits no obvious seasonality and appears highly-noisy.

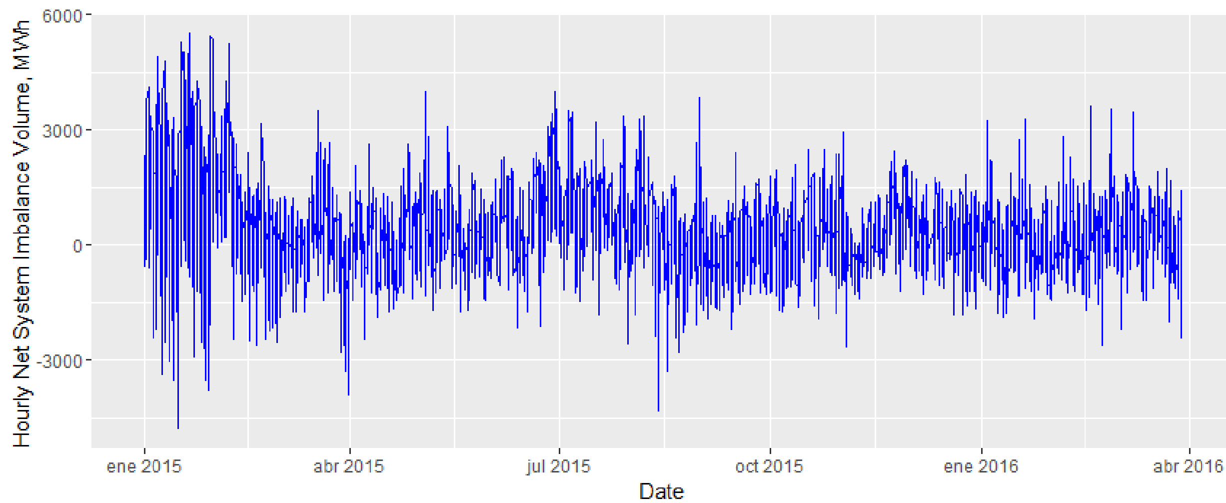


Figure 4.1: Hourly Imbalance Volume of the Spanish electricity system from Jan 1, 2015 to April 1, 2016

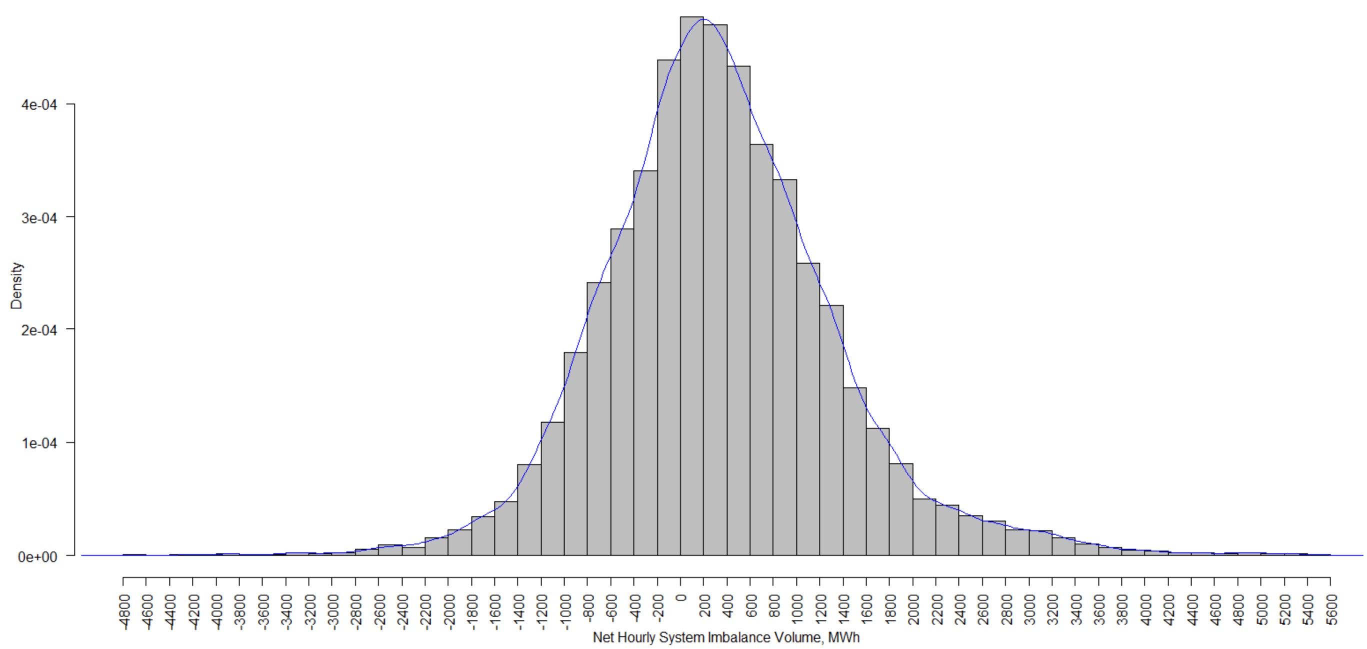


Figure 4.2: Histogram of system imbalance volumes in the Spanish electricity system from Jan 1, 2015 to April 1, 2016.

The **<u>hourly system imbalance volume</u>**, defined as the net energy needed by the system in a specific hour, is calculated as follows:

$$\Delta Psys_h = \sum_h SecReg^{u,d} + \sum_h TerReg^{u,d} + \sum_h DevMan^{u,d} \tag{4.1}$$

where:

$\Delta Psys_h = Net\ balancing\ energy\ needs\ of\ the\ system, \quad MWh$

$SecReg = Secondary\ regulation\ energy\ used, \quad MWh$

$TerReg = Tertiary\ regulation\ energy\ used, \quad MWh$

$DevMan = Deviation\ Management\ energy\ used, \quad MWh$

$h\ =\ hour$

$u\ =\ upwards\ energy\ (energy\ increase\ -> \ positive\ sign)$

$d\ =\ downwards\ energy\ (energy\ decrease\ -> \ negative\ sign)$

As a result, the direction of the net system imbalance volume, $IMBsys_h$, indicates whether is system was short (needed more production), or "long" (needed less energy production).

Therefore, the sign convention is as follows:

➔ "Short" = $\Delta Psys_h$> 0 or <u>POSITIVE</u> - > system needs more energy
➔ "Long" = $\Delta Psys_h$ < 0 or <u>NEGATIVE</u> - > system needs less energy

Table 4.1 Sign convention for the direction of the system imbalance

Table 4.2 contains a summary of the system´s imbalance direction for the period of June 2015 through February 2016. The system is most likely to be need energy (positive imbalance) than have an excess, with over 61% of the hours in that period having a positive system imbalance, and only over 38% of the hours observing a negative imbalance. With the exception of two months - September and February –most months displayed the same tendency of more positive than a negative system imbalance hours. This fact can be better appreciated in Figure 4.3.

	No. Obs.	Direction, %	
	Total	*ΔPsys>0*	*ΔPsys<0*
June	618	78.80%	21.20%
July	648	82.41%	17.59%
August	696	58.76%	41.24%
September	696	44.83%	55.17%
October	744	63.58%	36.42%
November	657	55.56%	44.44%
December	576	59.38%	40.63%
January	574	59.41%	40.59%
February	204	40.20%	59.80%
9-months	5413	61.80%	38.20%

Table 4.2: Per month Summary of direction of hourly system imbalance for June 2015 – February 2016.

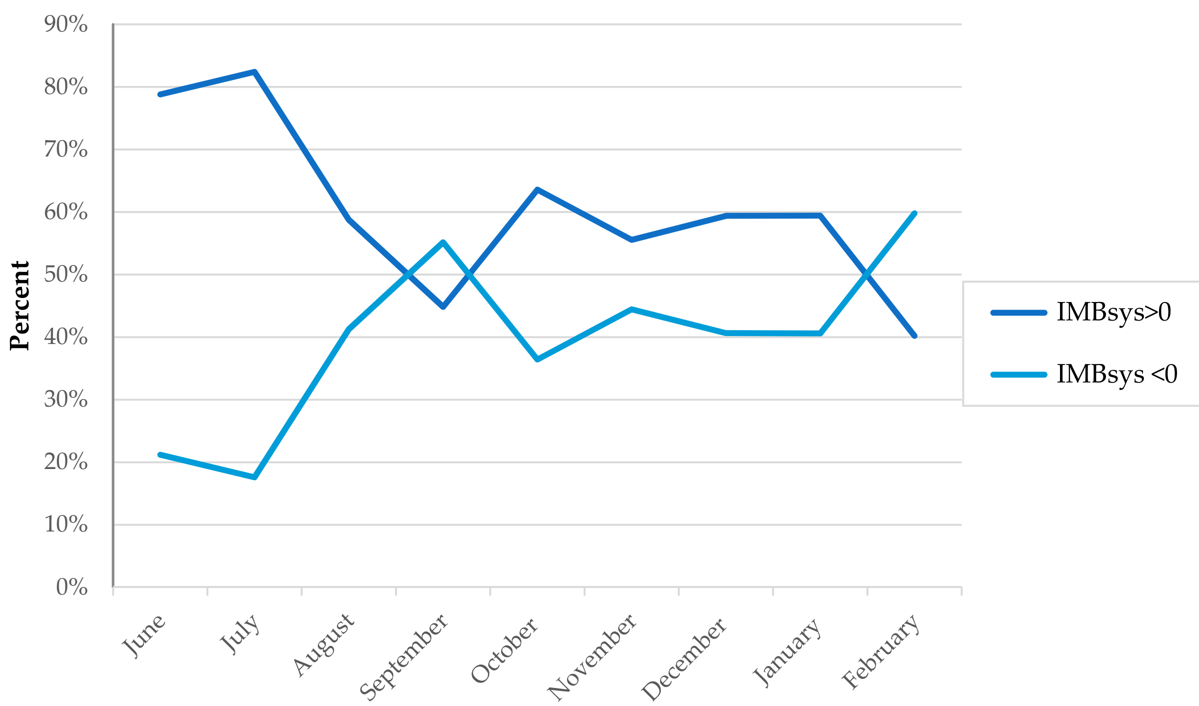


Figure 4.3: Graphical representation of monthly summary of system imbalance direction for June 2015 – February 2016.

The **<u>hourly deviation volume for each generating unit and agent´s portfolio</u>** is determined as follows:

$$\Delta Punit_h = PMESunit_{h,} - PHOunit_h \quad (4.2)$$

$$\Delta Pport_h = \sum \Delta Punit_h \quad (4.3)$$

where:

$\Delta Punit_h =$ $hourly\ deviation\ from\ scheduled\ program\ for\ a\ production\ unit, MWh$

$\Delta Pport_h = hourly\ deviation\ for\ portfolio\ of\ production\ units$

$PMESunit_h =$ $hourly\ measure\ of\ actualy\ energy\ generated\ by\ production\ unit, MWh$

$PHOunit_h$= *final hourly program for production unit after intra-day market adjustments minus energy restrictions, if any, from grid congestions, MWh.*

The sign convention of the imbalance volume for a production unit or portfolio of units ($\Delta Punit_h$ or $\Delta Pport_h$) indicates the type deviation:

- ➔ "Long" = $\Delta Punit_h$ or $\Delta Pport_h$> 0 or <u>POSITIVE</u> - > more energy produced than the program schedule; also referred to as a **"upwards" (u)** imbalance.
- ➔ "Short" = $\Delta Punit_h$ or $\Delta Pport_h$< 0 or <u>NEGATIVE</u> - > less energy produced than program scheduled; also referred to as a **"downwards" (d)** imbalance

4.2.2 SYSTEM IMBALANCE COST

In the Spanish electricity market, the imbalance prices are highly variable and difficult to forecast (Bueno-Lorenzo, Moreno, & Usaola, 2013). Due to the unavoidable imbalance between scheduled and generated energy, the imbalance prices are very important for a wind power producer because much of the source of the imbalance is due to wind. Imbalance prices also have high volatility because the number of participants and the amount of energy exchanged are relatively low and (in dual pricing systems, as is the case for Spain) because of the random nature of the overall imbalances (Hirth, Lion & Inka Ziegenhagen, 2015).

The per-unit of energy imbalance cost for a market agent is based on the imbalance price (see Chapter 2 for imbalance price determination): 1- if imbalance supports the system (same direction), the day-ahead market price is applied (not penalized) or 2- if the imbalance is against the system a price which includes a "premium" or cost in relation to the day-ahead market price - based on the cost of the balancing energies used – is applied.

<u>This cost or "premium", on a per-unit basis, for agents whose imbalance is ***against/opposite*** the system's balancing needs is the following</u>:

$$COST\Delta P_h = PRICE\Delta P_h^d - PRICE\Delta P_h^u \tag{4.4}$$

where:

$$COST\Delta P_h = per\ unit\ cost\ for\ unit\ \ or\ portfolio\ \ with\ imbalance\ against\ the\ system, € / MWh$$

$$PRICE\Delta P_h^d, PRICE\Delta P_h^u = \ Hourly\ price\ for\ downwwards\ (d)\ and\ upwards\ (u) imbalance$$

The imbalance cost is generally highest when the system imbalance is negative, meaning there was an excess of energy – the system was "long". Below is a summary of average imbalance costs for a nine-month period which shows the imbalance cost to consistently be more than twice the cost applied to positive system imbalances.

	No. Obs	**Average Hourly Imbalance Cost, €/MWh**		
		monthly	*ΔPsys>0*	*ΔPsys<0*
June	618	9.67	7.95	16.05
July	648	7.50	6.11	13.99
August	696	12.25	7.33	19.27
September	696	11.12	5.62	15.58
October	744	10.43	6.07	18.04
November	657	11.66	8.24	15.92
December	576	10.04	5.88	16.11
January	574	12.10	7.95	18.19
February	204	11.43	13.29	10.18
9-months	5413	10.63	7.05	16.44

Table 4.3: Average monthly cost of market agent´s imbalance, per-unit of energy volume.

4.3 Data Selection

The variables selected in this project are based on their availability for adjustments in the intra-day market, and include a combination of market and production variables. These variables are defined in the table below, and ordered based on the time-frame that they become available. Their values correspond to each hour of the day:

> x_h^D = *hourly value for variable X corresponding to D*
>
> *D = Represents the* ***target day,*** *encompassing its 24 hours of imbalances, for which we seek to reduce the imbalance cost.*

The variables in this section can considered as belonging to two types:

1. those available pre-day ahead market gate closure; and
2. those available post-day ahead market gate closure.

For target day *D*, values for variables 5 through 14 ($x_{v5:v14}$) are only available *after* gate closure of the day-ahead market. For pre-gate-closure, those variables correspond to D-1, ($x_{v1:v15,D-1}$).

Variable (V)	LEGEND	NAME	DEFINITION	DATA SOURCE OR EQUATION	DATA AVAILABILITY: PRE-GATE CLOSURE *(V5-V14 correspond to delivery day D-1)*	DATA AVAILABILITY: POST-GATE CLOSURE *(V5-V14 correspond to delivery day D)*
1	**DF**	Demand Forecast, MWh	Hourly forecast of demand.	Source: REE	Continuously updated	Continuously updated
2	**WF**	Wind Forecast, MWh	Hourly forecast of wind production.	Source: private third- party vendor	Continuously updated.	Continuously updated.
3	**TEMP**	Temperature, °C	Hourly actual and forecast temperatures.	$TEMP_h = Tmean_h$ Source: AEMET (Spain´s National Meteorological Agency)	Forecast is continuously updated. Actuals are available immediately.	Forecast is continuously updated. Actuals are available immediately.
4	**WindRAMPS**	Wind RAMP Forecast , MW/h	Hourly rate of increase/decrease in forecasted wind power output relative to the hour prior	WindRAMPS $= WF_h - WF_{h-1}$	Continuously updated.	Continuously updated.
5	**DemandPBF**	Demand PBF, MWh	Hourly accepted bid volume in day-ahead market (PBF).	Source: REE	D-2 at approx 12-13 hrs	D-1 at approx 12-13 hrs
6	**porcenDF**	Demand Porcent, MWh	Hourly porcent difference between the demand forecast and demand met by the day-ahead market (PBF schedule)	porcenDF = (DF - demandBF)/DF	D-2 at approx 12-13 hrs	D-1 at approx 12-13 hrs
7	**WindPBF**	Wind in PBF	Hourly accepted wind production bid volume in day-ahead market (PBF schedule).	Source: REE	D-2 at approx 12-13 hrs	D-1 at approx 12-13 hrs
8	**porcenWF**	Wind Porcentage difference in PBF, MWh	The hourly porcentage difference of forecasted wind production and accepted volume of wind production bids in day-ahead market (PBF schedule).	porcenWF = (WF-WindPBF)/WF	D-2 at approx 12-13 hrs	D-1 at approx 12-13 hrs
9	**DAMP**	Day-Ahead Market Price, €/MWh	Hourly price clearing the day-ahead market.	Source: OMIE	D-2 at approx 12-13 hrs	D-1 at approx 12-13 hrs
10	**TCEnergy**	Constraints: Energy, MWh	Energy volume auctioned in Phase II of technical constraints market.	Source: REE	D-2 at approx 15-16hrs	D-1 at approx 15-16hrs
11	**TCPrice**	Technical Contraints: Price, €/MWh	Hourly price for congestion energy in Phase II of technical contraints market.	Source: REE	D-2 at approx 15-16hrs	D-1 at approx 15-16hrs
12	**RPAS**	Additional Upwards Reserve	Hourly occurrence of Additional Upwards Reserves market.	Source: REE	D-2 at approx. 16-17hrs	D-1 at approx. 16-17hrs
13	**SRBdown**	Secondary Regulation Band: Downwards, MW	Hourly contracted capacity for downwards secondary regulation band.	Source: REE	D-2 at approx 21:30 hrs	D-1 at approx 21:30 hrs
14	**SRBup**	Secondary Regulation Band: Upwards, MW	Hourly Contracted capacity for secondary upwards regulation band.	Source: REE	D-2 at approx 21:30 hrs	D-1 at approx 21:30 hrs

Table 4.4: Variables description and availability.

4.4 VARIABLE ANALYSIS

4.4.1 CROSS-CORRELATION

A nine-month cross- correlation of the variables was performed, reults under

Variable No.	Variable	Correlation
1	DF	0.040038565
2	WF	-0.013376171
3.1	Temp (Forecast)	0.090739168
3.2	Temp (Actual)	0.089570937
4	WindRAMP	0.015570115
5	DemandPBF	-0.02569405
6	porcenDF	0.37020217
7	WindPBF	-0.021885455
8	porcenWF	0.101509643
9	DAMP	0.027242249
10	TCenergy	-0.024040173
11	TCPrice	0.030559866
12	DF	0.024333969
13	SRBdown	0.060403179
14	SRBup	-0.009596265

Table 4.5: Cross-correlation of Variables.

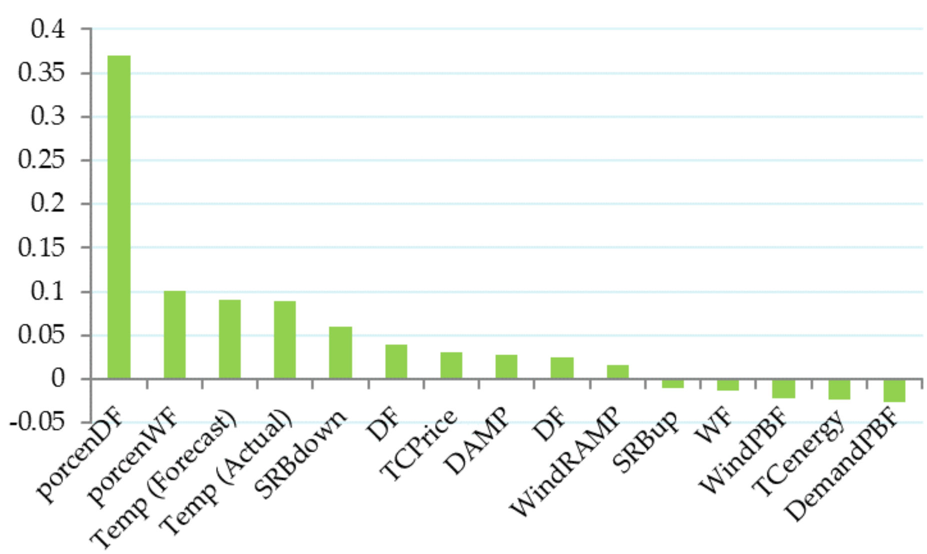


Figure 4.4: Ordered graphical representation of variable´s cross-correlation.

Annex A contains full cross-correlation results for all variables.

4.4.2 VARIABLES

4.4.2.1 DEMAND: FORECAST, IN PBF AND PERCENT IN PBF

As indicated earlier, the demand load is one the sources of the system´s imbalance. We will consider three demand related load and market variables: the demand forecast, demand load matched in the day-ahead market (PBF schedule), and the percent difference of demand forecast matched in the PBF.

The figure below is a scatter plot matrix of these variables color-coded to highlight positive imbalances in magenta and negative imbalances in cyan for the month of January. A similar plot for the 14-month period is found in Annex B.1.

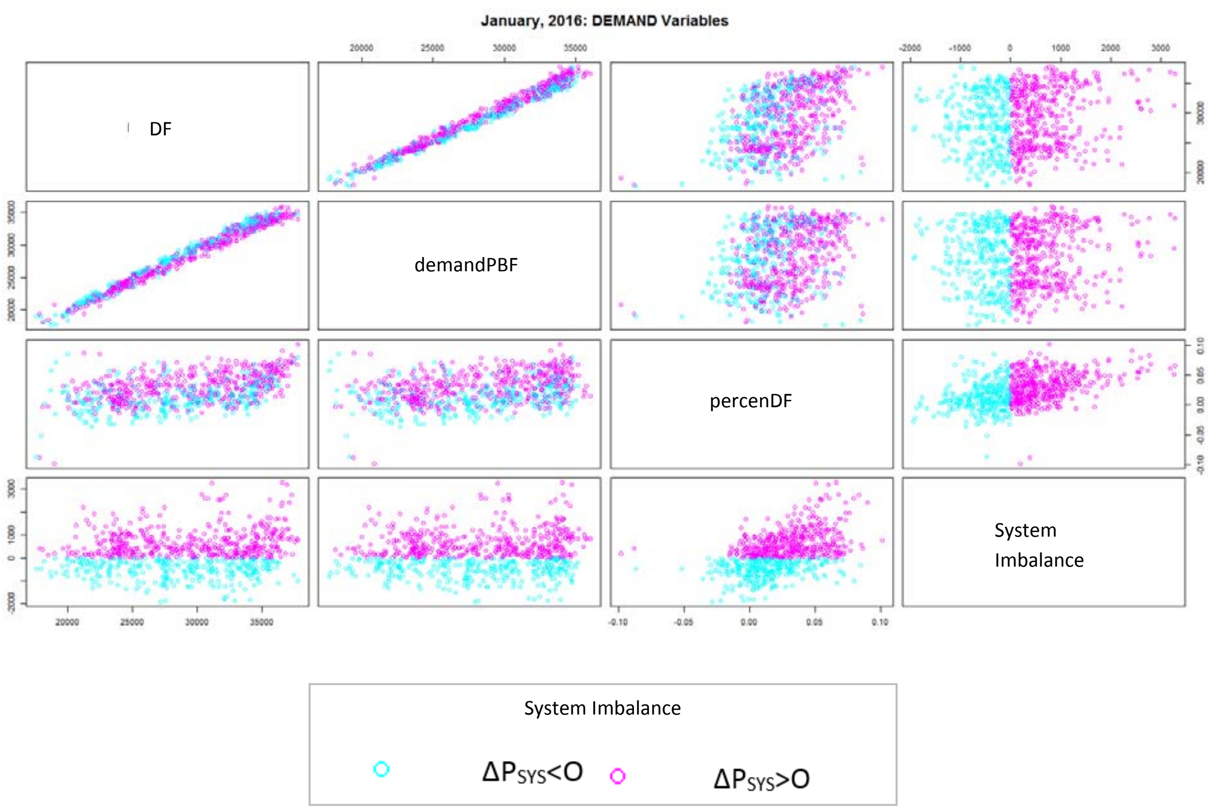


Figure 4.5: Matrix scatter plot of demand variables for January 2016.

Demand Forecast (DF)

A scatter plot of the demand forecast and the system imbalance volume is shown in Figure 4.6. Although no obvious relationship could be discerned, it can observed that during the month of January, the very large negative imbalances (greater than -2,000 MWh) occurred alongside higher demand forecast volumes (exceeding 35,000 MWh).

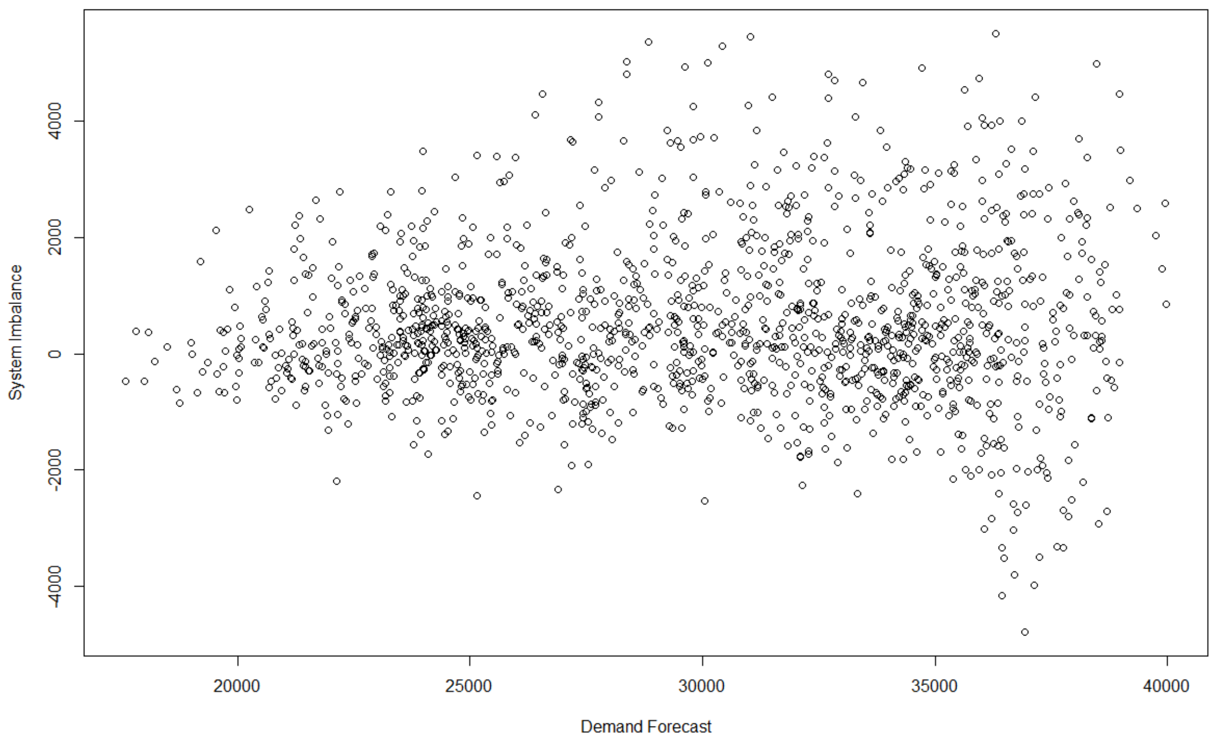


Figure 4.6: Scatter plot system imbalance volume and demand forecast for Jan 2016.

Demand met in PBF schedule (demandPBF)

As can be observed from Figure 4.7, which contains a blow up of the DF and demandPBF variables plotted in Figure 4.5, some sort of relationship between these variables and the direction of the system imbalance appears to exist. Aside from the expected linear relationship between DF and demandPBF, it seems that the occurance of negative system imbalances – in cyan - is greater the higher the value for demandPBF is in relationship to the DF variable, and vice versa (the lower the demandPBF value in relation to the DF value, the more occurrences of positive system imbalances – in magenta). This relationship is further accentuated through the analysis of the percenDF variable below.

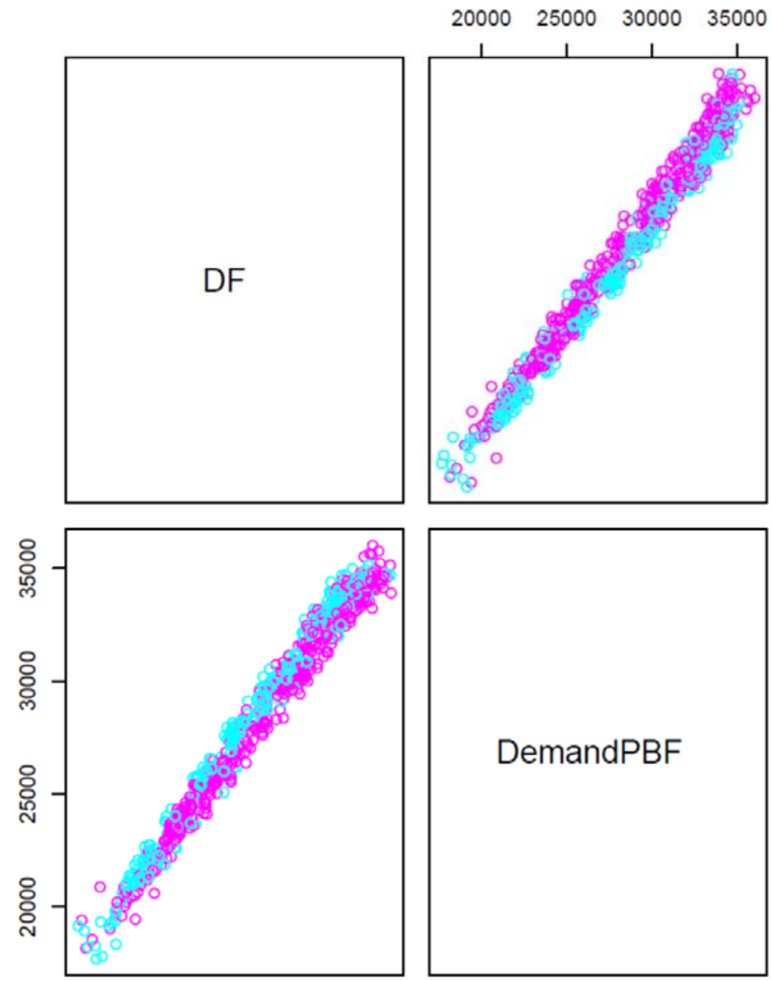


Figure 4.7: Scatter plot of the demand forecast and the demand met in the day-ahead market.

Percent difference of demand forecast met in PBF schedule (percenDF)

The percenDF variable yielded particularly interesting results. As can be observed from the histograms of this variable in Figure 4.8 and Figure 4.9 for the positive and negative system imbalances occurring during the same 14-month period analyzed earlier (January 2015 to April 2016), it appears that most of the positive system imbalances occur when the percenDF is greater or equal to 2% (or when 98% or less of the demand forecast is met in the day ahead market (PBF)). Similarly, most of the negative imbalances appear to occur when the percenDF is less than 2% (or when the demand forecast met in the day-ahead market (PBF) exceeds 98%).

Figure 4.10 is a scatter plot of the system imbalance volume and the demand forecast, with each observation color coded to represent percenDF with a threshold of 2%. From that figure, it can be noted more positive imbalances appear to occur when 98% or less of the demand forecast is met in the day-ahead market (PBF schedule).

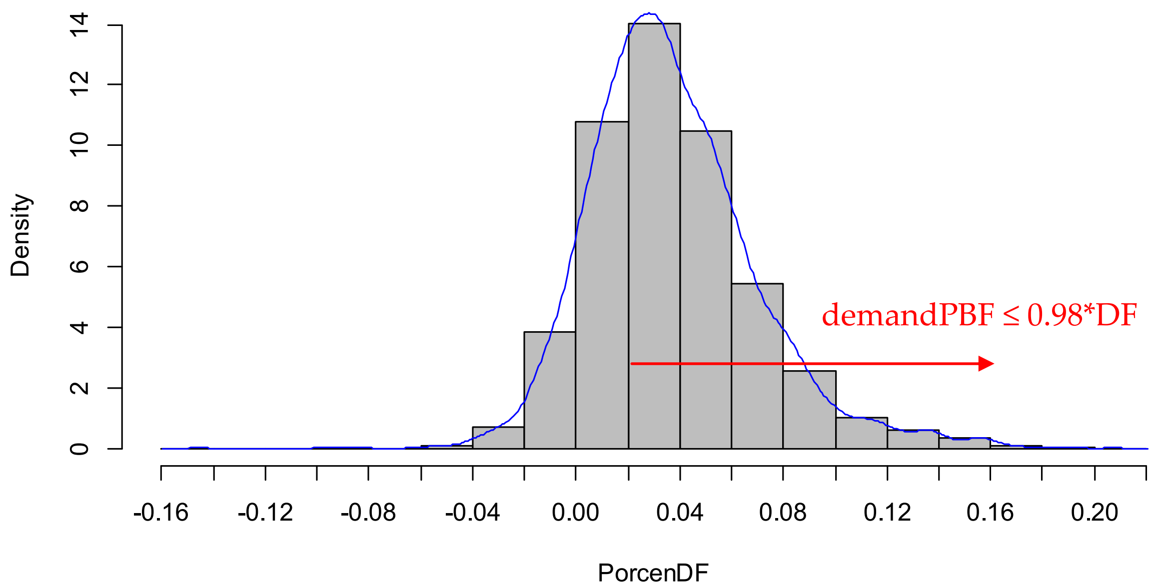


Figure 4.8: Histogram of percenDF for positive system imbalances.

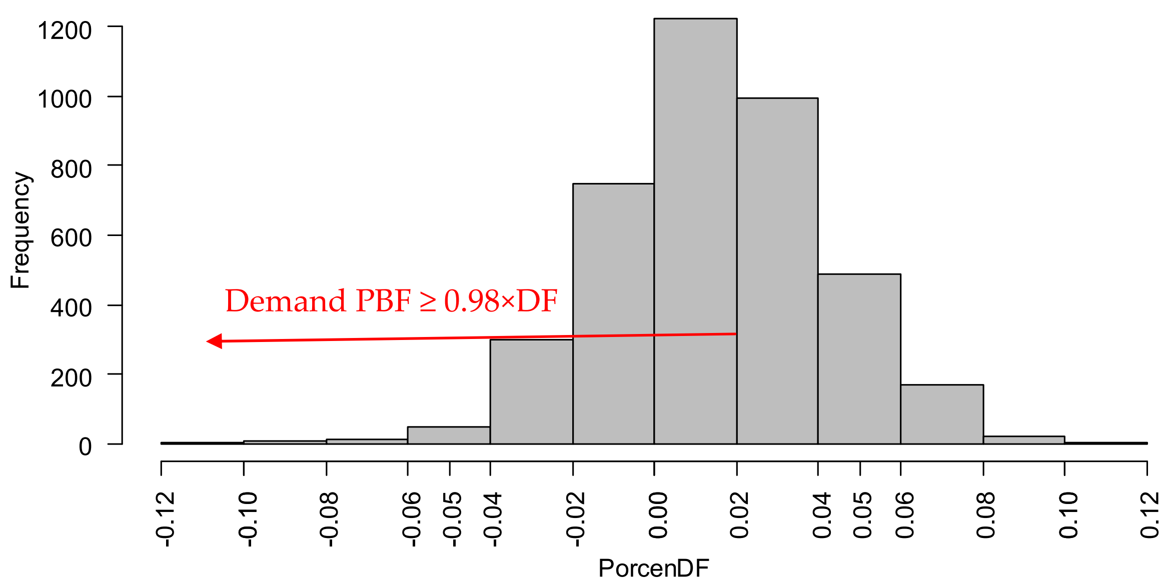


Figure 4.9: Histogram of percenDF for negative system imbalances

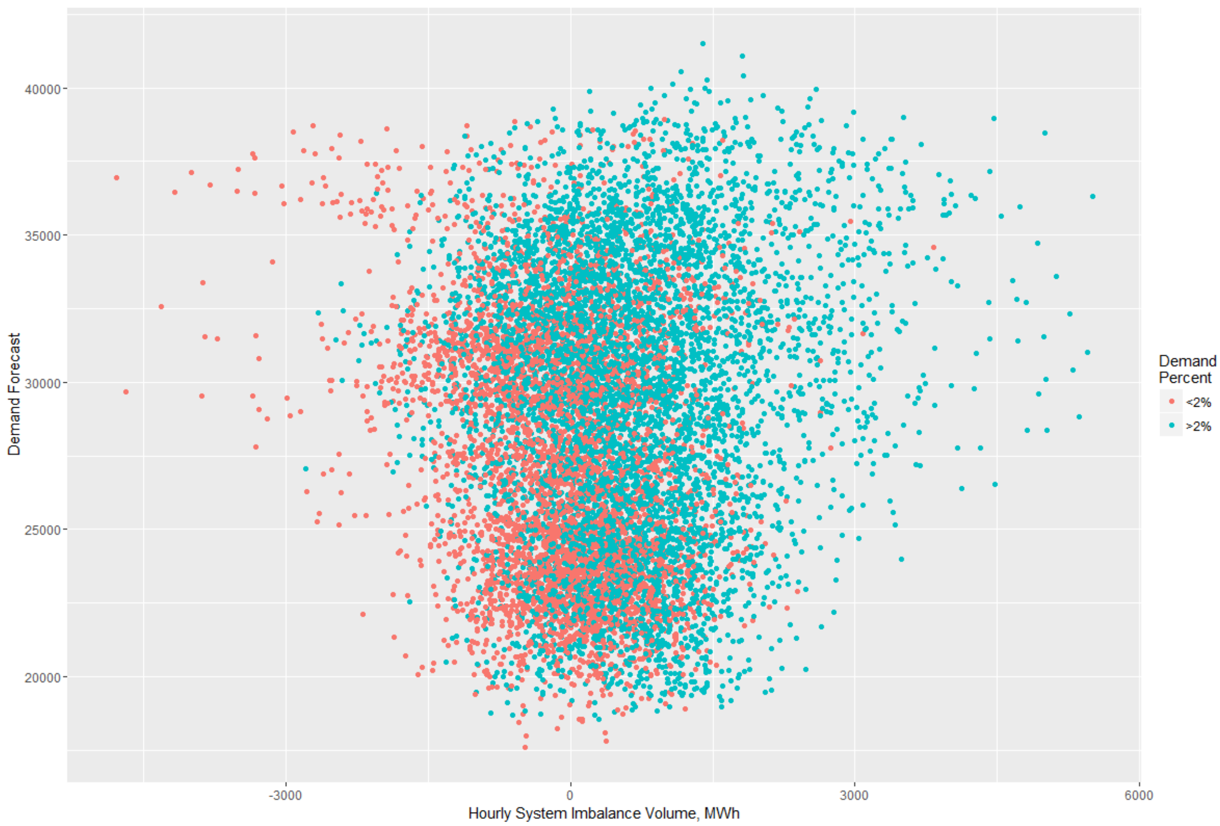


Figure 4.10: Scatter plot of the system imbalance volume and demand forecast as they relate to the percenDF variable.

4.4.2.2 WIND: FORECAST, IN PBF, PERCENT AND RAMPS

In Spain, as mentioned previously, wind production is one of the major causes of the system imbalance. We will consider four wind related production and market variables: the wind forecast, wind production bids accepted in the day-ahead market (PBF schedule), percent excess/deficit of wind forecast matched in the day-ahead market (PBF schedule), and wind ramps. Annex B.2 contains a matrix scatter plot of the these wind related variables.

Wind Forecast (WF) & Wind in PBF (windPBF)

Wind forecasts contain uncertainty which could lead to the reasonable assumption of a potential relationship between the *level* of the wind forecast and the system imbalance volume itself. However, no obvious (linear) relationships could be discerned from the scatter plot of those two variables as shown in the last plot in Figure 4.11 below.

Following the structure of previous plots in this chapter, the direction of the imbalance has been color coded with cyan for negative imbalances and magenta for positive imbalances. Figure 4.12 contains the density distribution histograms for these variables.

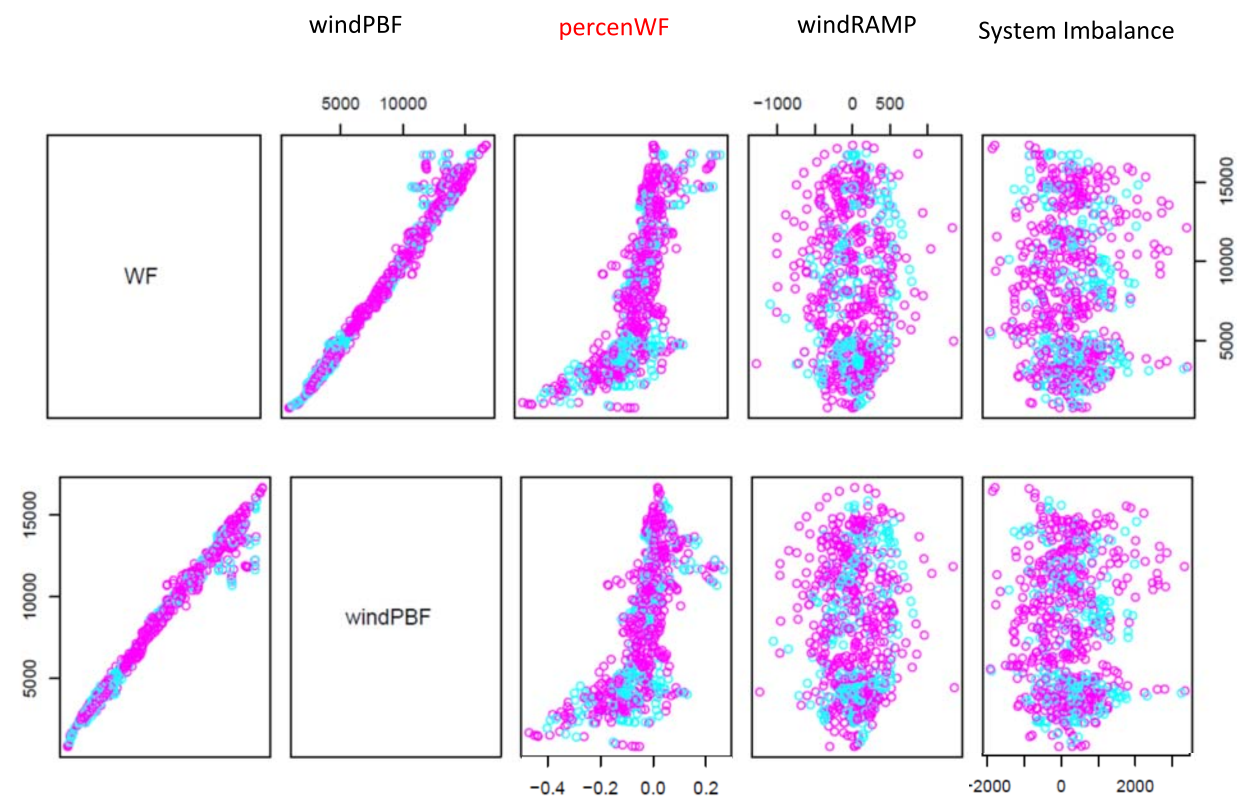


Figure 4.11: scatter plot matrix of wind forecast and wind demand met in day-ahead market variable for Jan 2016.

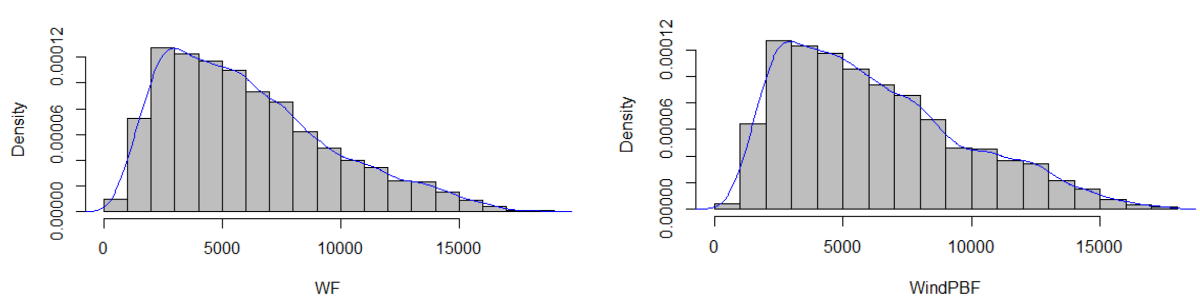


Figure 4.12: Density distribution histogram of the wind forecast and wind in PBF variables for Jan 2015 through March 2016.

Wind Percent (percenWF)

As with the "percenDF" variable, the wind production forecast in excess/deficit of the aggregated wind production offers accepted in the day-ahead market and represented by the PBF schedule, could provide indication of the system imbalance.

From the scatter plot of the wind forecast (and windPBF) with the percent difference of the wind forecast met in the day-ahead market (percenWF), depicted as the third scatter plot column in Figure 4.11, relationships could be discerned. It appears that the lower the hourly wind forecast values is, the wind demand met in the day ahead market (windPBF) will tend to exceed the wind forecast.

Basically, the percent of wind forecast met in the day-ahead market tends to increase (negative percenWF value means windPBF exceeds WF) as the wind forecast value decreases.

Wind Ramps (windRAMP)

Ramp events are a significant source of uncertainty in wind power generation. A ramp represents a large increase or decrease in wind power within a limited time view, as shown in the figure below.

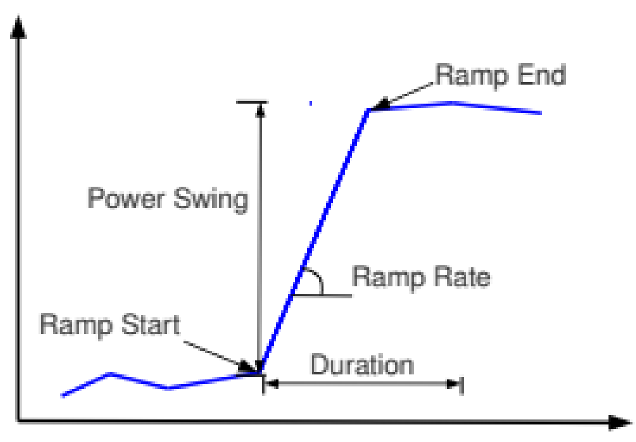


Figure 4.13 Ramp representation.

The wind ramp variable used herein is calculated as follows:

$$windRAMP_h = WF_h - WF_{h-1} \qquad (4.5)$$

Figure 4.1 below shows this density distribution of this variable.

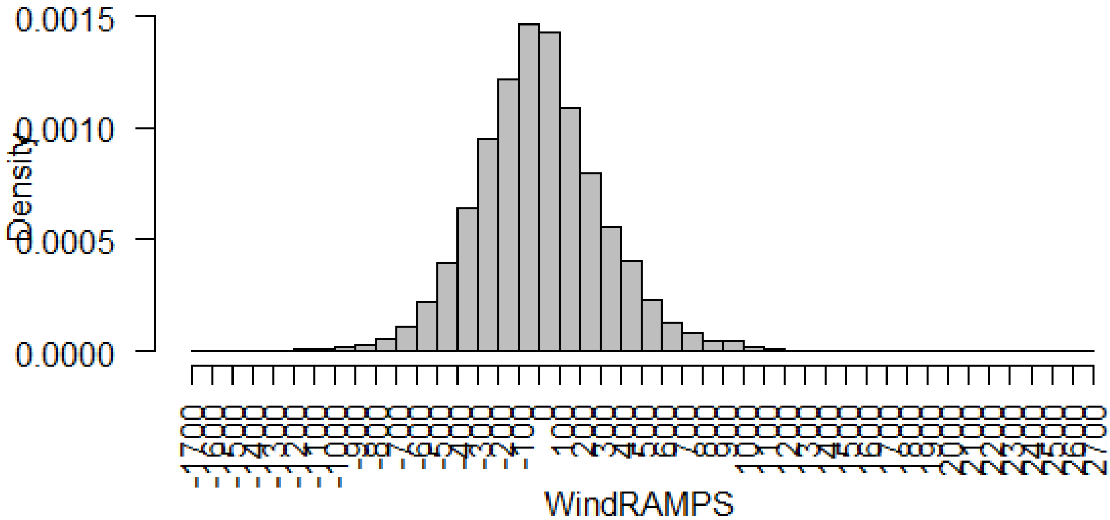


Figure 4.14: Density distribution histogram of windRAMPS variable from Jan 2015 to March 2016.

From the scatter plot in Annex B.1, no obvious relationships could be identified for this variable.

4.4.2.3 TEMPERATURE

It is commonly accepted and known that the temperature affects the demand load, which itself has been identified as a source of the system imbalance.

To obtain one hourly temperature variable for the Spain, temperatures from (10) different major geographical locations were considered:

1. Madrid
2. Santander
3. Barcelona
4. Valencia
5. Malaga
6. A Coruna
7. Badajoz
8. Valladolid
9. Zaragoza
10. Albacete

The average using each locations hourly temperature was used and calculated as follows:

$$TEMP_{h,t} = T_{h,t}^{mean} \quad (4.6)$$

where,

$$T_{h,t}^{mean} = \frac{1}{n} \sum_{i=1}^{n} T_{t,h,l} \quad (4.7)$$

$$t = temperature\ type, measured\ (m) or\ forecast\ (f)$$

$$T_h = hourly\ temperature, ℃$$

$$l = gerographical\ location$$

$$n = number\ of\ geographical\ locations$$

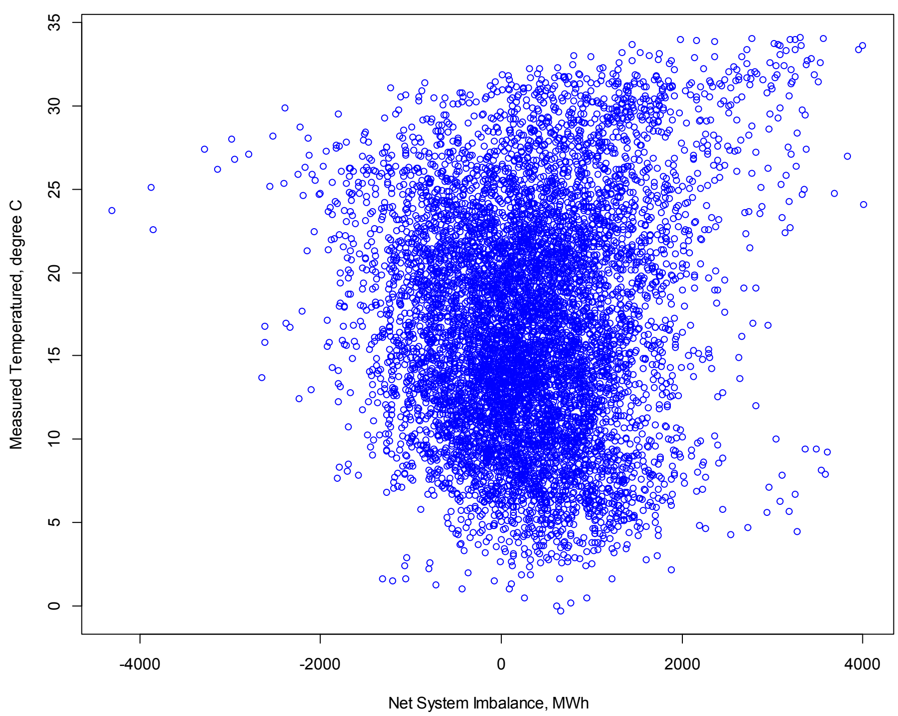


Figure 4.15: Scatter plot of measured temperature and net system imbalance.

4.4.2.4 *Technical Constraints: Energy and Price*

On any particular day, increased wind production can be associated with increased network constraints. For the technical constraints variable, the energy and price resulting from the *Phase II* technical constraints resolution process managed by the TSO was used. For ease of analysis, the magnitude (absolute energy value) was used.

From Figure 4.16 and Figure 4.17 it can be noted that the more extreme system imbalance events, especially for negative system imbalances, occurred when the technical constraints energy was around 500 MWh.

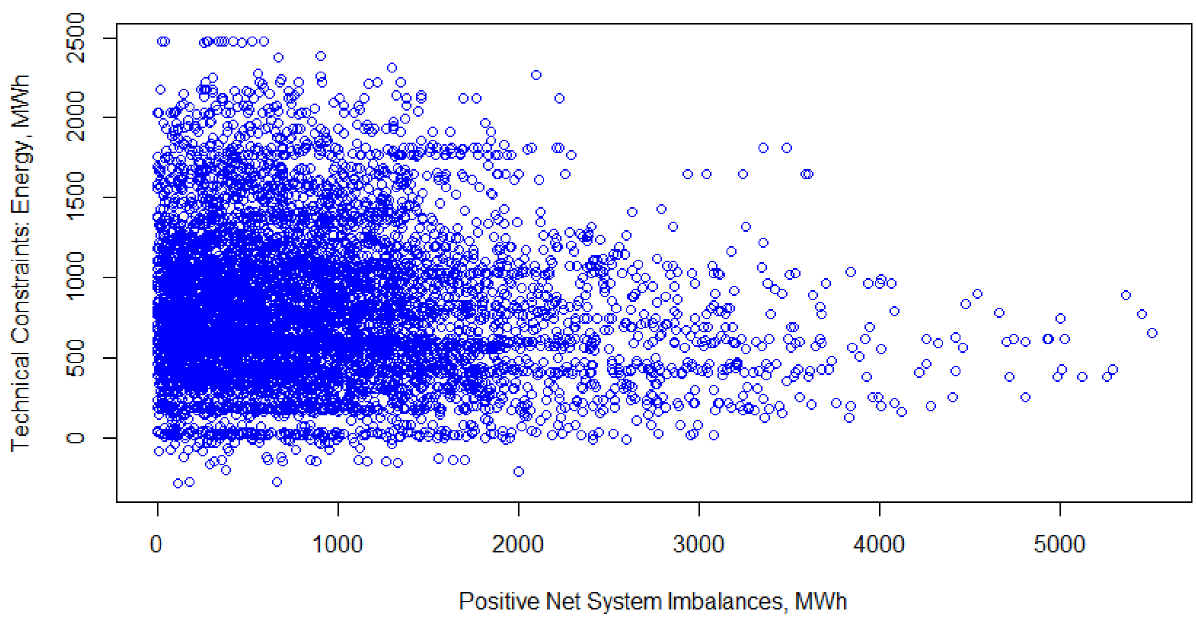


Figure 4.16: Scatter plot of energy network constraints and positive system imbalance volumes from Jan 2015 to March 2016.

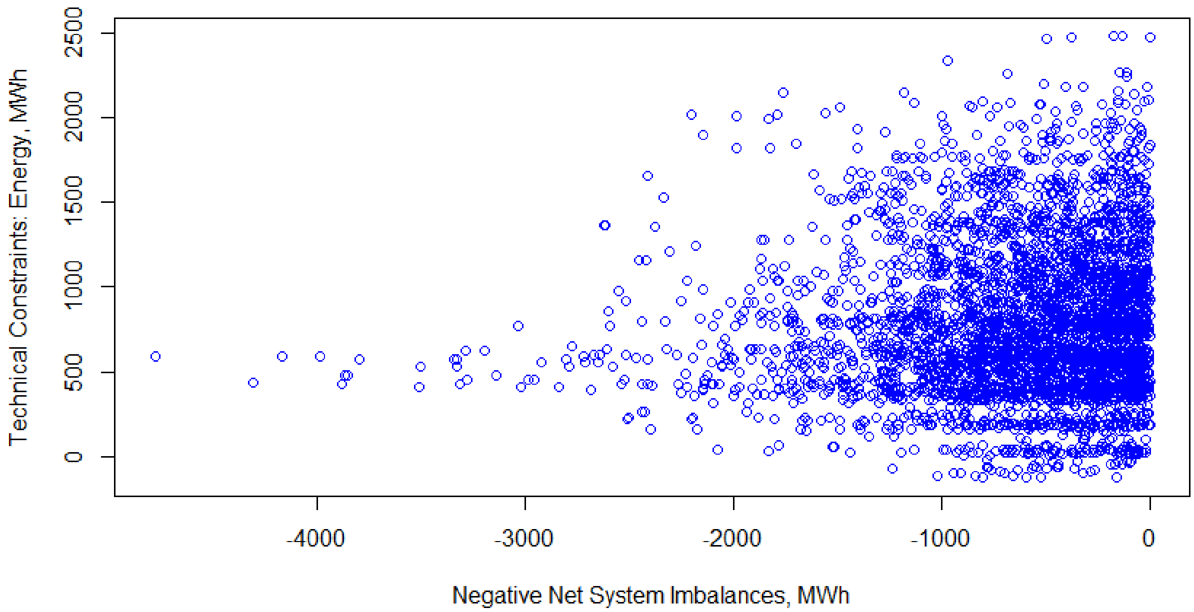


Figure 4.17: Scatter plot of network constraints and negative system imbalances.

4.4.2.5 Day-Ahead Market Price

The day-ahead market price is often an indicator of market conditions. It can provide insight to the volume of demand and supply matched, and to the types of technologies participating in the matching process through the economic merit order system, among others. As an example, high levels of production from RES-E in any given hour are associated with lower day-ahead market prices. The day-ahead market price is also the benchmark by which the *cost* (not necessarily price*)* of the imbalance is determined. As was explained in section 4.2.2, the imbalance price is related to the cost of the balancing energy procured by the TSO (secondary and tertiary regulation, and deviation management) to mitigate the system ´s imbalance.

DAMP, Imbalance Direction and Cost

From Figure 4.18, it can be observed that <u>the negative imbalance cost tends to increase as the day-ahead market price decreases</u>. This relationship is further accentuated if we consider the negative imbalance cost as a fraction of the day-ahead market price, as done in Figure 4.19, where an exponential increase of the negative imbalance cost can be observed as the day-ahead market price decreases.

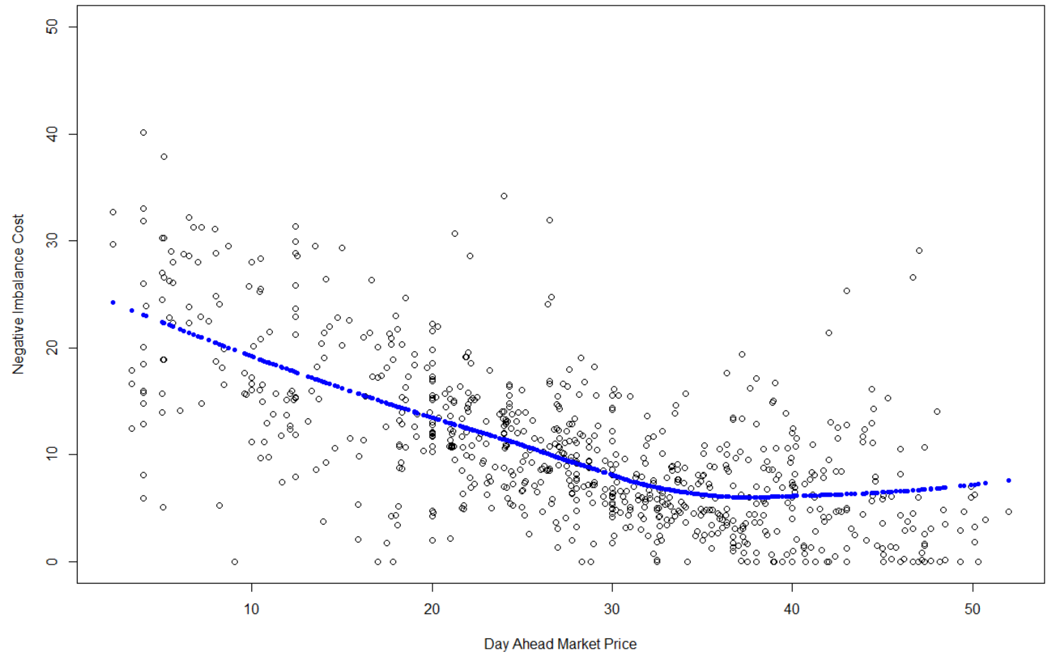


Figure 4.18: Scatter plot of the DAMP and the negative imbalance cost (when system imbalance is positive) for February to March, 2016.

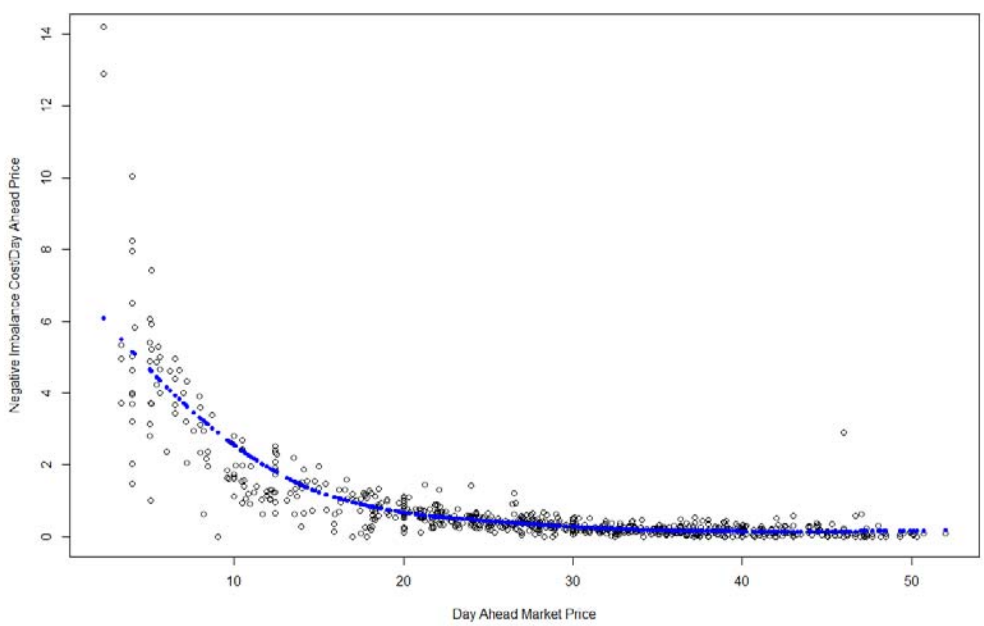


Figure 4.19: Scatter plot of the DAMP and the negative imbalance cost (when system imbalance is positive) as a fraction of the DAMP for February, 2016 to March, 2016.

In terms of the positive imbalance cost (when system imbalance is negative and portfolio imbalance is positive – actual production is greater than scheduled), no obvious trends could be discerned. As the day-ahead price acts as a “ceiling” or maximum of the possible positive imbalance cost, the apparent linear trend observed in Figure 4.20 just represents that ceiling.

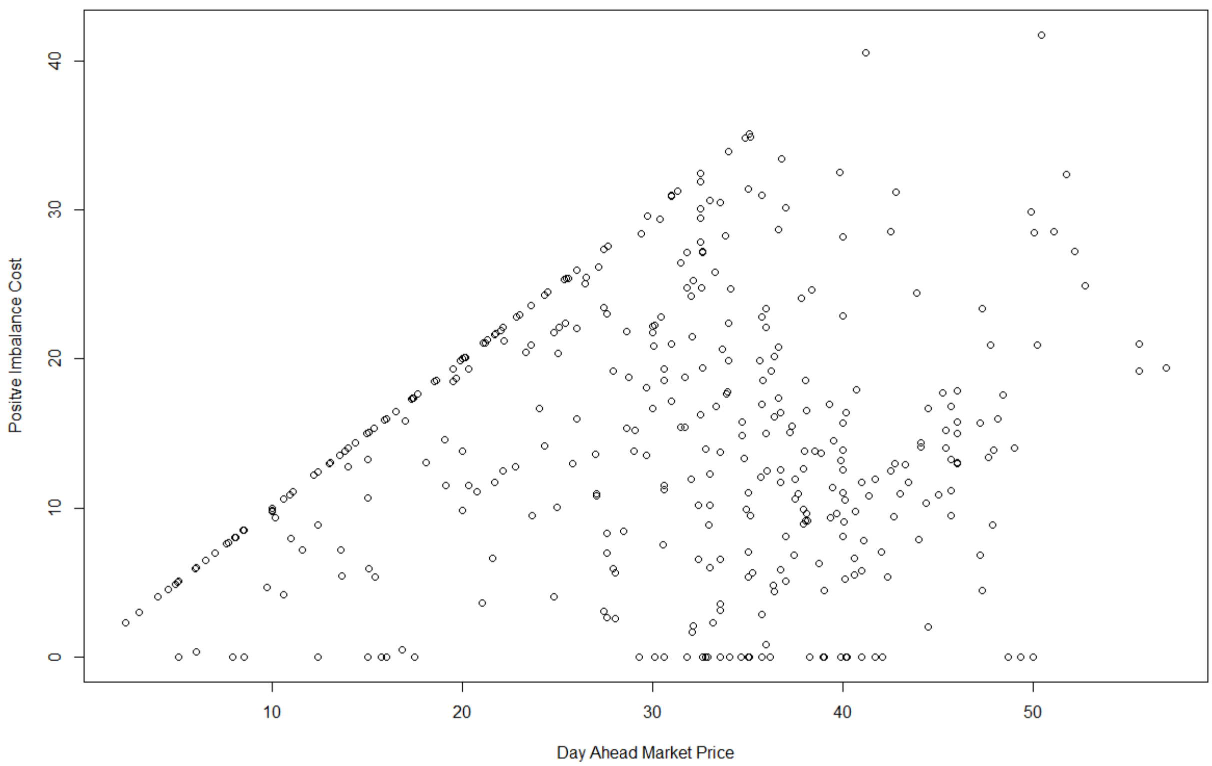


Figure 4.20: Scatter plot of positive imbalance cost (negative system imbalance) and DAMP for February, 2016 to March, 2016.

Of particular significance, is the <u>finding that the imbalance cost (both positive *and* negative), tended to be highest for the lower DAMPs</u>. Figure 4.21 shows how, for the month of January, 2016, the bottom quartile (25^{th} percentile) of DAMP – below 32.09€/MWh – not only had the largest number of observations with the highest imbalance costs – above 19.22€/MWh – but 1- the frequency (and likelihood) of these occurrences consistently and rapidly decreased as the DAMP increased (bin within dotted red line), and 2- within the lowest DAM prices (bottom quartile) the occurrence of imbalance costs greater than the median was over 85%, of which over 60% of the observations corresponded the highest imbalance costs experienced in that month (quartile with trend line).

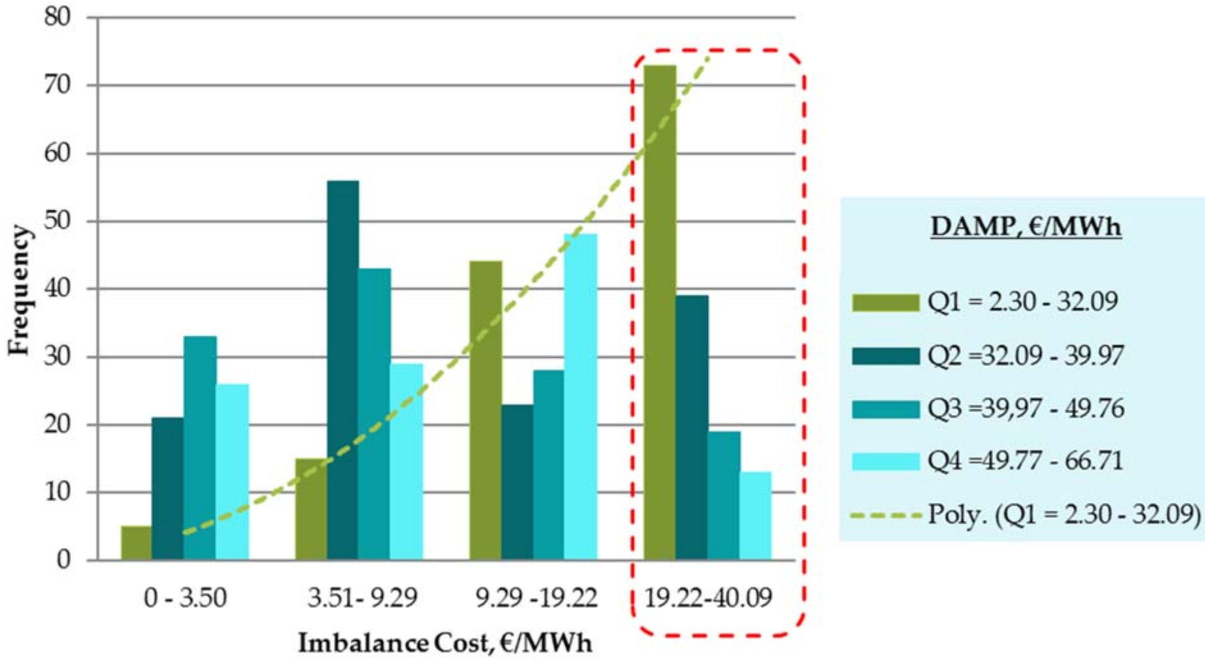


Figure 4.21: Quartile frequencies of Imbalance cost and DAMP for January, 2016

4.4.2.6 Additional Upwards Reserves (RPAS)

As described in Chapter 3, the additional upwards reserve is a market is called upon on an as-needed basis, explaining the zero values in Figure 4.22 for the RPAS variable. For only about 9.2% of the hours in the period studied was the additional upwards reserve called upon.

No obvious trends were identified for this variable.

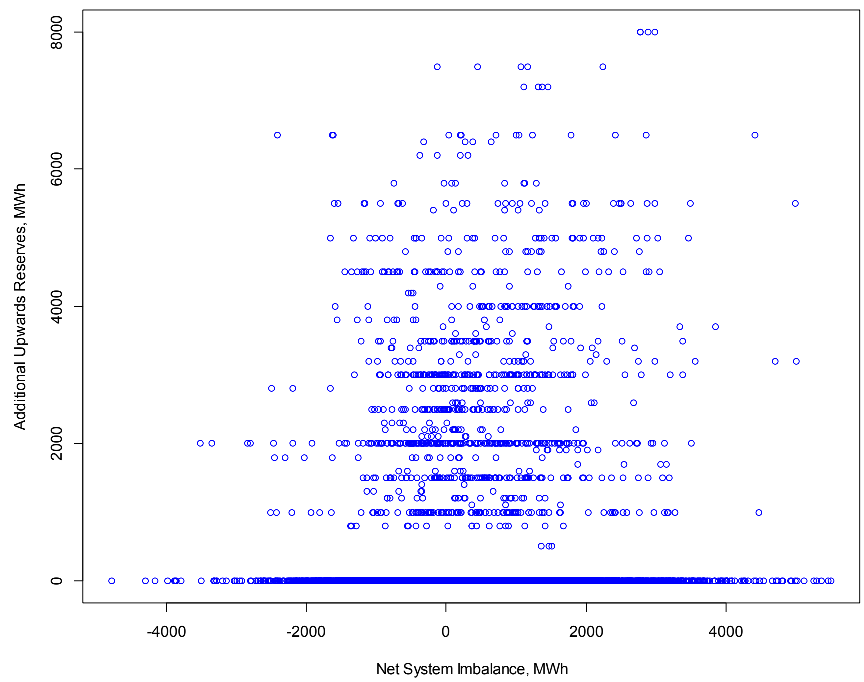


Figure 4.22: Scatter plot of the additional upwards reserve and the system imbalance for Jan 2015 to March 2016.

For use in the model, this variable will considered as binary: 1 for market called upon and 0 for absence of market.

4.4.2.7 Secondary Regulation Band

The secondary regulation band variable (MW), both downwards and upwards, are represented in Figure 4.23 and Figure 4.24 with the system imbalance, respectively. The band value seemed to be issued in steps of around 100 MW, and for the period studied, ranged from around 400 to 700 MW.

From those two figures, we can observe that the downwards band exhibits somewhat of a clearer trend in terms of the magnitude, and to a limited extent, to the direction of the imbalance. More specifically, the larger negative imbalances (e.g. greater than 2,000 MWh) seemed to occur most frequently when the downwards regulation was around 500 MW. In fact, 89% of the negative imbalances with a magnitude greater than 2000MWh occurred when the secondary regulation band was between 450 and 550 MW. Given the larger costs implications associated with negative system imbalances, this trend can be useful in implement a bidding strategy.

On the other hand, around the lowest downwards band of 400MW, the magnitude of imbalances generally remained under 2,000 MWh.

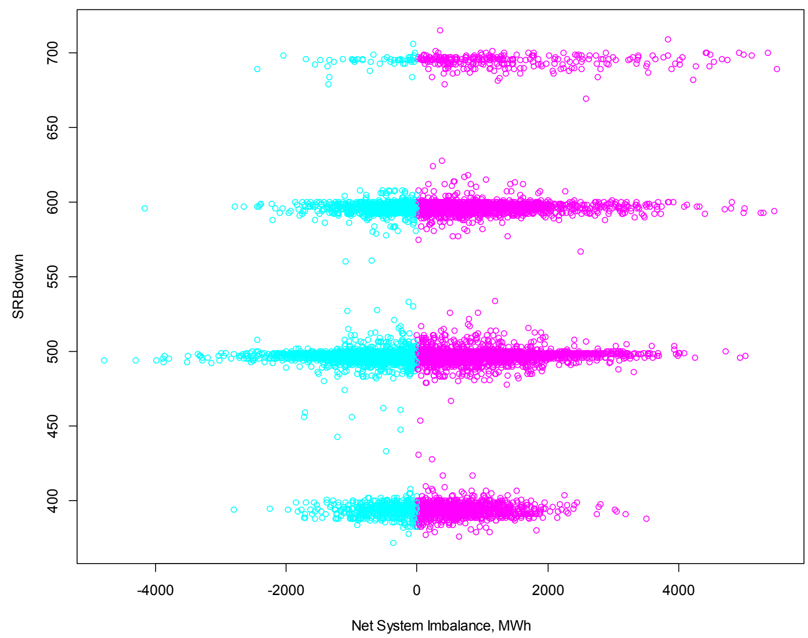


Figure 4.23: Scatter plot downwards secondary regulation band and system imbalance for Jan 2015 to March 2016.

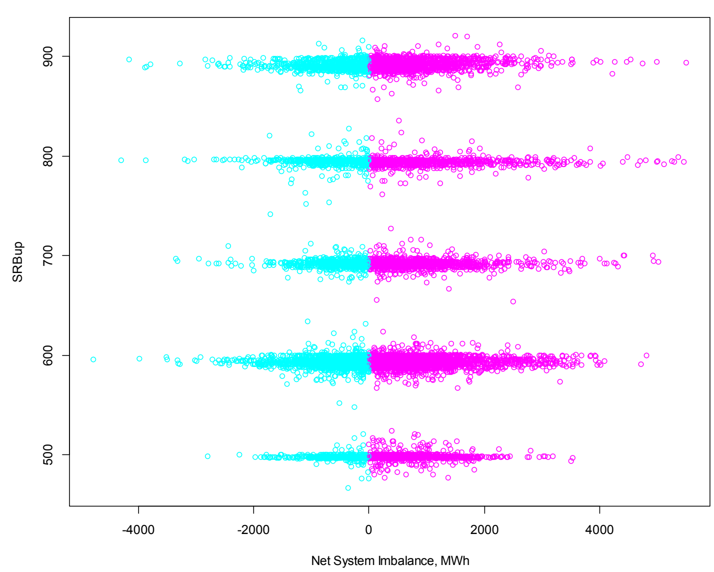


Figure 4.24: Scatter plot upwards secondary regulation band and system imbalance.

4.4.3 OBSERVATIONS

From the previous analysis, the main observation is as follows:

- With the exception of *percenDF*, very weak liner-relationships were encountered, if any.

However, certain trends and tendencies were observed. Most notably:

- **percenDF:** There appears to be a relationship between the system imbalance volume and the percentage difference of demand forecast met in the day-ahead market - *percenDF*. This variable had a significantly higher correlation than all other variables. Particularly noticeable is the trend of the system imbalance being positive when the level of demand met by the day-ahead market (PBF schedule) is lower than the forecasted demand.
- **Day-ahead market price and the imbalance cost:** It was observed that the imbalance cost (both positive *and* negative), tended to be highest for the lower DAMPs.
- **Technical constraints energy:**
- **Downwards secondary regulation band:** the larger negative imbalances (e.g. greater than 2,000 MWh) most frequently occured when the downwards regulation band was around 500 MW. For the period studied, this was equivalent to approximately 85% of those imbalances.

It is worth noting, however, that the lack of linear dependency does not negate other types of relationships that could improve the model´s performance. Therefore, the exploratory data analysis takes on more of an informational role.

CHAPTER 5:

MODEL DEVELOPMENT

This chapter discusses the main components of the model, namely development of the forecasting model and development of the optimized bidding strategy application, followed by a description of the validation and assessment methodology.

5.1 INTRODUCTION

As mentioned in the previous Chapter, the main components to developing the optimized bidding strategy can be broken down as follows:

1. Exploratory data analysis of variables;
2. *Development of forecasting model;*
3. *Development of optimized bidding strategy application tool;*
4. *Validation;*
5. *Assessment.*

The exploratory variable analysis was detailed in the previous Chapter. That analysis can also be used as a discriminatory step to identify the input for the *forecasting model*. This second part consists of a multivariate forecast of the system´s hourly imbalance volume using the random forest technique where non-linear relationships between past values of the selected predictor variables and the system´s imbalance volume can be represented. Several short-term time-frames for the training horizons will simulated to determine the best performing time-frame.

The third part consists of developing an *optimized bidding strategy application*. Using the forecasting model´s predictions as one of the input variables, a genetic algorithm optimization is applied to determine an hourly bid amount to adjust the agent´s schedule that minimizes the imbalance costs (maximizes the savings).

Validation of the model is performed by running simulations for each training horizon. Finally, *assessment* of the model´s performance is done to evaluate 1- the forecasting model and 2- application of the bidding strategy. A sensitivity analysis is also conducted.

5.2 DEVELOPMENT OF FORECASTING MODEL

The forecasting model was developed by applying the random forest technique described in Chapter 3. The main development stages of the model are described following sections.

5.2.1 FEATURE SELECTION

In machine learning, features refer to predictor variables, and we'll refer to them as such throughout this section. For feature selection it has been recognized that the combination of the *m* best features does not necessarily lead to better performance[1]. Moreover, in machine learning a variable that may appear of no use on its own may provide significant performance improvement when taken into account with others, even two different – and apparently useless – variables used together (Guyon & Elisseeff, 2003). Therefore all variables were included in the model. Different variable combinations were tested as part of the sensitivity analysis described later on.

5.2.2 DATA COLLECTION AND PREPARATION

Handling of large data sets often requires significant efforts related to acquiring and preparing the data. For example, a one-year simulation using all the variables identified in the preceding section (in addition to target variables) - could entail handling a data-set of well over 130,000 data points. Such volume of data requires a systematic approach were processes can be automated as much as possible. Some of the most critical elements of this process are described below.

First, a database, if not already available, has to be created to store the feature data necessary to conduct the forecasting portion. All external (i.e REE and OMIE) and internal (specific to the Energy Trader in the case-study) data was stored in a central SQL server database local to the Energy Trader. Obtaining and aggregating local temperatures into national values, for both average forecast and measured values, is a

[1] Peng H; Long F; Ding C: "Feature selection based on mutual information: criteria of max-dependency max-relevance, and min-redundancy", IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) 1226-1238, 2005.

lengthier process that requires processing data for specific geographical locations from among a database containing dozens of different locations.

Next, formatting of data, especially for the handling of dates and times, was imperative to creating a time-series based input dataset digestible in the programming language of choice. For this task, the lubridate package available for R programming was used.

The data were merged into a single set with all the variables and respective dates/hours corresponding to the simulation period. In order to run any simulation-type validation, as was performed in this project and described later on in this section, all missing values for the time-range in question had to be identified. For the sake of purity, no substitution - by imputation or otherwise - was performed (which greatly reduced the number of predictions made).

Finally, as the magnitude of the features selected differs widely, a certain amount of data preparation is needed to normalize the data. Pre and post processing of the data is necessary to:

i. normalize input to the random forest algorithm for training,; and
ii. transform prediction output to original data range.

5.2.3 Model Parameters

The random forest model was fit with 1500 regression trees using the random forest package available for the R programming environment. The package´s standard default parameters were also applied (i.e. default value for the number of variables randomly sampled as candidates at each split $\lfloor p/3 \rfloor$ and the minimum size of terminal nodes [5] (Liaw & Wiener, 2015)).

- Number of trees = 1500
- Default settings of R random forest library will be used.

5.2.4 Training and Testing

Training

From observing the day-to-day evolution of the system´s imbalance volume it is apparent that its behavior is more likely related with very short term indicators, as in what happened yesterday or the last couple of days, as opposed to what happened several weeks or even a month ago. Therefore very short to short-term training time-frame horizons were used to train the model. Eight time-frames, ranging from 1 to 8 days, were simulated as a separate training horizon case to determine which of these parameters yields better model performance. The training horizons correspond to the time-range represented in the training data set. The eight simulations are illustrated in Figure 5.1, and formulated below.

$$\textit{Training Horizon} = D - H \tag{5.1}$$

where,

$H = 1{:}8$ *days***Error! Not a valid link.**

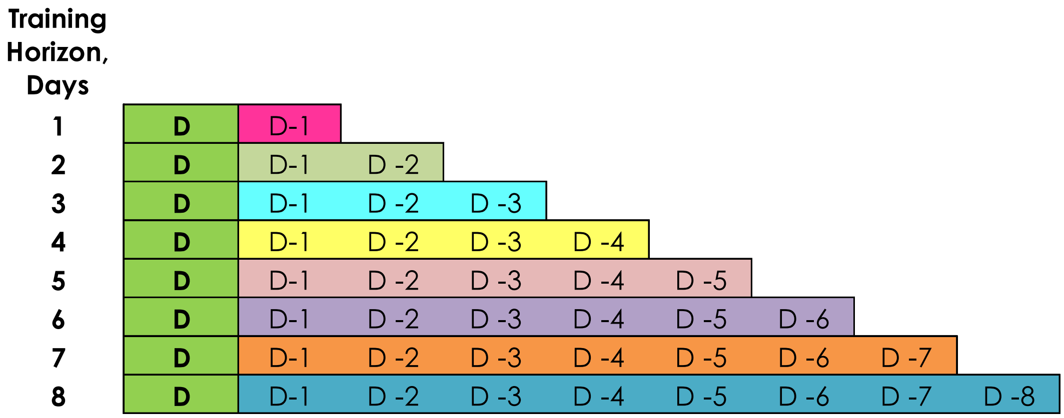


Figure 5.1: Illustration of training horizon time-frames for post-gate closure model (pre-gate closure

Test/Forecasting

For the post-gate closure forecasts, the test data set consisted of the predictor variables for target day D, whereas these test data set for the pre-day ahead gate closure forecast consisted of the predictor variables for D-1.

5.3 Development of Bidding Strategy Application

The insight gained from obtaining a forecast of the system's imbalance must be translated into useful bidding activity by determining the optimal hourly bid that will minimize the imbalance costs for the producer/trader.

To this end, a bidding strategy was defined through a function containing a set of parameters that express bidding conditions and whose optimal values simultaneously minimize the imbalance cost based on forecasts and outcomes for the preceding week.

In order to implement this strategy, a genetic algorithm (GA) is applied to obtain a solution set for the strategic parameters. The GA is based on an optimization function (fitness function) that minimizes the imbalance cost based on hindsight consideration (7 days) of: a) the forecasting model´s past predictions, and b) the actual observed variables (operational and market) of the system´s imbalance corresponding to those predictions.

Along with an objective function minimizing the cost, the parameters describing the bidding strategy conditions are incorporated as conditional constraints into the GAs fitness function. By doing so, the genetic encoding and the basic genetic operators will include feasible members of the population.

Upon obtaining a solution set for the parameters, these are substituted into the strategy function, which is subsequently applied to the target day variables, and an hourly bid amount is obtained.

5.3.1 Forecast Certainty in Bidding Strategy

The fundamental idea behind the bidding strategy is to explore links between indicators, the model´s forecasted values, and actual outcomes in order to identify circumstances of increased (or decreased) certitude in the prospective outcome and benefits from the bidding activity.

With that in mind, a preliminary simulation of the forecasting model, including simulations for the different training horizons, was performed to gain insight on how a potential bidding strategy could be based on certitude of the model´s accuracy. An analysis of the model´s output, as it relates to the actual observed outcomes, was conducted on the output data from the model with a training horizon of 5 days[2].

PorcenDF

The percent difference of the demand forecast matched in the day-ahead market (*porcenDF*) appeared to be one of the indicators providing the most promising insight to the direction of the system´s imbalance volume. As can be observed from Figure 4.8 and Figure 4.9 in the previous Chapter, the majority of positive system imbalances appear to

[2] From preliminary runs, a training horizon of 5 days provided the best performance. The final results are presented later on.

happen when *porcenDF* is above 2% or higher, and conversely, most negative system imbalances appear to occur when the *porcenDF* is below that level.

If we set a 2% threshold on the *porcenDF* variable to group the data in the simulation, we obtain the results shown in Table 5.1. We observed that 65% of the system imbalance volumes were positive when 98% or less of the forecasted demand (porcenDF>=2%) is met by the PBF schedule. For over 71% of those occasions, the model accurately forecasted the direction of the imbalance if the magnitude of the forecasted volume was over 200 MWh.

Output Information	No. Obs.	Percent
Total Data	5685	
Demand PBF $\leq 0.98 \times$ Demand Forecast	3135	55.14%
$\Delta Psys_h > 0$; positive	3678	64.69%
$\Delta Psys_h < 0$; negative	2007	35.53%
Model predicts sign	3862	67.93%
Model output magnitude > 200 MWh	4325	
Model predicts sign for output magnitude > 200 MWh	3098	71.63%
Demand PBF $\geq$ Demand Forecast	969	17.04%

Table 5.1: Model output results, from April, 2015 to March ,2016.

If we only consider the hours were demand forecast met by the PBF schedule is less than or equal to 98%, we get the results in Table 5.2 below. In this case, the model accuracy in predicting the direction is just over 72%, increasing to almost 75% for instances were the forecasted volume was of over 200 MWh. Although not indicated in the table, this accuracy increases to over 80% for a magnitude of 300 MWh, but decreases to around 71% for magnitudes of 150 MWh.

Output Information	No. Obs.	Percent
Demand PBF ≤ 0.98 × Demand Forecast	3135	
$\Delta Psys_h > 0$; positive	2292	73.11%
$\Delta Psys_h < 0$; negative	843	26.89%
Model predicts sign	2266	72.28%
Model output magnitude > 200 MWh	2642	
Model predicts sign for output magnitude > 200 MWh	1978	74.86%

Table 5.2: Model output results for hours were the demand forecast met by the PBF schedule is less than or equal to 98%.

The table below shows what happens with the model´s output when the demand forecast met by the PBF schedule exceeds the forecasted demand (*percenDF<0)*. For this criteria the occurrence of negative imbalances increases from to over 55% from the 35.56% under Table 5.1. The model accuracy on predicting the imbalance sign decreases by almost 7%.

Output Information	No. Obs.	Percent
Demand PBF ≥ 0.98 × Demand Forecast	969	
$\Delta Psys_h > 0$; positive	428	44.16%
$\Delta Psys_h < 0$; negative	541	55.86%
Model predicts sign	594	61.30%
Model output magnitude > 200 MWh	773	
Model predicts sign for output magnitude > 200 MWh	495	64.03%

Table 5.3: Model output results for hours were the demand forecast met by the PBF schedule is greater than 98%.

The results of this analysis is the basis of the bidding strategy.

5.3.2 Bid Volume Strategy Function

The bidding strategy is based on setting a bid volume whose magnitude increases linearly as the forecast imbalance magnitude also increases. In other words, a large system imbalance may handle a larger bid value without flipping the imbalance´s direction, while greater certainty on the imbalance direction is also represented. Considering the analysis made in the previous section, among other factors, a set of parameters that describe bidding conditions bounding the linear function, as is shown in Figure 5.2, are also incorporated to manage risks, express risk-appetite and portfolio needs, and increase certainty.

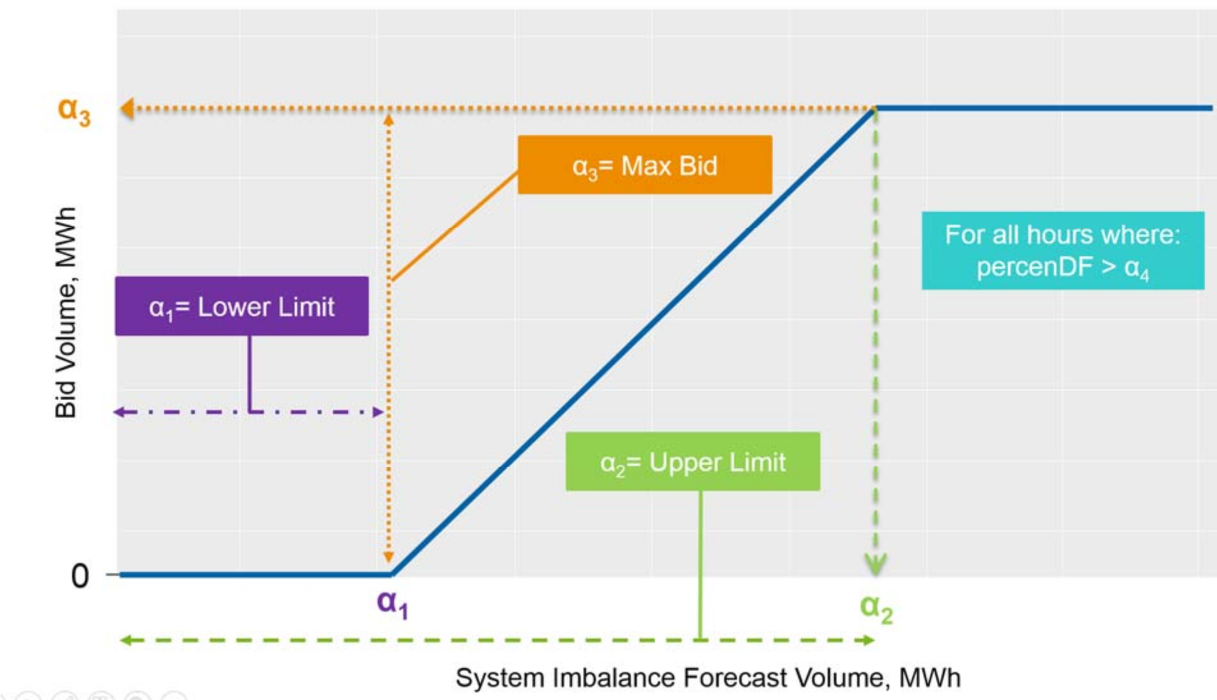


Figure 5.2: Graphic representation of bidding strategy for all hours where PercenDF >= α_4.

Any bid activity must meet the following condition, otherwise bid volume is null:

$$(\boldsymbol{PercenDF_h} \geq \boldsymbol{\alpha_4}) \tag{5.2}$$

where,

α_4 = *optimal* percent difference of the demand forecast matched in the PBF schedule

For all hours meeting condition (($\boldsymbol{PercenDF_h} \geq \boldsymbol{\alpha_4}$)
(5.2), the bid volume is determined as follows:

$$(\,\Delta \boldsymbol{P_h sys} \geq \boldsymbol{\alpha_1}) \;\; \rightarrow \; \boldsymbol{BV_h} \;=\; \boldsymbol{\alpha_3} \tag{5.3}$$

$$(\boldsymbol{\alpha_2} < \Delta \boldsymbol{P_h sys} < \boldsymbol{\alpha_1}) \;\; \rightarrow \boldsymbol{BV_h} \;=\; \boldsymbol{\alpha_3}\, .\, \frac{(|\Delta \boldsymbol{P_h sys}| - \boldsymbol{\alpha_2})}{\boldsymbol{\alpha_2} - \boldsymbol{\alpha_1}} \tag{5.4}$$

else,

$$\boldsymbol{BV_h = 0} \tag{5.5}$$

where,

BV_h = magnitude of hourly bid volume.

α_1 = an upper limit on the magnitude of the system imbalance forecast volume.

α_2 = a lower limit on the magnitude of the system imbalance forecast volume.

α_3 = maximum magnitude of the bid volume.

5.3.2.1 OPTIMAL PARAMETERS

To key about the strategic parameters is that their *optimal* values simultaneously minimize the imbalance cost. To obtain a solution set for these strategic parameters, a genetic algorithm (GA) is applied, as described in section 94

$\boldsymbol{\alpha_1}$ **,** $\boldsymbol{\alpha_2}$ **:** These are the *optimal* upper and lower bounds that represent the level of comfort with the magnitude of the imbalance volume forecast.

1. A forecast volume may be uncharacteristically large and thus unlikely. This can be an issue considering that the bid value increases linearly with the forecast value, so an upper bound is established to mitigate the risk of error with large bid values which can carry significant losses.

2. On the other hand, low forecast values also pose increased risk for being too close to the opposite direction. This may represent a higher potential for the actual outcome to be of the opposite sign. This risk will be mitigated by placing a lower bound under which no bid will be placed.

$\boldsymbol{\alpha_3}$**:** This parameter is the *optimal* maximum bid value and it represents the risk appetite of the market agent and the properties of its portfolio´s imbalance.

$\boldsymbol{\alpha_4}$**:** Based on the analysis conducted in the preceding section, certainty on the outcome may be increased by considering the *optimal* percent difference of the demand forecast matched in the PBF schedule, and requiring that all bid activity correspond to hours where its variables meet or exceed this parameter.

The bidding strategy was developed from the perspective of an energy trader assuming its production portfolio´s hourly imbalance is always against/opposite the system´s needs. This is a reasonable assumption considering that an imbalance that supports the system´s needs is not penalized. However, it can also be applied from the perspective of a retailer by just changing the sign convention discussed in the next section, and of course, considering the specific agents own portfolio imbalance to determine the maximum bid.

5.3.3 Genetic Algorithm Optimization

To key about the strategic parameters is that their *optimal* values simultaneously minimize the imbalance cost. To obtain a solution set for these strategic parameters, a genetic algorithm (GA) is applied.

The GA is based on an optimization function (the fitness function) that determines the best solution set for the strategic parameters to be used for target day D, that would minimize[3] the imbalance costs for the preceding week (7 days, from D-7 to D-1), using as input variables listed in

Variable	Name	Equation or source
$\widehat{\Delta P_h} sys$	Forecast: system´s net imbalance volume, MWh	Source: Forecasting model output
$\Delta P_h sys$	Observed: system´s net imbalance volume, MWh	Source: REE
percenDFh	Percent difference of the demand forecast matched in the day-ahead market.	See Table 4.4
$COST\Delta P_h$	Imbalance cost, €/MWh	$COST\Delta P_h = PRICE\Delta P_h^d - PRICE\Delta P_h^u$ (4.4)

[33] It is worth noting that the GA package for R programming only offers a maximization feature option, thus the trivial transformation to maximization was performed when implementing the optimization.

Table 5.4: Input variables to determine optimal parameters.

The GA package available for the R programming environment was used with 150 iterations.

5.3.3.1 Fitness Function

In GAs, an explicit objective function is not necessary. However, the specification of an appropriate fitness function is crucial. The fitness function is a black box for the GA, which in our case is achieved by specifying a mathematical function that incorporates:

- An objective function.
- Conditional constraint parameters.

A. The objective function minimizing the cost of the imbalances, from the perspective of a producer/trader, is represented below.

$$min\ C_w = \sum_{h=1}^{N} (BV_h \times COST\Delta P_h \times b_h)$$

(5.6)

where,

b_h = binary variable that denotes cost (1) or savings (-1), based on accurate prediction of the imbalance direction. The sign of variable is equivalent to the sign of - ($\widehat{\Delta P}_h sys \times \Delta P_h sys$).

$\widehat{\Delta P}_h sys$ = forecast of system imbalance.

$COST\Delta P_h$= see Eqn 4.4.

N = number of hours in W.

W = Week from $\mathrm{D} - 7$ to $\mathrm{D} - 1$, the optimization period.

B. Parameter Optimization. GAs have a flexible constraint handling method. In our case the fitness function is built to search for a set of parameters that we have defined to describe the behavior of the bid activity, and whose solution minimizes the objective function above. In other words, the bid activity behavior defined through these parameters is the bidding strategy.

$$(\alpha_1^D \leq \widehat{\Delta P}_h sys) \ \ \& \ (PercenDF \geq \ \alpha_4^D) \rightarrow \ BV_h \ = \ \alpha_3^D \tag{5.7}$$

$$(\alpha_2^D < \widehat{\Delta P}_h sys < \alpha_1^D) \ \ \& \ (PercenDF \geq \alpha_4^D) \rightarrow BV_h \ = \ \alpha_3 \ . \frac{(|\widehat{\Delta P}_h sys| - \alpha_2^D)}{\alpha_2^D - \alpha_1^D} \tag{5.8}$$

$$\text{else,} \quad BV_h^D \ = \ 0 \tag{5.9}$$

5.3.3.2 *OPTIMAL PARAMETER RANGES*

Based on the data summarized in Table 5.5, corresponding to the simulation period range of April 1, 2015 – March 2016, the pre-determined value ranges for the strategic parameters were selected, as are shown in Table 5.6. The portfolio values are based on the Energy Trader´s own data.

Table 5.6 specifies the range of values for the strategic parameters. These ranges were determined based on historical data, and portfolio imbalance data specific to the energy trader in the case. For a summary of the data considered, see the table below.

Net Imbalance Volume, MWh	**Min**	**1st Qu**	**Median**	**Mean**	**abs(Mean)**	**3rd Qu**	**Max**
System: Observed	-3283	-235.2	302.5	336.4	739.05	872.2	4011
System: Forecast	-2545	-66.51	308.6	338	553.66	708.5	3172
Portfolio	-819	-120.8	-20.36	-27.35	120.06	70.44	696.1

Table 5.5: Data summary, No. Obs = 5685 period range= April 1, 2015 – March 31, 2016.

Strategic Parameter		**RANGE**	
		Min	**Max**
lower limit	α_1	50	400
upper limit	α_2	550	1250
Max Bid	α_3	10	150
Percen	α_4	input min	input max

Table 5.6: Value ranges for strategic parameters.

5.3.1 APPLICATION OF BIDDING VOLUME STRATEGY FUNCTION

Upon obtaining a solution set for the parameters, these are substituted into the strategy function which is applied to the variables corresponding target day D. Each parameter will have the same value for the 24 hours of day D. Subsequently, and hourly bid is obtained for each hour of the day.

Bid Value sign convention

Market agent transactions to adjust production program will be conducted from the perspective of a seller, therefore the sign convention depicted Table 5.7 is applied, conditioned to the sign (direction) of the system imbalance forecast in the last column.

Bid Sign (BS)	Transaction	Effect on Portfolio	System Imbalance Forecast Sign Condition
+	Sell (S)	Increase program schedule	-
-	Purchase (P)	Decrease program schedule	+

Table 5.7: Bid value sign convention for market agent transacting as seller.

The hourly bid value to be submitted, based on the type of transaction, is as follows:

$$\boldsymbol{B_h^D = BV_h^D \times BS} \tag{5.10}$$

where,

B_h^D =Final hourly bid value.

BS = Binary variable that determines the type of transaction, S/sell (1) or P/purchase (-1).

5.4 MODEL VALIDATION

5.4.1 SIMULATIONS

To validate the bidding strategy a series of simulations were performed for both components of the model and hourly bids were estimated for a 12-month period from April 2015 to March 2016.

(i) **Forecasting component.**

The model combinations to be simulated were first and foremost selected based on the availability of information to transact in the intraday markets. Therefore two main models are considered based on the forecasting horizon, and a third sub-set related to sensitivity analysis:

I. pre-day (ahead market) gate closure (Model 5),
II. post -day ahead market gate closure:
    a. *Pre- gate closure of first-intraday market session.* For transacting in *first intraday* market session without secondary regulation band information, as it is not yet available. (Model 3)
    b. *Pre- gate closure of second-intraday market session,* for transacting in *second intraday market session* with secondary regulation band information available by then. (Model 2)
    c. As the wind RAMP variable was a later addition to the forecasting model, a second set of simulations was run to for the post-day ahead models that incorporated the variable (Model 7 to transact in first intraday session, and Model 6 to transact in second intraday session).Other model variations were run as part of the sensitivity analysis
    d. Simulations for other model variations were also run as part of the *sensitivity analysis* described in section 5.4.4.

Each training horizon case, from 1 to 8 days (8 per model type) were also simulated.

(ii) Bidding strategy application component.

a. The bidding strategy was applied to all forecasting model typed and their respective training horizon cases listed above.
b. A second bidding strategy was applied after review of the results, which is discussed in the following chapter.

5.4.2 DATA

As specified earlier, all data is stored centrally at the Energy Trader´s local database.

The input variable data for the forecasting model corresponding to the simulation period was merged into a single data-set after the required processing was performed (see Section 4.3). However, due to missing data points, - particularly full days without temperature forecasts - the actual input data set was reduced to 7782 obs for the 14 variables listed in Table 4.4. As training horizons of up to eight days were applied, the number of model output points was further reduced, and more so in order to conduct an equivalent pre and post day ahead market gate closure comparison.

The input variables for the bidding strategy application are listed in Table 5.4.

5.4.3 PERFORMANCE ASSESSMENT WITH REAL-DATA.

To assess the performance of the model, a set of metrics were used for each model component.

(i) Forecasting model results: RMSE, MAE

The main metric used to evaluate the forecasting model´s predictions is the root mean square error (RMSE). This error measure gives more weight to larger residuals than smaller ones.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n}} \tag{5.11}$$

where,

$\widehat{y_i}$ = predicted value
y_i = observed value for the ith observation
n = number of observations

The mean absolute error (MAE) will be used a secondary metric. Although the MAE gives equal weight to the residuals, it is relatively easier to interpret.

$$\boldsymbol{MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}|} \tag{5.12}$$

(ii) Bidding Strategy: Savings (€, €/MWh)

As noted by Garcia et al (Garcia & Kirschen, 2006), “the true measure of improvement when forecasting market imbalance volumes is not an abstract error index but rather the savings in balancing costs that this improvement makes possible”, therefore the bottomline net savings will be the evaluation metric for the whole strategy.

<u>Savings, €</u>

The main global metric used to evaluate the bidding strategy as a whole is the cost reduction or savings that it achieves. The bidding strategy was applied to real data from the Spanish Energy Trader and its portfolio imbalance volumes were used to estimate the cost reduction/savings.

The energy trader´s portfolio imbalance data required additional processing to remove the effect of previous strategies implemented during the simulation period which affected the portfolio´s imbalance volume.

Determining if a cost reduction resulted from applying the strategy to any given hour depends on whether: a) the bid activity reduces or increases the portfolio´s (pre-strategy) imbalance; and b) the portfolio´s imbalance (pre-strategy) is supporting or opposing the system. Unless noted otherwise, the portfolio imbalance will always refer to the *pre-strategy* volume.

a) The bid value with respect to the portfolio´s imbalance (BP):

$$\boldsymbol{BP_{h} = SIGN(B_h \times \Delta Pport_h)}$$

BP_h^+ = Bid activity reduces portfolio imbalances. ($B_h, \Delta Pport_h \rightarrow$ the same sign)

BP_h^- = Bid activity increases system imbalance. ($B_h, \Delta Pport_h \rightarrow$ opposite sign)

b) The portfolio imbalance with respect to the system´s imbalance:

$$\boldsymbol{PS_{h} = SIGN(\Delta Pport_h \times \Delta Psys_h)}$$

PS_h^+ = pre-strategy portfolio imbalance supports system´s balancing needs.

PS_h^- = pre-strategy portfolio imbalance opposes system´s balancing needs.

Four possible savings/cost scenarios arise from applying the bidding strategy having different effects on savings, as depicted in Figure 5.3 (green arrows represent beneficial relationship for the strategy).

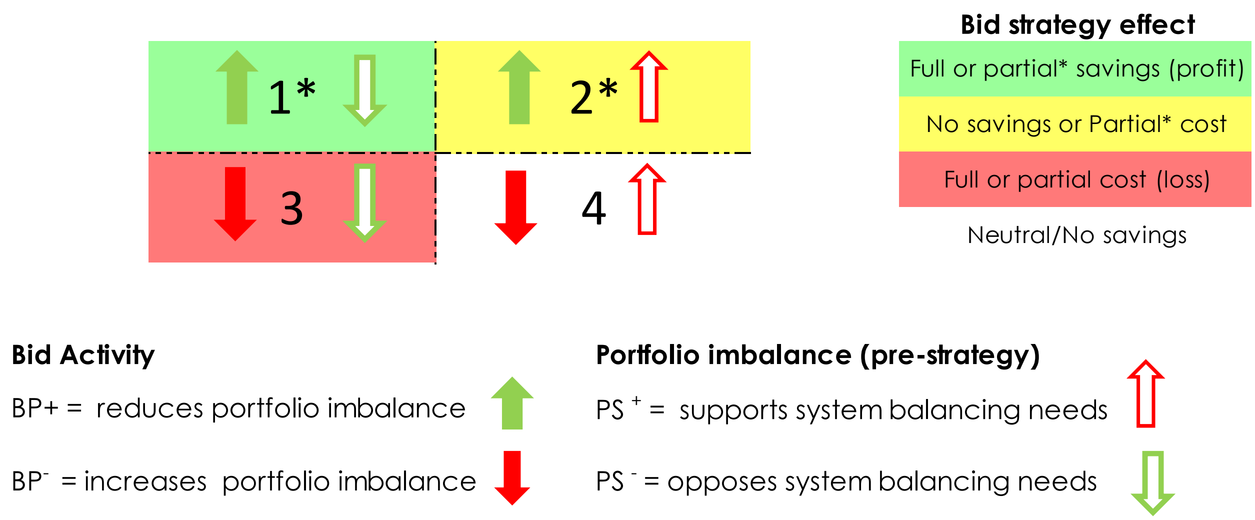


Figure 5.3: Cost-reduction scenarios for outcome of bidding strategy.

Hourly savings/cost are calculated as follows:

$$|b_h| < |\Delta P_h \boldsymbol{port}| \rightarrow PNL_{\boldsymbol{h.S}} = B_h \times COST\Delta P_h \times \boldsymbol{CR} \qquad (5.13)$$

where,

$PNL_{h,s}$ = hourly bidding strategy profit or loss/savings or cost for each cost-reduction scenario.

S = Cost reduction scenario (1-4)

CR = binary variable representing the cost-reduction scenarios 1-4;
scenario 1 = 1; sensation 2= 0; scenario 3 = -1; scenario 4 = 0

The partial cost or savings in scenarios 1 and 2 are due to the bid value flipping the imbalance direction of the portfolio, so either partial savings are realized, or a partial cost is incurred as represented in figure (5.6)

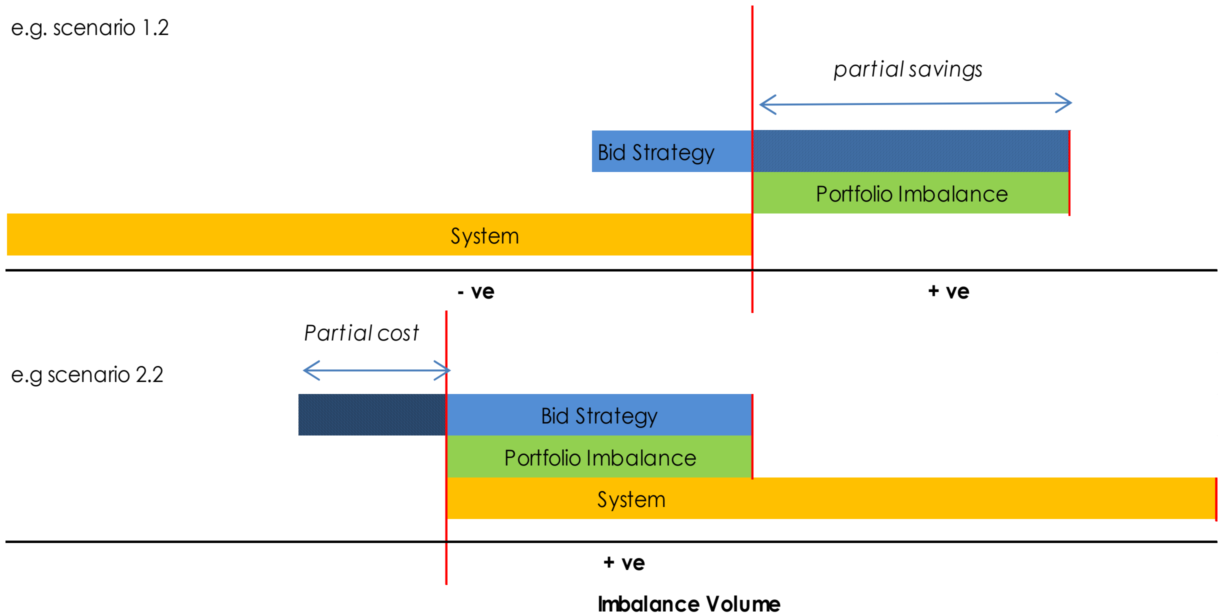


Figure 5.4: Partial cost or savings scenarios.

$$|b_h| > |\Delta P_h port| \rightarrow PNL_{h.S} = B_{h,s} \times COST\Delta P_{h,s} \times CRP_s \qquad (5.14)$$

where,

CRP = binary variable representing the partial cost-reduction scenarios 1 and 2; Scenario 1 = 1, Scenario 2 = -1;

The total savings is estimated with $PNL^{Total} = \sum_{S=1,}^{N_S} \sum_{h=1}^{N_h} PNL_{S,h}$ (5.15).

$$PNL^{Total} = \sum_{S=1,}^{N_S} \sum_{h=1}^{N_h} PNL_{S,h} \qquad (5.15)$$

where

PNL^{Total} = Total profit/losses or savings/costs for applying bidding strategy, €.

Savings, €/MWh

The savings per unit of energy adjustment will be used as a secondary metric.

$$PNL^{PerUnit} = PNL / \sum_{h=1}^{N} |B_h| \qquad (5.16)$$

where,

PNL^{PeUnit} = per unit savings of applying bidding strategy, €/MWh

5.4.4 Sensitivity Analysis

As part of the assessment process, a sensitivity analysis was performed to evaluate the relative importance of the input variables on the accuracy of the forecast and performance of the bidding strategy application. This evaluation was based on the effect of omitting a predictor/s from forecast model.

Predictors were omitted based on the variable importance information (%IncMSE) calculated by the random forest package for R programming. According to (Liaw & Wiener, 2015), the %IncMSE variable importance measure is computed from permuting out-of-the bag (OOB): "for each tree, the error prediction for the out-of-bag portion of the data is recorded (MSE for regression). Then the same is done after permuting each predictor variable. The difference between the two are then averaged over all tress and normalized by the standard deviation of the differences. If the standard deviation of the differences is equal to 0 for a variable, the division is not done (the average is almost always equal to 0 in that case)".

The models were ranked based on the performance metrics described in the previous section.

CHAPTER 6: RESULTS AND OBSERVATIONS

A total of 104 simulations were run, corresponding to 8 training horizon cases for 13 different models. Upon review of the first set of results, another 104 simulations were run with a modification to the bidding strategy application.

This chapter presents the results of these simulation, starting with the variable importance measure to select model variations, followed by a description of the model variations, continued by the results of the forecasting model and the bidding strategy applications (both original and modified), including a brief overview of the effects of the intra-day market prices on the strategy, ending with observations.

6.1 VARIABLE IMPORTANCE

The variable importance measure (%IncMSE) was saved from running the post-day ahead model with all the variables. The measure was obtained for each variable for each forecasted value and aggregated to represent the full simulation period. A summary of the results of this aggregation are presented in Table 6.1 for a training horizon of 4 days (H=4). A plot of the mean %IncMSE ranking the variables in order of importance is shown in.

The variables ranked in the top 10 all had a %incMSE between 15% to just over 20%, with three more or less distinct groups that could be identified, as shown in Figure 6.1 which contains a plot of the mean %IncMSE ranking of the variables in order of importance. Group 1 variables are primarily related to meteorological conditions – wind and temperature-, group 2 variables are predominantly related to demand, group 3 variables to price, and the bottom variables to balancing services.

Group 1 is the top ranked group containing three very closely ranked variables: the wind forecast (WF) – top ranked variable-, the wind forecast matched in the day-ahead market (WindPBF) and the temperature (TEMP), followed by the percent of demand forecast met/not met in the day-ahead market (percenDF). This group highlights the connection between the wind related variables and the system´s imbalance, and confirms the significance of percenDF as expected in the previous chapter.

The second group consists of the technical constraints energy (TCEnergy), demand load met by day-ahead market (DemandPBF) and the demand forecast (DF), all with a very close ranking. The third group of the top 10 variables is rounded of by the percent of the wind forecast in excess/deficit of the wind production matched in the day-ahead market.

The bottom 4 ranking variables were not grouped as their %IncMSE difference was larger among them than it was within the top three groups. The wind ramps follows group three, then secondary regulation band (upwards and then downwards, SRBBu and SEBd, respectively). The additional upwards reserves (RPAS) was the lowest ranked variable trailing far behind all others.

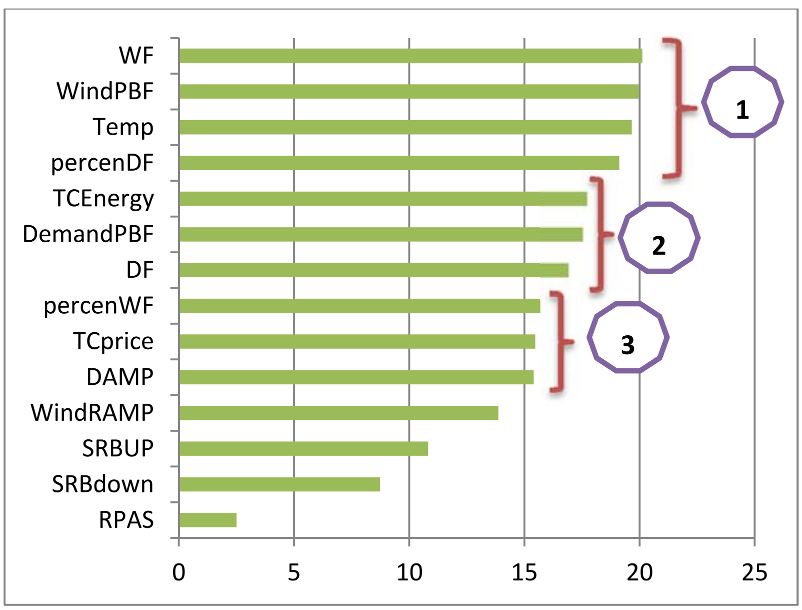


Figure 6.1: Aggregated variable importance (%IncMSE) plot for Model 6 with H=4.

Variable	Mean Norm % IncMSE	Mean %IncMSE	Median Norm % IncMSE	Median %IncMSE	STD %IncMSE
WF	0.7118515	20.122661	0.70927675	19.936942	6.5337547
WindPBF	0.7062795	19.965151	0.71503654	20.098843	6.8167821
Temp	0.6955443	19.661687	0.67011663	18.836197	7.2820799
percenDF	0.6763435	19.118917	0.65257166	18.343029	8.4403636
TCEnergy	0.626788	17.718081	0.59509307	16.727372	6.2629539
DemandPBF	0.6200425	17.527398	0.60214741	16.925662	5.8458833
DF	0.5981612	16.908855	0.57937469	16.285547	5.2036697
percenWF	0.554933	15.686879	0.51935217	14.598384	6.3801681
TCprice	0.5474395	15.475053	0.52038418	14.627393	4.333044
DAMP	0.5448385	15.401527	0.52485156	14.752965	4.0168081
WindRAMP	0.4907284	13.871939	0.45818808	12.879133	6.0537538
SRBUP	0.3827281	10.818981	0.35408954	9.9530441	5.1087571
SRBdown	0.3090343	8.7357983	0.28790322	8.0926238	4.8924408
RPAS	0.088657	2.5061601	0.05272413	1.4820138	3.4004251

Table 6.1: Summary of aggregated variable importance measure (%IncMSE) for Model 6 with H=4.

From the next section, Models 10 to 16[1] were selected based on the variable importance ranking by omitting the lower ranked variables (Models 12-15), starting with the bottom most variable first, then adding the second lowest to the omission and so no. Then, group 2 variables were omitted (Model 16), followed by omitting group 2 as well, which means only the top group of variables were used (Model 10).

6.2 MODEL VARIATIONS

The model variations were selected based on 1- their forecasting horizon (pre- vs post day-ahead market) and 2- as part of the sensitivity analysis (discussed above). In order to understand the potential intra-day market effect on savings, the post-day ahead models were grouped into further categories: pre-intraday market (IM) session 1 and pre-IM session 2 (based on time frame availability of secondary regulation band information). The variable combinations for each model considered are contained in the table below.

[1] Model No. 11 is not included in the results of this Project.

VARIABLE	MODEL 1	MODEL 2	MODEL 3	MODEL 4	MODEL 5	MODEL 6	MODEL 7	MODEL 10	Model 12	Model 13	Model 14	Model 15	Model 16
Demand Forecast	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
Demand PBF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
DemandPercen	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind Forecast	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wind PBF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
WindPorcen	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗
DayAhead Price	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗
Constraints (price)	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗
Constraints (energy	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗
RPAS	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
SecondaryRegBand	✓	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗
SecundariaBand-Up	✓	✓	✗	✗	✓	✓	✗	✗	✓	✓	✗	✗	✗
Temp	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
WindRAMP Forecast	✗	✗	✗	✗	✗	✓	✓	✗	✓	✓	✓	✗	✗

A = Pre-day Ahead
B.1= Post day-ahead, pre-ID session 1
B.2= Post day-ahead, pre-ID session 2

Two main model categories were simulated based on the forecasting horizon. A third sub-category for the post-day ahead category classifies forecasts as either pre intraday (ID) market session 1 or 2.

Table 6.2: Variable combinations for Models' simulated.

These are summarized below:

A. **Pre-DA model -** ***Benchmark model:***

- Model 5: all variables except windRAMPS.

B. **Post-DA models:**

- Model 2 & 3: Pre IM 1 and IM 2; all variables ***except*** **windRAMPS[2].**
- Model 6 & 7: PreIM1 and IM2; all variables**.**
- Model 10-16: Variations based on variable importance[1].
  - Model 12 – omits bottom variable
  - Model 13- omits bottom 2 variables (RPAS, SRBdown)
  - Model 14 – omits bottom 3 variables (RPAS, SRBdown, SRBup)

[2] As mentioned in the previous chapter, windRAMPS was identified *after* the first simulation set.

  - Model 15 omits bottom 4 variables (RPAS, SRBdown, SRBup, windRAMP). Includes group 1-3 variables.
  - Model 10: omits bottom and group 3 variables (RPAS, SRBdown, SRBup, windRAMP, DAMP, TCprice, percenWF). Includes group 1 and 2 variables.
  - Model 16: omits bottom and group 3 and 2 variables (RPAS, SRBdown, SRBup, windRAMP, DAMP, TCprice, percenWF, DF, DemandPBF, TCenergy,). Includes group 1 only.
- Model 1, 4, Ad hoc variations pre-variable importance analysis.
  - Model 1: to gauge effect of omitting temperature.
  - Model 4: to incorporate demand and wind production variables only.

6.3 Forecasting Model Results

The results presented in this section correspond to the 128 simulations using input data from April 2015 to March 2016, yielding 5,685 forecast hours – approximately 7.8 months' worth of forecast points for each simulation. The reduction in forecast points from the number in the input data set data are a result of missing variable data needed for the combination of training horizons with corresponding test data sets, reducing the amount of forecast points.

The simulation results for the top performing training horizon for each model are included in this section´s tables. The first column of the results' tables ranks the model in order of performance, from highest to lowest, based on the *value* for the first listed evaluation metric. As the performance ranking for the remaining metrics in the same table may or may not coincide with the first metric, the highest and lowest performances for each metric have been color coded, with green and red, respectively. The median metric corresponds to the median value for all training horizon case results of each model, whereas the horizon listed corresponds to the top performing case for the metric and model in question.

6.3.1 Model Forecasting Performance

The simulation results for the top performing training horizon for each model are included in Table 6.2 and Table 6.5, for error and forecast accuracy metrics, respectively.

6.3.1.1 *Forecast Volume Error Measures*

The RMSE and MAE included in Table 6.3 are error measures: the lower the value, the higher the performance, and vice versa. All post-day-ahead market models outperformed the benchmark Model 5 - the pre-day ahead gate closure model. The post-day ahead market model using all 14 variables (Model 6) resulted in the highest RMSE performance. The error slowly increased from there with each consecutive omission of the bottom three variables (RPAS and Secondary band regulation in both directions) based on the variable´s importance. Different omission variations of the bottom variables deteriorated de RMSE only slightly, and not quite like the single omission of TEMP or just keeping the top variables which yielded the highest error of the post day ahead models.

Model 5, the pre-day ahead benchmark model was the lowest performer (higher error value results in lower performance). Although the error variance among the post-day ahead models is small, it is much higher when it comes Model 5 - an RMSE about 10% higher than the top model. In terms of the median error for all of training horizon cases simulated for each model, Model 7 exhibited the best performance (omits secondary regulation band). In terms of training horizons, 8 days consistently yielded the better results for all models.

A description of the model in ranked order for RMSE performance is given in Table 6.4 below.

Model	**RMSE**			**MAE**		
	H	**Value**	**Median**	**H**	**Value**	**Median**
Model 6	*8*	825.32	859.54	*8*	644.23	664.44
Model 12	*8*	825.55	860.89	*8*	644.05	664.76
Model 13	*8*	828.02	861.82	*8*	646.23	666.57
Model 14	*8*	829.23	860.09	*8*	648.38	665.52
Model 2	*8*	829.46	863.54	*8*	648.44	667.98
Model 7	*8*	830.87	830.87	*8*	649.48	667.46
Model 10	*8*	831.51	860.36	*8*	650.24	665.85
Model 3	*8*	832.71	864.53	*8*	652.98	669.56
Model 15	*8*	834.48	867.04	*8*	654.19	670.81
Model 4	*8*	843.34	863.97	*8*	658.96	667.26
Model 1	*8*	851.51	878.30	*8*	664.93	679.24
Model 16	*8*	859.29	876.09	*1*	670.27	682.67
Model 5	*8*	909.48	949.81	*8*	711.15	738.48

Best performer
Worst performer
H = Model Training Horizion

Table 6.3: Forecasting model results metrics: RMSE and MAE

RMSE	
Model No.	**Variables**
6	all variables
12	omission bottom variable (RPAS)
13	omission bottom 2 variables (RPAS, SRBdown)
14	omission of bottom 3 varibles (RPAS, SRBdown
2	omission of windRAMP
7	omission of SRBdown, SRBup
10	omission of botton half of variables (RPAS, SRBdown, SRBup, windRAMP, DAMP, TCprice, percenWF)
3	omission of windRAMP, SRBdown, SRBup
15	omision of bottom 4 varibles (RPAS, SRBdown SRBup,
4	omission of all except wind and demand variables
1	omission of Temp
16	omission of all variables except top group
8	omission of all except (RPAS and DemandPBF)
9	omission of all except (RPAS, DemandPBF, SRBup)
5	BENCHMARK/Pre-day ahead

Table 6.4: Model description in order of RMSE performance.

6.3.1.2 *Forecast Correlation and Direction Accuracy*

Model 12 (omits RPAS) followed closely by Model 6 (all variables) results represent the highest correlation between the forecasted value and the observed imbalance, although only marginally better than the other post-day ahead models. On the other hand, Model 4 – which is a simplified post-day ahead model with only demand and wind related variables (6 total) – was best at accurately forecasting the direction of the imbalance, at over 68%m followed by Model 14 (omits the bottom three variables). The remaining post-day ahead models followed closely behind all with percentages in the range of 67%. Yet again Model 5 exhibited the lowest performance in both metrics, with a correlation at least 10% lower than all the post-day ahead models, and a decrease in foresting accuracy of the imbalance direction of over 3% with respect to the other models.

The longest training horizon (8 days) yielded the highest correlations, except for Models 1, 4 and 16 which were the exact opposite (1 day training horizons). However, in terms of forecasting the imbalance direction, the shorter training horizons yielded the best results.

Model	Correlation			Accuracy of Forecast Direction,%		
	H	Value	Median	H	Value	Median
Model 12	8	0.4487	0.4183	5	67.810%	67.441%
Model 6	8	0.4474	0.4190	5	67.933%	67.529%
Model 13	8	0.4463	0.4167	2	67.669%	67.318%
Model 7	8	0.4437	0.4151	4	67.792%	67.537%
Model 2	8	0.4433	0.4158	2	67.863%	67.432%
Model 14	8	0.4431	0.4155	4	68.056%	67.634%
Model 16	1	0.4418	0.3881	1	67.423%	66.095%
Model 4	1	0.4406	0.4140	2	68.303%	67.590%
Model 10	8	0.4394	0.4165	5	67.968%	67.731%
Model 3	8	0.4388	0.4131	2	67.757%	67.546%
Model 15	8	0.4376	0.4131	4	67.704%	67.203%
Model 1	1	0.4266	0.3951	1	67.617%	67.010%
Model 5	8	0.3131	0.2842	3	64.714%	63.641%

Table 6.5: Forecasting model results metrics: Correlation (forecast/observed) and accuracy of forecast direction.

Accuracy of Forecast Direction	
Case Rank **(best=1 worst=8)**	*All models:* **Average, H**
1	3.08
2	3.67
3	3.83
4	3.17
5	4.67
6	4.42
7	6.58
8	7.17

Table 6.6: Forecast accuracy of system imbalance direction for all model case simulations.

6.3.2 Best Performing Models

Observed values and model output for Model 6 are contained in Figure 6.2.

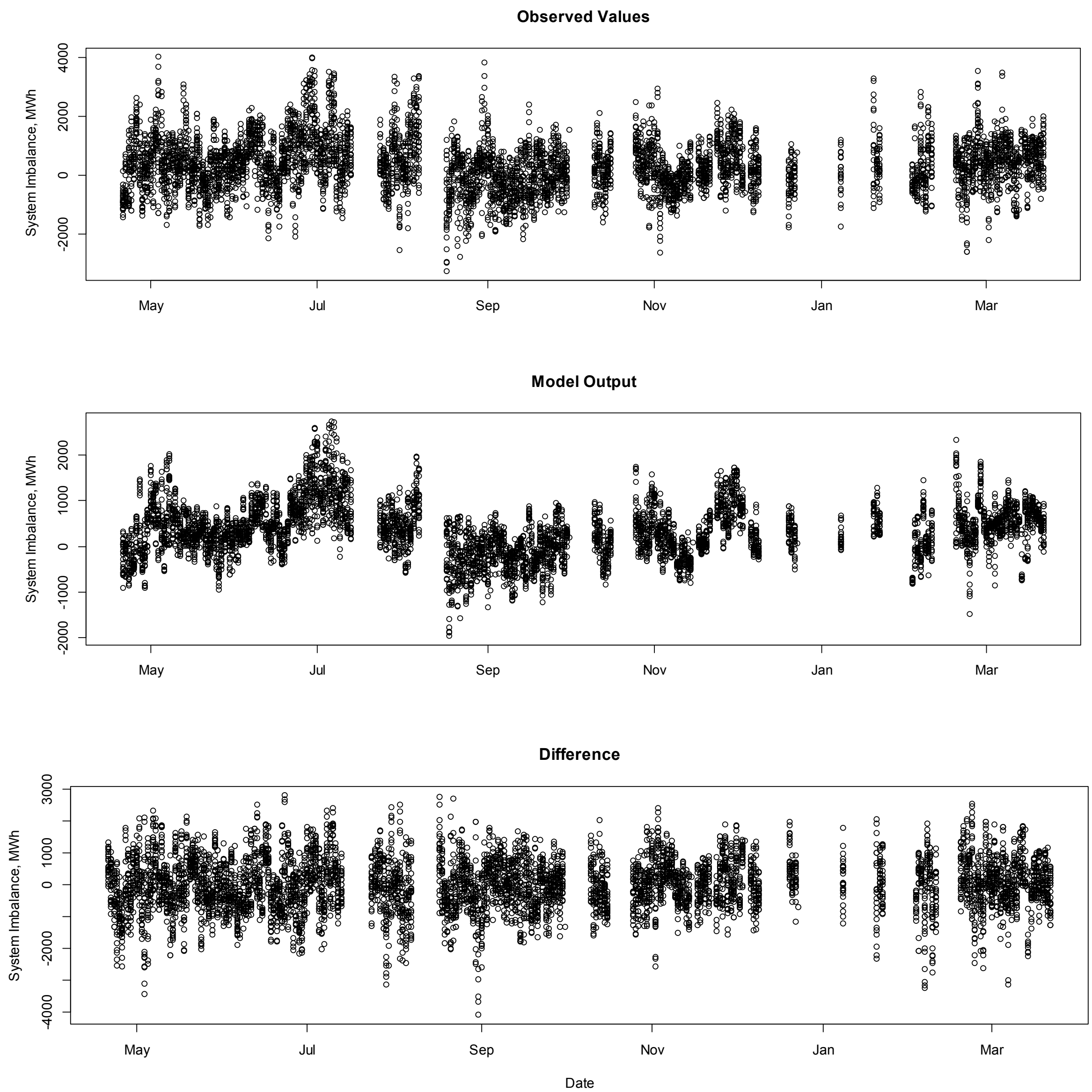


Figure 6.2: Results for Model 6, Training Horizon 8.

The results for Model 6 (lowest RMSE) and Model 12 (lowest MAE) were analyzed to gain additional insight on properties of the forecasting results. Table 6.7 and Table 6.8 contain the evaluation metrics for all the simulated training horizon cases, with each metric ranked from top to bottom by highest to lowest performing case, and the corresponding horizon case listed in the column to the left of each metric. From the error metrics we can observe a tendency for the error to generally increase – a total of almost 6% for model 6 - as the training horizon shortens. No similar tendency could be inferred from the direction accuracy, although it would appear that mid to lower horizons performed better.

MODEL 6					
Horizon, H	RMSE	Horizon, H	MAE	Horizon, H	Direction Forecast Accuracy, %
8	825.317	8	644.225	5	67.93%
7	838.172	7	652.476	2	67.88%
6	849.209	6	656.52	6	67.85%
5	855.506	5	662.691	1	67.53%
1	863.57	4	666.195	4	67.53%
4	864.113	2	667.766	3	67.37%
2	866.169	1	668.543	8	67.30%
3	876.796	3	676.539	7	67.07%

Table 6.7: Model 6 results metrics for all training horizon cases.

MODEL 12					
Horizon, H	RMSE	Horizon, H	MAE	Horizon, H	Direction Forecast Accurate, %
8	825.547	8	644.045	5	67.81%
7	839.376	7	653.5	2	67.72%
6	849.952	6	657.355	6	67.69%
5	856.437	5	662.929	1	67.48%
4	865.349	4	666.584	3	67.41%
1	865.442	2	669.218	4	67.35%
2	868.264	1	669.827	8	67.14%
3	879.136	3	678.29	7	67.12%

Table 6.8: Model 12 results metrics for all training horizon cases.

Although in actuality positive imbalances were observed to be more common than negative ones, the model forecasts overestimated the number of positive imbalances (7% more than observed for Model 6). Yet it accurately forecasted the direction of almost 73% of those positive imbalances. A data summary of these results is found in Table 6.9 and Table 6.10. Conversely, the negative imbalances were underestimated and the direction forecast accuracy was almost 18% lower than for the positive imbalances for Model 6.

MODEL 6, H=5		
Imbalance Direction	No. Obs	5685
	Forecasted	*Observed*
Positive		
No.	4061	3678
%	71.43%	64.70%
MAE	786.6493	
PredDir	72.84%	
Negative		
count	1642	2007
%	28.88%	35.30%
MAE	618.5461	-
PredDir	55.05%	

Table 6.9: Data summary for Model 6.

MODEL 12				
	No. Obs 5685			
Imbalance Direction	*Forecasted*			*Observed*
	H = 8	H = 5	H = 4	
Positive	-			
No. Hours	4132	4050	3976	3678
%	72.68%	71.24%	69.94%	64.70%
MAE	634.78	670.91	674.86	
PredDir No. Hour	2971	2949	2899	
%	71.90%	72.81%	72.91%	
Negative				
No. Hours	1553	1635	1709	2007
%	27.32%	28.76%	30.06%	35.30%
MAE	668.69	643.15	647.32	-
PredDir No. Hour	846	906	930	-
%	54.48%	55.41%	54.42%	

Table 6.10: Data summary for Model 12.

In terms of the actual level of imbalances, the forecast volumes tended to be lower than the observed values. The frequency distribution of both value types can be seen in the histogram below.

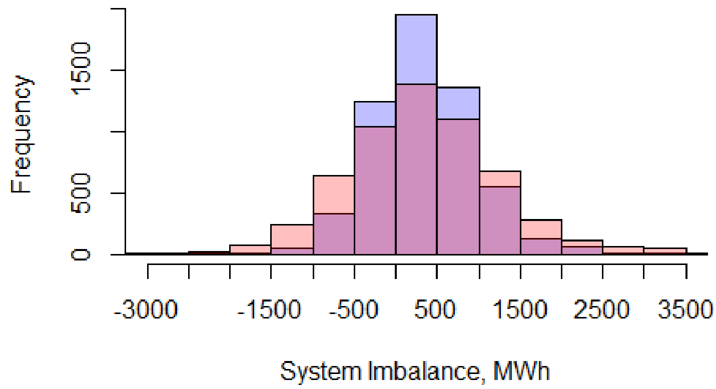


Figure 6.3: Frequency distribution of forecasted and observed imbalance volumes: Model 12, Training Horizon 8.

Finally, as the magnitude of the imbalance increases, the accuracy of forecasting the imbalance direction appears to increase as well, as shown in Table 6.11.

MODEL 6, H=5			
Forecast Magnitude >	**IMB >0**	**IMB<0**	**Total**
0	0.728392	0.5566502	0.6793316
200	0.7565033	0.5720891	0.7163006
300	0.7648489	0.5915895	0.7312569
400	0.7845414	0.596926	0.7556799
500	0.7927412	0.6131997	0.7719228
600	0.7952846	0.6282311	0.7800995
700	0.8049793	0.6364709	0.7957055
800	0.8142123	0.6444543	0.8051044
900	0.8252119	0.6502848	0.8178332
1000	0.8236915	0.6581972	0.8212005
1100	0.825046	0.663944	0.8276451
1200	0.8333333	0.6672358	0.8329621
1300	0.858006	0.6682854	0.8502825
1400	0.8830645	0.6700386	0.8825758
1500	0.895288	0.6718238	0.8926829

Table 6.11: Forecasting accuracy of imbalance direction as forecast magnitude increases.

6.3.3 OBSERVATIONS

The post-day ahead models in all instances outperformed the pre-day ahead model. The improvement of post vs pre is significant enough that their application would most certainly be warranted. Although by a small margin, the results obtained from using all the variables (Model 6) generally yielded the lowest error index of all post-day-ahead models. Removing a combination of the bottom ranked variables showed the lowest impact on the error measures, with RPAS exhibiting the lowest impact. Omission of the meteorological related variables – i.e temperature – has a larger impact on the error, but preserving just top ranked variables results in the largest deterioration of the results.

The longest training-horizon was best at decreasing error, meaning more accurate forecasts of volume magnitudes were obtained with their application. However, if we consider the forecast as a classification problem instead, the shorter term horizons proved superior at forecasting the direction of the system imbalance.

6.4 BIDDING STRATEGY RESULTS

Upon obtaining the forecast results, the bidding strategy component of the model was applied. As noted by Garcia & Kirschen, 2006, "the true measure of improvement when forecasting market imbalance volumes is not an abstract error index but rather the savings in balancing costs that this improvement makes possible".

The cost savings is used as the main evaluation metric for the strategy as a whole, and presented as the "gross savings" in this section. The effects of transacting in the different intra-day markets was evaluated separately (referred to as the net savings herein) and intended as complementary information to further comprehend the potential effect on gross savings.

A second bidding strategy was simulated after evaluating the results of the original strategy, as modification of the first one.

The cost overruns resulting from the Trader's portfolio imbalance are estimated at over 10€ million for the year 2015. The Trader´s actual portfolio cost overruns corresponding to the hours evaluated in the simulation is 6,527,859 € providing further context to the results presented in the following sections.

Real portfolio cost overruns (CO) *without* strategy:

6.4.1 Gross Savings

Table 6.12 contains the imbalance cost reduction results from applying the bidding strategy component to the imbalance forecast results. As with the results tables for the forecasting model, the results presented in this sections correspond to the best performing case for each model type, and ranked in order of highest to lowest performance based on the "Value" column for the first metric presented in table. The median is also provided in the same manner as was done earlier.

Other complementary metrics provided are the savings in euros (€) per each energy unit (MWh) transacted through this bidding strategy and percentage savings the cost reduction represent to the Trader´s real cost overruns. The former metric enables us to gauge the efficiency of the strategy, while the latter provides further real –world context. The final column shows the potential yearly savings estimated by escalating the calculated savings to a yearly amount. Although not quite as insightful as having hard results for the full year, it does at least provide an idea of the potential.

Model	Total Gross Savings[1]					Improv. over lowest perfomance[4]	Est. Yr. Gross Savings	
	H	Value, €	Median, €	€/MWh[2]	% of CO		€	% of CO
Model 12	4	483,457	378,717	1.1842	7.41	88.75	739,404	11.33
Model 16	1	470,476	196,920	1.1209	7.21	83.68	719,552	11.02
Model 6	4	459,497	379,494	1.1530	7.04	79.39	702,761	10.77
Model 13	4	420,511	377,354	1.0275	6.44	64.17	643,134	9.85
Model 7	4	406,538	386,789	0.9974	6.23	58.72	621,764	9.52
Model 10	4	405,110	365,934	1.0164	6.21	58.16	619,580	9.49
Model 14	7	399,897	363,800	1.0580	6.13	56.13	611,607	9.37
Model 4	4	387,606	322,987	0.9576	5.94	51.33	592,810	9.08
Model 2	1	387,201	345,264	0.9612	5.93	51.17	592,189	9.07
Model 15	4	371,047	310,987	0.9142	5.68	44.86	567,483	8.69
Model 3	1	365,066	318,037	0.9133	5.59	42.53	558,336	8.55
Model 1	5	360,606	315,266	0.7580	5.53	40.79	551,514	8.45
Model 5	8	256,137	175,657	0.7645	3.92	0.00	391,740	6.00%

1. Gross Savings do not consider the intraday market cost/benefit.
2. € of savings per MWh of energy applied to correct the portfolio´s imbalance.
4. Based on Gross Savings.

Table 6.12: Bidding strategy application results: Gross savings.

Once again *all* post-day ahead models outperformed the day-ahead model, with savings over 88% higher for the top performing model. Model 12, which excludes the RPAS variable, yielded the greatest savings at $483,457, or approximately 7.41% of imbalance cost overruns, whereas the day-ahead market model yielded almost half that amount.

Model 12 was also superior in the per unit savings measure, saving on average 1.18€ for every MWh transacted through this strategy. Interestingly, the two models that followed in ranking included either 1- just the top 4 variables (group 1, Model 16) or all variables (Model 6). Table 6.12 shows the variance in savings from the top to bottom performing *post-day ahead* models to be around 25%.

Four days appeared to be the training horizon case yielding the highest savings for the majority of models. Only two models achieved its highest results by using a longer training horizon, of which two are the bottom ranked models.

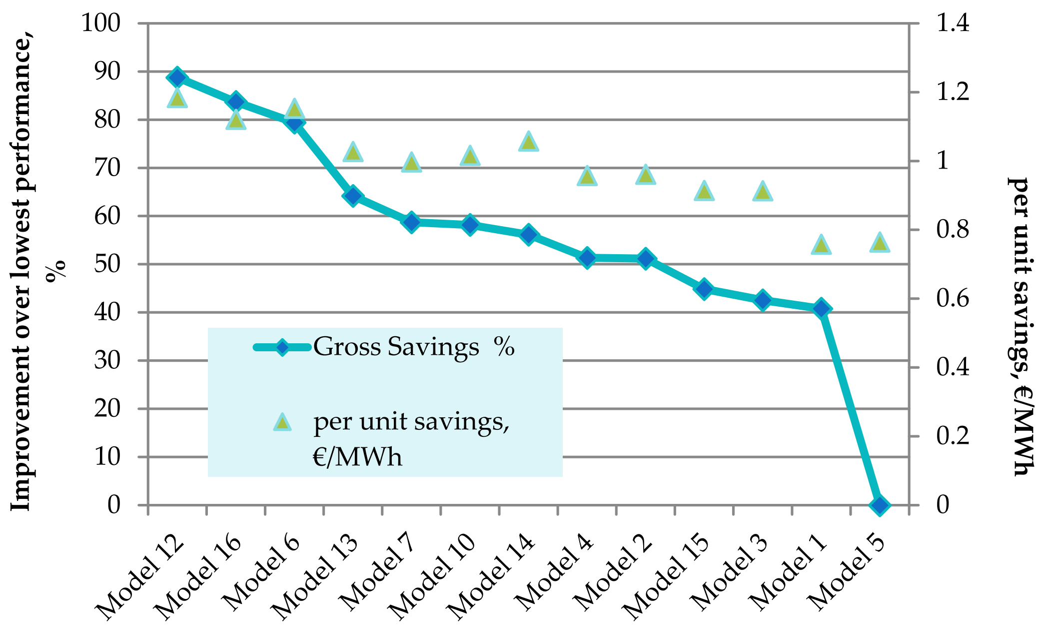


Figure 6.4: Plot of performance improvement (gross savings and per unit savings results

Gross Savings	
Model No.	**Variables**
12	omission bottom variable (RPAS)
16	omission of all variables except top group
6	all variables
13	omission bottom 2 variables (RPAS, SRBdown)
7	omission of SRBdown, SRBup
10	omission of botton half of variables (RPAS,
14	omission of bottom 3 varibles (RPAS, SRBdown
4	omission of all except wind and demand variables
2	omission of windRAMP
15	omission of bottom 4 varibles (RPAS, SRBdown
3	ommisison of windRAMP, SRBdown, SRBup
1	omission of Temp
5	BENCHMARK/Pre-day ahead

Table 6.13: Ranked model description for Gross Savings.

From observing the plot of the daily evolution of the profits and losses obtained by applying the strategy, as done in Figure 6.5, the outcome can be perceived as volatile. Except for the pre-day ahead model (Model 5), the general shape appears to be similar for all models. The beginning of the simulation experiences fast gains which peak twice before two significant drops.

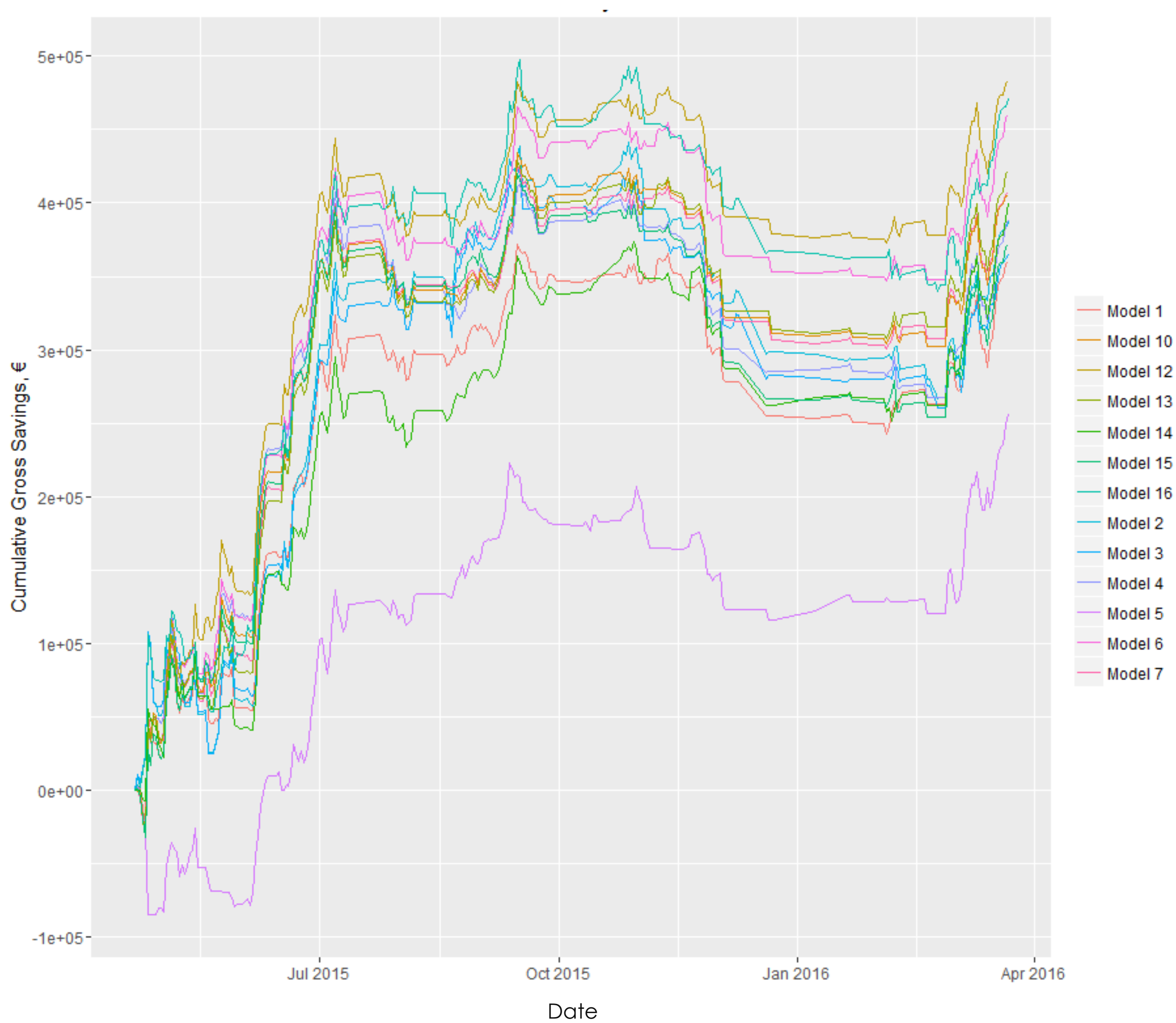


Figure 6.5: Daily cumulative evolution of gross savings results for all models.

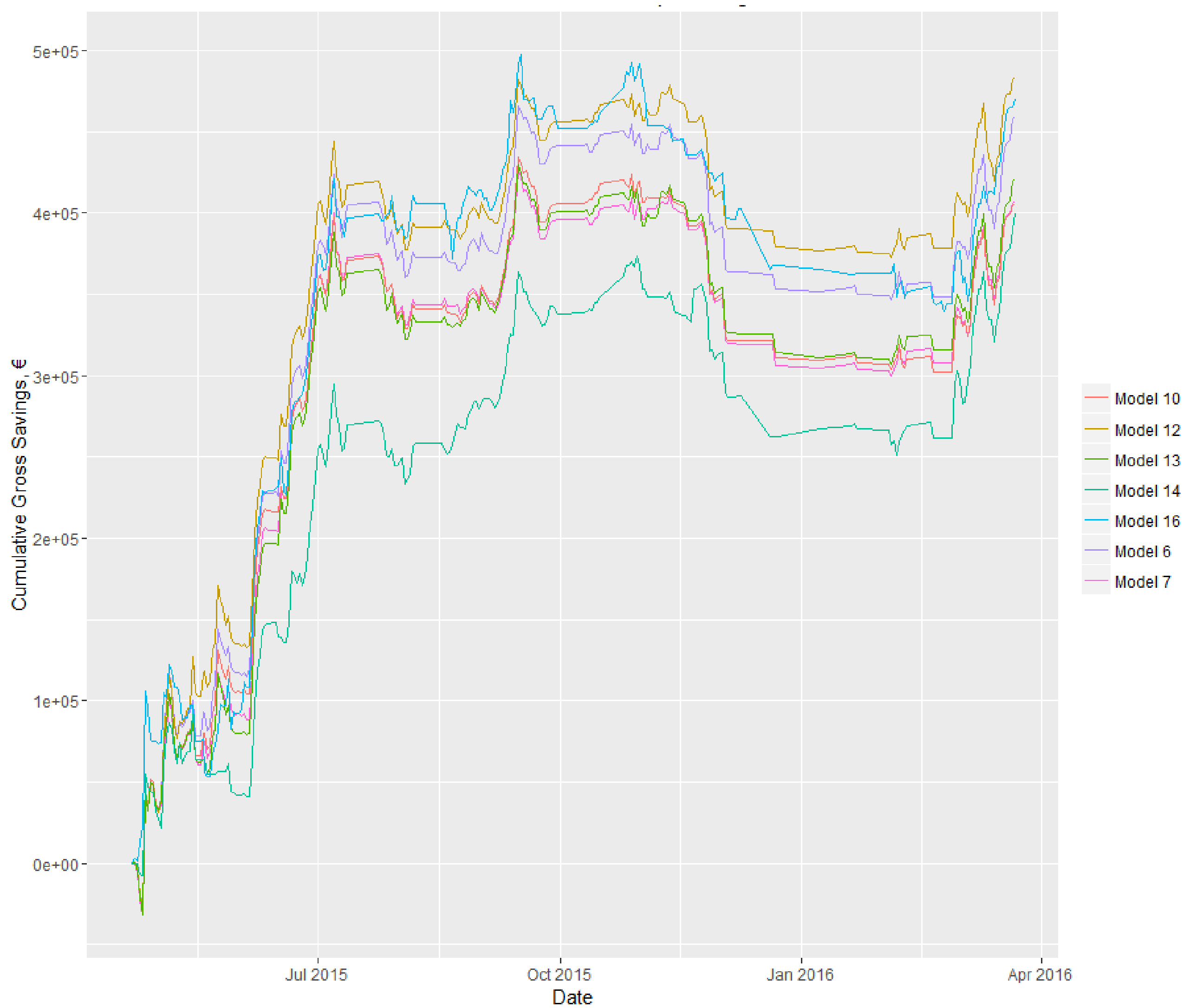


Figure 6.6: Daily cumulative evolution of gross savings results for top half of models.

6.4.2 Intra-Day Market Effects on Savings

To gain better understanding of the potential effects that the intra-day market prices could have on the savings, the net savings were determined based on the spread of the intra-day market price (IDMP) with respect to the day-ahead market price. This will have no impact on the day ahead model as adjustments would be internalized within the bid in that market.

There are four different scenarios that can occur when adjusting the program in the intra-day market:

1- DAMP < IDMP
   a. If ***increasing program***/*selling more energ*y → the spread is a per unit ***profit.***
   b. If **decreasing the program**/*buying more energy* → the spread is a per unit ***loss***.

And vice versa,

2- DAMP>IDMP
   a. If ***increasing program***/*selling more energ*y → the spread is a per unit ***loss.***
   b. If ***decreasing the program***/*buying more energy* → the spread is a per unit ***profit***.

The net savings incorporating the spread of the intraday/day-ahead market, as represented by the cases above, was calculated each model and case, and top results for each model are shown in Table 6.14. All the savings results were reduced. For example, the highest gross savings achieved were reduced to by about 18%. The models that required transacting in the second ID market session experienced the greatest reduction from the ID effect, as shown in Table 6.15 which contains the savings reduction ordered from highest to lowest reduction in gross savings. The impact for those models was an average reduction in savings of around 47%. Savings from the ID session 1 models were reduced on average by 18%. Considering that market liquidity in Spain is often reduced with each consecutive ID market session, it is not surprising that the first ID session models would experience lower impact than the second ID session.

Model 16 (only group 1 variables) and Model 12 (excepting RPAS) still round up the top spots. However, Model 6 -which yielded the largest gross savings - experienced a reduction of over 32%

Model	Total Net Savings[3]		
	H	Value	Median
Model 16	1	394,904 €	149,102 €
Model 12	4	342,983 €	195,251 €
Model 10	4	327,265 €	258,442 €
Model 4	4	326,524 €	256,622 €
Model 14	4	324,863 €	271,810 €
Model 7	4	324,263 €	295,980 €
Model 6	4	308,791 €	181,053 €
Model 15	4	283,658 €	214,872 €
Model 3	1	273,253 €	229,550 €
Model 13	4	268,881 €	181,074 €
Model 5	8	256,137 €	175,657 €
Model 2	4	225,712 €	173,264 €
Model 1	5	195,355 €	106,713 €

3. Net Savings internalize the IM market price cost/benefit w/ respect to the day ahead price.

Table 6.14: Strategy results: Total Net Savings.

ID market Effect on Savings				
Model	**Gross, €**	**Net,€**	**Δ€**	**Δ%**
Model 1	360,606	195,355	-165,251	-45.83%
Model 2	387,201	225,712	-161,489	-41.71%
Model 13	420,511	268,881	-151,630	-36.06%
Model 6	459,497	308,791	-150,706	-32.80%
Model 12	483,457	342,983	-140,474	-29.06%
Model 3	365,066	273,253	-91,813	-25.15%
Model 15	371,047	283,658	-87,389	-23.55%
Model 7	406,538	324,263	-82,275	-20.24%
Model 10	405,110	327,265	-77,845	-19.22%
Model 14	399,897	324,863	-75,034	-18.76%
Model 16	470,476	394,904	-75,572	-16.06%
Model 4	387,606	326,524	-61,082	-15.76%
Model 5	256,137	256,137	0	0.00%

A = Pre-day Ahead
B.1= Post day-ahead, pre-ID session 1
B.2= Post day-ahead, pre-ID session 2

Table 6.15: Effect of ID market price on Gross Savings.

Even after considering the savings reduction from the ID effect, the strategic bids result in savings that outweigh the ID cost effect. Although the pre-day ahead model is not affected by the ID prices, its results are also outperformed by the vast majority of the post-day ahead models.

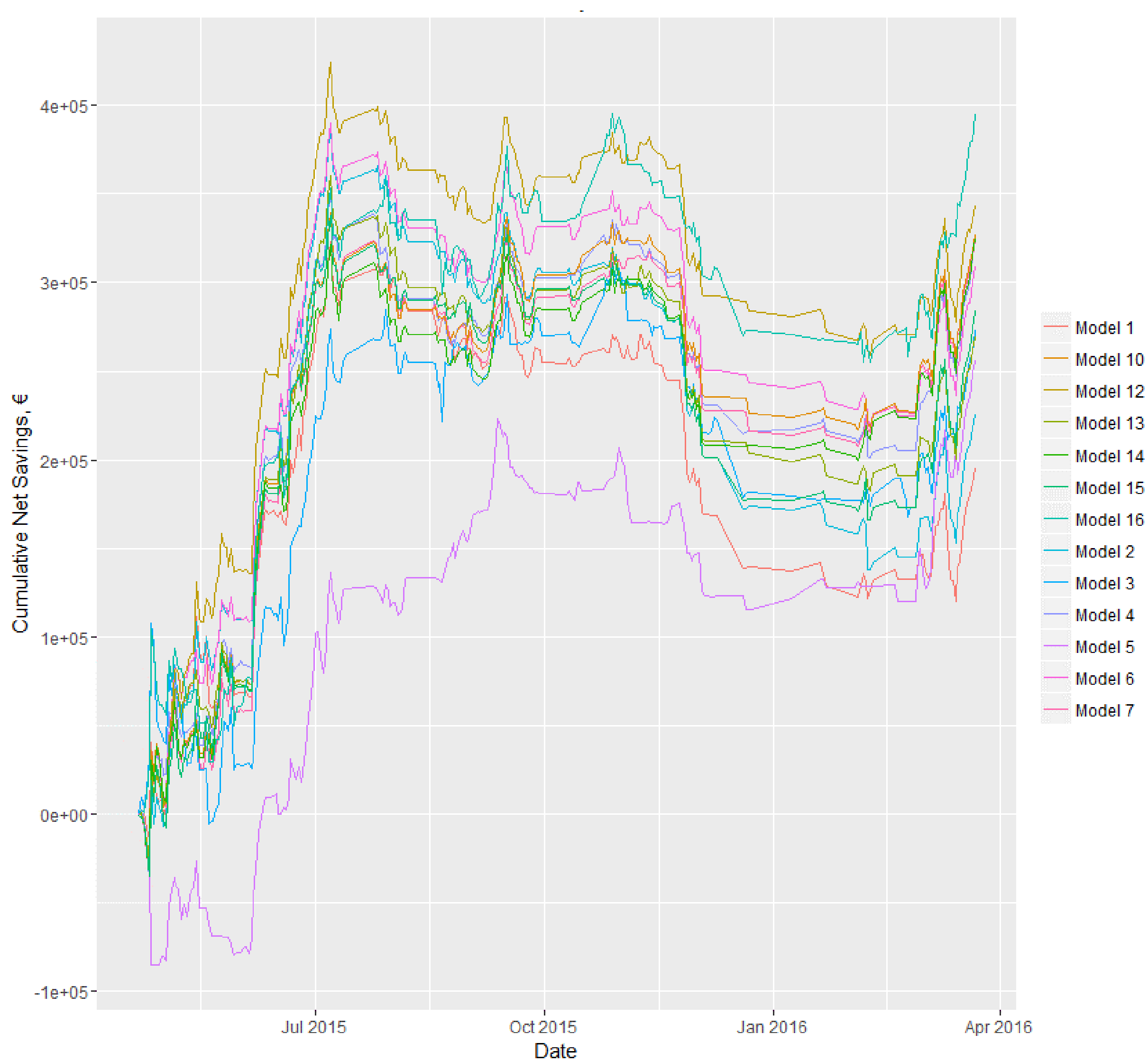


Figure 6.7: Daily cumulative evolution of Net Savings results for all models

6.4.3 MODEL 12 AND 16

Model 12 and Model 16 yielded the highest net savings, despite the fact that the former transacts in the second ID market session. Additional details about their results are presented in this section.

6.4.3.1 *RESULTS OF ALL HORIZONS*

Table 6.16 contains results of all the training horizons for Model 12 (excludes RPAS variable).

The best performing horizon is 4 days, and the difference between its and the next best results are quite significant: 17% improvement in Gross Savings and over 36% improvement in Net Savings from using a day training horizon,

MODEL 12							
Horizonte	Gross Savings, €	Horizon, H	% of OC	Horizon, H	Savings, €/MWh	Horizon, H	Net Savings, €
4	483,457	4	7.41%	4	1.1842	4	342,983
1	401,033	1	6.14%	8	1.0370	5	250,392
8	393,463	8	6.03%	1	0.9828	6	197,222
5	385,815	5	5.91%	5	0.9585	8	195,652
2	371,619	2	5.69%	2	0.9252	1	194,850
3	346,688	3	5.31%	7	0.9034	7	166,497
7	345,591	7	5.29%	6	0.8916	2	151,907
6	338,984	6	5.19%	3	0.8825	3	147,115

Table 6.16: Strategy results for Model 12.

The daily evolution of profit and losses for the gross savings, ID market effects (referred to as IM savings in the graph) and the resulting net savings are plotted in Figure 6.8. Aside from the steep rise in savings at the beginning of the period, the losses from the intraday market are evident from the blue curve, as are two peaks followed by losses continuous losses. The latter peaks are circled in the plot below. These characteristics are better appreciated with the plot in Figure 6.9 which smooths out the daily fluctuations into monthly savings-

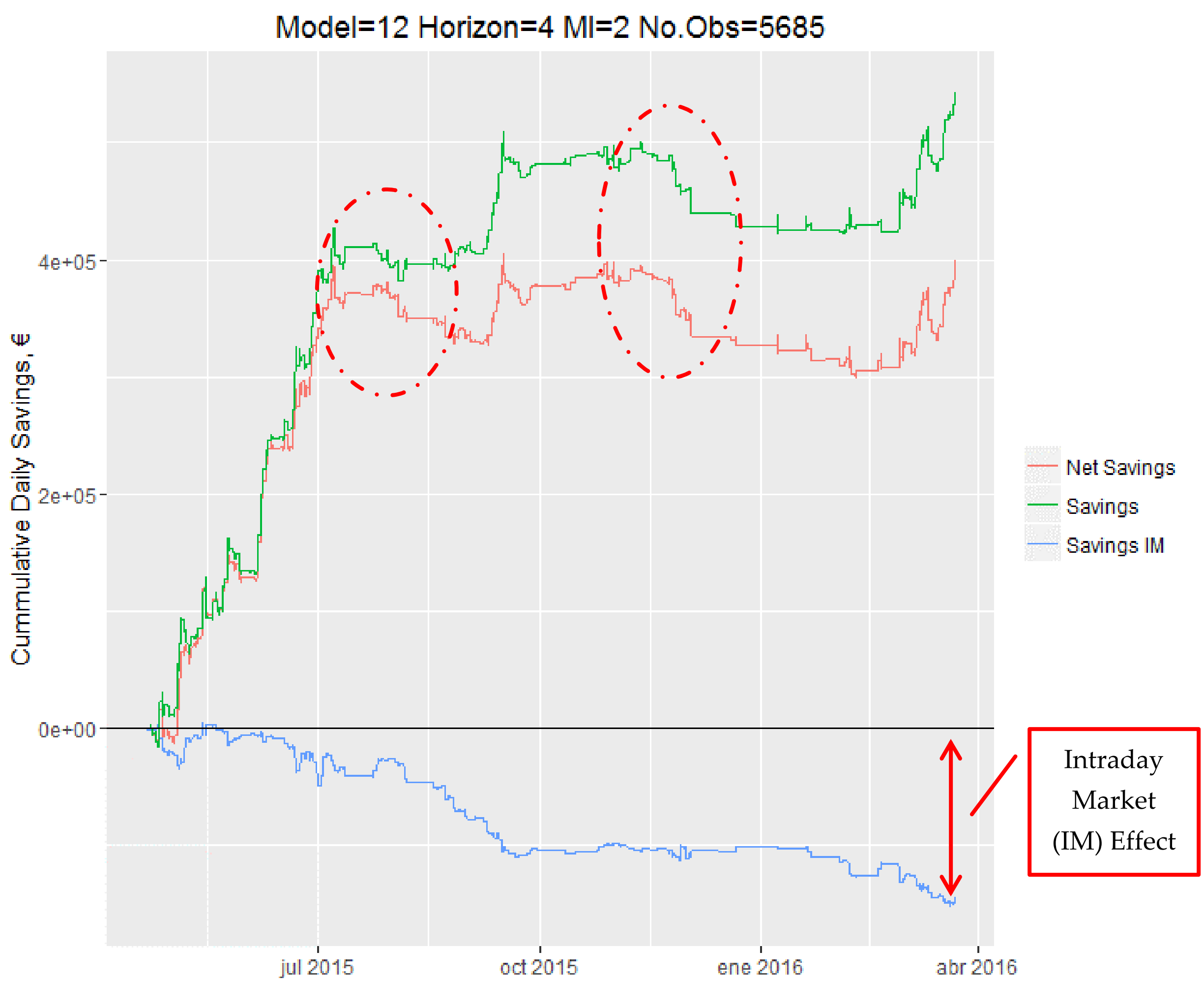


Figure 6.8: Cumulative Daily Gross and Net Savings for Model 12.

Month	No.Obs, hrs.	Approx. No.Days	Monthly Gross Savings	Cum. Gross Savings
Apr-15	240	10	33,183	33,183
may-15	744	31	102,087	135,270
jun-15	714	30	253,068	388,337
jul-15	504	21	-1,334	387,004
aug-15	528	22	9,829	396,833
sep-15	720	30	59,737	456,570
oct-15	334	14	11,008	467,578
nov-15	624	26	-56,133	411,445
dec-15	237	10	-32,512	378,933
jan-16	95	4	-2,757	376,176
feb-16	443	18	36,257	412,433
mar-16	502	21	71,024	483,457

Table 6.17: Monthly savings for Model 12

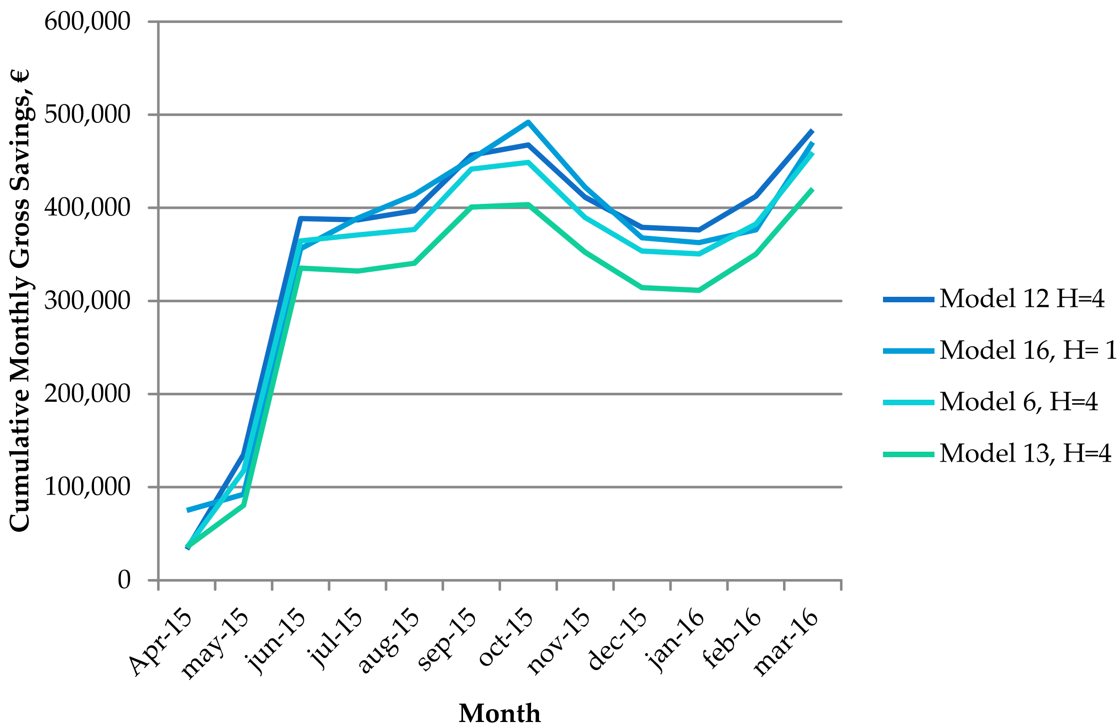


Figure 6.9: Cumulative gross monthly savings from strategy application for top 4 performing models.

6.5 MODIFIED BIDDING STRATEGY RESULTS

Upon identifying the peaks and consequent falls in savings brought up in the previous section, in addition to continually observing the system´s imbalance data on day to day basis, a second or modified strategy was applied

With the intent to reduce continued day-after-day losses, the strategy was modified to bid only bid half of the amount suggested by the strategy function following a day of losses. This also means that benefits following a day of losses, should they occur, would also be reduced in half.

The results of simulating the modified strategy are presented in this section

6.5.1 GROSS AND NET SAVINGS

Table 6.18 contains the gross savings results for the modified strategy. The modification results in increased of savings of up to 8.17% of cost overruns for the Trader. Table 6.19 contains the net savings results. The top performing models are the same as those in the in the original strategy.

Model	Gross Savings[1]				
	H	Value	Median	€/MWh[2]	% of OC
Model 12	4	533,476 €	423,209 €	1.5405	8.17%
Model 16	2	521,632 €	303,237 €	1.4494	7.99%
Model 6	4	498,332 €	420,884 €	1.4631	7.63%
Model 13	4	472,749 €	416,657 €	1.3649	7.24%
Model 10	4	460,886 €	413,894 €	1.3644	7.06%
Model 7	4	455,567 €	429,331 €	1.3192	6.98%
Model 4	2	454,079 €	377,385 €	1.3092	6.96%
Model 14	6	449,953 €	415,788 €	1.3983	6.89%
Model 2	4	431,330 €	399,938 €	1.2427	6.61%
Model 15	4	427,719 €	367,999 €	1.2421	6.55%
Model 1	5	412,461 €	354,151 €	1.1900	6.32%
Model 3	6	407,706 €	379,517 €	1.2747	6.25%
Model 5	8	351,734 €	248,694 €	1.2580	5.38%

1. *Gross Savings do not consider the intraday market cost/benefit.*
2. *€ of savings per MWh of energy applied to correct the portfolio´s imbalance.*

Table 6.18: Modified Bidding strategy application results: gross savings.

Model	Net Savings[3]		
	H	Value	Median
Model 16	8	455,745 €	257,344 €
Model 10	8	391,252 €	337,227 €
Model 12	5	389,601 €	259,404 €
Model 7	4	382,575 €	350,316 €
Model 14	7	381,355 €	349,818 €
Model 4	2	378,773 €	323,191 €
Model 5	8	351,734 €	248,694 €
Model 15	5	351,652 €	289,536 €
Model 6	4	344,840 €	250,947 €
Model 3	6	334,030 €	302,907 €
Model 13	3	318,222 €	254,946 €
Model 2	4	294,807 €	220,561 €
Model 1	5	256,687 €	168,885 €

2. *Net Savings internalize the IM market price cost/benefit w/ respect to the day ahead price.*

Table 6.19: Modified Bidding strategy application results: Net Savings.

6.5.2 Comparison to Original Strategy

The modified strategy resulted in improved gross savings and net savings for all models. The improvement in gross savings for the post day-ahead models ranged from 10% to 17% (10% for the top performing model) and increasing to a range of 12 to 31% for net savings. The pre-day ahead model saw the greatest improvement, yet in terms of gross savings, all other models still outperformed it. The improvement results are contained in Table 6.20.

Model	**Improvement in Gross Savings**		**Improvement in Net Savings**	
	Δ€	**%**	**Δ€**	**%**
Model 12	50,019 €	10%	46,618 €	14%
Model 16	51,156 €	11%	60,841 €	15%
Model 6	38,835 €	8%	36,049 €	12%
Model 13	52,238 €	12%	49,341 €	18%
Model 10	55,776 €	14%	63,987 €	20%
Model 7	49,029 €	12%	58,312 €	18%
Model 4	66,473 €	17%	52,249 €	16%
Model 14	50,056 €	13%	56,492 €	17%
Model 2	44,129 €	11%	69,095 €	31%
Model 15	56,672 €	15%	67,994 €	24%
Model 1	51,855 €	14%	61,332 €	31%
Model 3	42,640 €	12%	60,777 €	22%
Model 5	95,597 €	37%	95,597 €	37%

Table 6.20: Modified strategy improvement

6.5.3 MODEL 12:

Model 12 also yielded the best results under this strategy. The evolution its daily profits and losses for the gross, IM, and net savings, are shown in the figures below. From comparing the evolution of the profit and losses for the original and modified strategy, as plotted below, it can be seen how the first strategy yields greater savings for few first months of the simulation. However, it also shows how the losses surrounding the two peaks of concern are reduced (see items no. 1 and 2 blown up for additional detail). At the first peak the second strategy outpaces the first one and remains as such through the end of the simulation period.

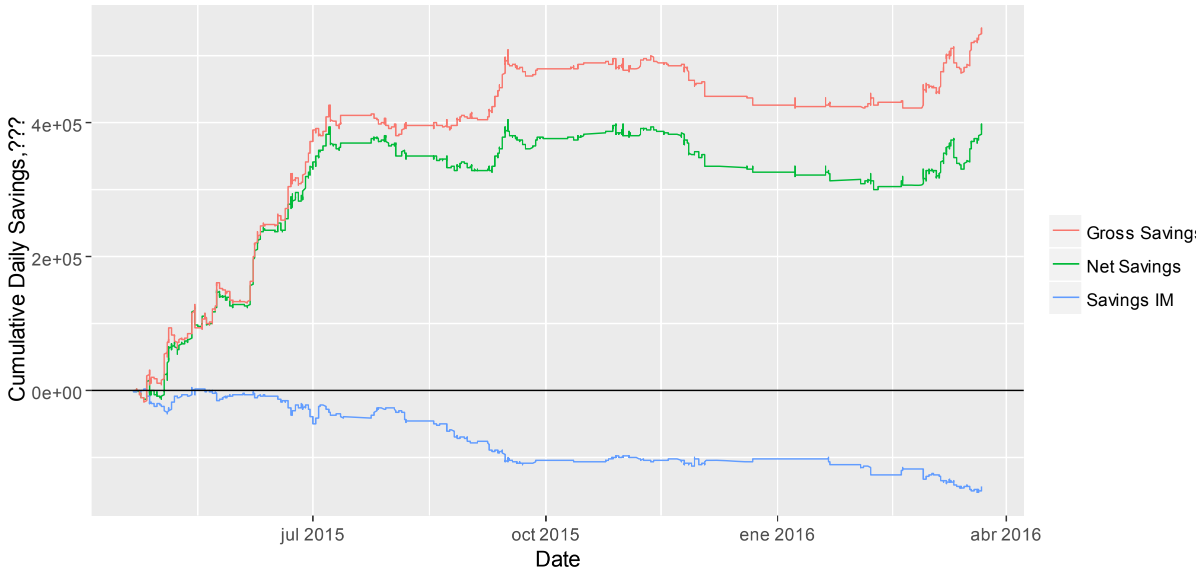


Figure 6.10: Cumulative Daily Gross and Net Savings for Model 12, H=4, with Strategy 2.

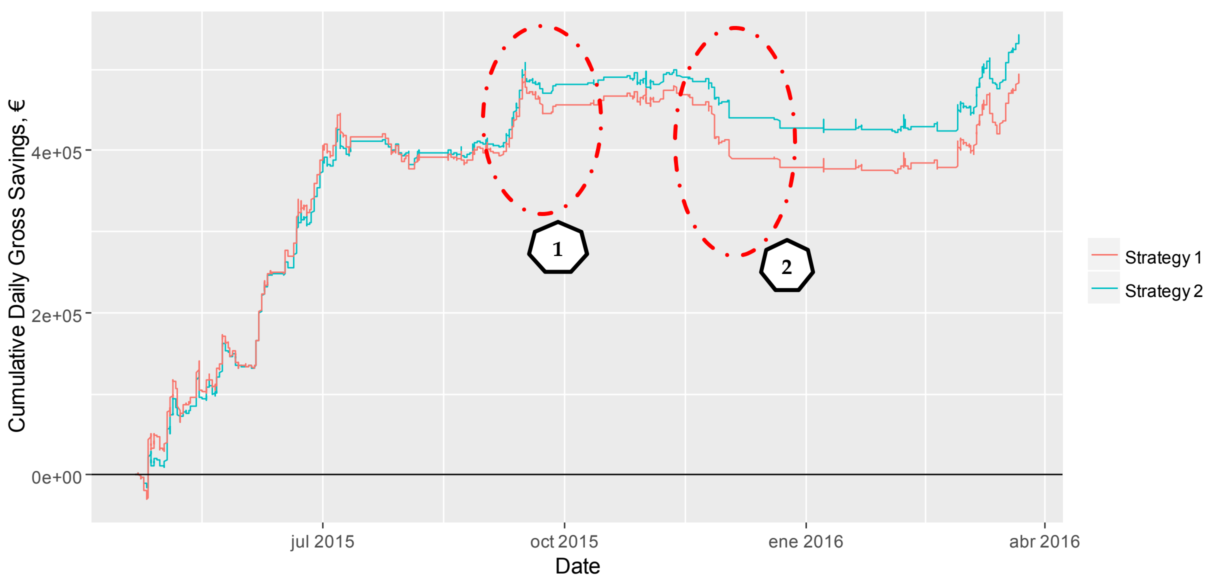


Figure 6.11: Evolution of Gross Savings comparison for Strategy 1 and Strategy 2.

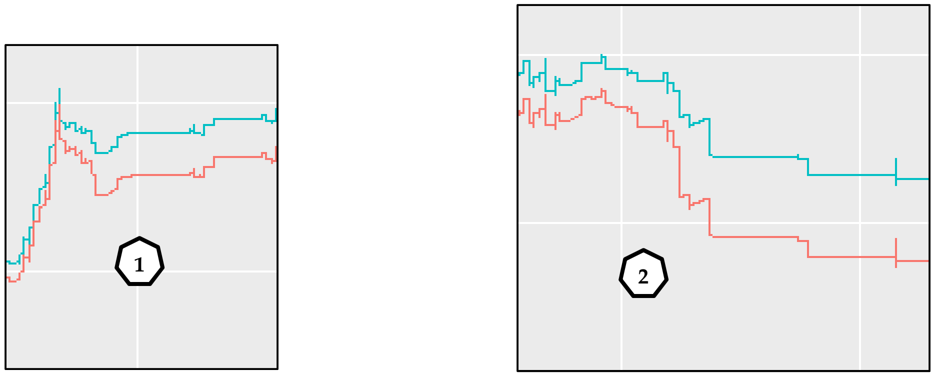


Figure 6.12: Loss reduction details for peaks 1 and 2 of profit and loss evolution.

6.6 OBSERVATIONS

From review of the simulation results, the following observations were inferred:

- Sufficiently significant imbalance cost reductions can be achieved by applying the bidding strategy developed in this project, even when using all the forecasting horizons within the scope. Considering both 1- the potentially large magnitude of the imbalance costs for certain market participants (estimated over 10€ million for the Trader in 2015) and 2- the small margins typically associated with pure trading and/or retailing activities, the potential savings of 8.71% in portfolio imbalance costs for the Trader in the case-study are sufficiently significant.
- The post-day ahead models results consistently outperformed the day-ahead model, doing so by a very large margin for the main strategy evaluation metric. When compared to the day-ahead models, the post-day ahead models achieved:
    - o Volume Error decrease of up to 9.2%
    - o Accuracy in forecasting imbalance direction increase up to 5%
    - o Correlation increase by up to 43%
    - o *Gross savings increase to 51% and 89%, depending on the strategy applied*
- The effect of transacting in the intra-day markets negatively impacts the savings, especially when transacting in the second intra-day market session. However, the imbalance ***cost reduction achieved with the strategy still outweighs the cost of transacting in the intraday- markets***.
- In terms of forecasting results:
    - o The variable importance metric appears to provide useful insight into the importance of the variables in forecasting the system imbalance.

        - All of the higher performing models included the higher ranked variable groups (in ranking order):
            1. Meteorological (wind and temperature related, except WindRAMP and percenWF) variables,
            2. Demand and network constraints (energy only) related variables.
            3. Price related variables (day ahead and constraints)
        - Omission of the lower ranked variables (balancing services and WindRAMP), especially the additional upwards reserve variable, had the least negative impact on the forecasting error.
        - Omission of the temperature – from top ranked group variable - has the greatest negative impact on the accuracy of the forecast relative to the other variable combinations.
    - Increasing the training horizon consistently decreased the error in forecasting the imbalance *volume.* However, the shorter training horizons, resulted in increased accuracy of
    - The accuracy of forecasting the imbalance direction increases as the magnitude of the forecast volume increases, which supports the bidding strategy approach.
- <u>Excluding the additional upwards reserve variable (RPAS) consistently provides better results</u>. Based on both the main evaluation metric – gross savings – and from a holistic point of view considering forecasting and secondary performance measures, Model 12 (omits RPAS) provides the higher and more consistent performance. Considering intraday market effects, this model ranks 2nd with net savings 12% lower than model 16. However, model 16, which uses only the top 4 variables ranked near the bottom in terms of forecasting performance. It is worth noting, however, that variance in forecasting performance measures between the two models may be viewed as small, ranging from 1.4 to 3.7%.
- <u>The modified bidding strategy, which reduces the bid volume following a day of losses, increases the both the gross and net savings.</u>

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

7.1 CONCLUSIONS

Portfolio imbalance penalties can represent a significant cost for market participants. This Master Thesis is focused on developing a bidding strategy that minimizes the imbalance costs of an energy trader and/or retailer´s portfolio. Although the strategy was developed with the latter agents in mind, the strategy can be applied by any market participant. We have developed an artificial-intelligence based model that takes into account information available after gate-closure of the Spanish day-ahead market to forecast the system´s energy imbalance and incorporate the forecast information into a bidding strategy that adjusts the scheduled program through intraday markets in order to influence the direction of the portfolio´s agent´s imbalance towards the direction that is not economically penalized.

For the first component, we have developed a black box model based on random forest technique to predict the hourly system imbalance. We have chosen random forest technique due to its robustness, stability, and competitive accuracy when compared to other state of the art machine learning algorithms.

For the second component we developed an optimization model based on a genetic algorithm to determine the optimal parameter values that minimize the cost for a strategic bidding function that we also defined.

Although bidding strategies to reduce imbalance costs, have been widely discussed in the literature, especially for wind power trading, none - that we are aware of - do so based on forecasting the system imbalance. The literature is very limited on system imbalance forecasting and random forests was not one of the techniques applied.

We used publicly available data as input for the model and actual data from an energy trader to evaluate the strategy´s performance. After examining the results, we can conclude that:

- It is entirely feasible to use the random forest technique to forecast system imbalances and to base a bidding strategy on those results.
- Using new information available post-day ahead market gate closure increases both the accuracy of forecasting the system imbalance and the savings derived therefrom.
- Although forecasting error measures may higher than those of some other market variables, the savings derived from applying the strategy to the forecasts are sufficiently valuable to justify its application.
- The impact of intraday markets does not outweigh or invalidate the savings potential of the strategy.
- In general, we can conclude that advanced modelling techniques are inexpensive and effective tools to forecast system imbalances and optimize bidding strategies to reduce imbalance costs.

7.2 FUTURE WORK

As long as intermittent RES-E sources continue to increase in penetration and balancing mechanisms continue to apply economic penalties on imbalances, imbalance cost reduction will remain a focus for many market participants.

Several matters of interest arose during and after the development of this thesis project. One such area deals with the actual imbalance cost. From analyzing the imbalance cost as it relates to day-ahead market prices, some interesting trends were identified which could be used to expand the strategy to consider bids conditioned on price. Also consider the intraday market prices.

Also, to lessen the impact of the intraday market on savings, forecasts of the D-1 system imbalance could be used as input to the forecast model and a pre-adjusted day-ahead bid placed. Further adjustments can be made in the intraday market by running the models with new information, as done in this thesis project.

Finally, different time horizons for optimizing the function parameters could be applied.

References

Batalla-Bejarano, J., Costa-Campi, M. T., & Trujillo-Baute, E. (2015). ***Unexpected Consequences of Liberalisation: Metering, Losses, Load Profiles and Cost Settlement in Spain´s Electricity System**.* Barcelona Insititute of Economics.

Breiman, L. (2001). ***Random Forest**.* University of California at Berkley.

Bueno-Lorenzo, M., Moreno, M., & Usaola, J. (2013). ***Analysis of the imbalance price scheme in the Spanish electricity market: A wind power test case.*** Energy Policy.

Chari, R. (2015). ***Life After Privatization*** (First Edition ed.). Oxford University Press.

Chaves-Avila, J., Hakvoort, A., & Ramos, A. (2013). ***Short-term strategies for Dutch wind power producers to reduce imbalance costs***. Energy Policy.

Cheng, Y.-Y., Chan, P., & Qiu, Z.-W. (2012). ***Random Forest based ensemble system for short term load forecasting.*** Proceedings of the 2012 International Conference on Machine Learning and Cybernetics.

CNMC (2015). ***National Report to the European Commission**.*

de la Fuente, I. (2009). A***nscillary Services in Spain: dealing with High Penetration of RES***. *Red Electrica de España.* Retrieved from http://www.reshaping-res-policy.eu/

ENTSO-E. (2016). Retrieved from ***Balancing and ancillary services markets:*** https://www.entsoe.eu/about-entso-e/market/balancing-and-ancillary-services-markets/Pages/default.aspx

Garcia, M. P., & Kirschen, D. S. (2006). ***Forecasting System Imbalance Volumes in Competitive Electricity Markets***. *IEEE Transactions on Power Systems*.

Gonzales-Aparicio, I., & Zucker, A. (2015). ***Impact of power uncertainty forecasting on the market integration of wind energy in Spain.*** *Applied Energy, 159*, 334-349.

Guyon, I., & Elisseeff, A. (2003). ***An Introduction to Variable and Feature Selection***. *Journal of Machine Learning Research*, 1157-1182.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction*** (Second ed.). Stanford, California: Springer.

Hirth, L., & Ziegenhagen, I. (2015). ***Balancing Power and Variable Renewables: Three Links***. *Renewable & Sustainable Energy Reviews* , 50, 1035-1051.

International Energy Agency. (2015). ***Energy Policies of IEA Countries: Spain, 2015 Review***. Paris: OECD/IEA.

Liaw, A., & Wiener, M. (2015, 10 06). ***Package 'randomForest': Breiman and Cutler's Random Forests for Classification and Regression***. CRAN.

Matevosyan, J., & Söder, L. (2006). ***Minimization of Imbalance Cost Trading Wind Power on the Short-Term Power Market***. *IEEE Transactions on Power Systems*, VOL. 21.

NREL. (2010). ***Operating Reserves and Wind Power Integration: An International Comparison***. 9th Annual International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants Conference.

OMIE. (2016). Retrieved 2016, from http://www.omie.es/

Pinson, P., Chevallier, C., & Kariniotakis, G. N. (2007). ***Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power.*** *IEEE Transactions on Power Systems*.

Red Electrica de España. (2016). Retrieved from http://www.ree.es

Scrucca, L. (2013). ***GA: A Package for Genetic Algorithms in R***. *Journal of Statistical Software*, Vol 53 Issue 4.

van der Veen, R. A. (2012). ***Agent-based analysis of the impact of the imbalance pricing mechanism on market***. *Energy Economics*.

Annex

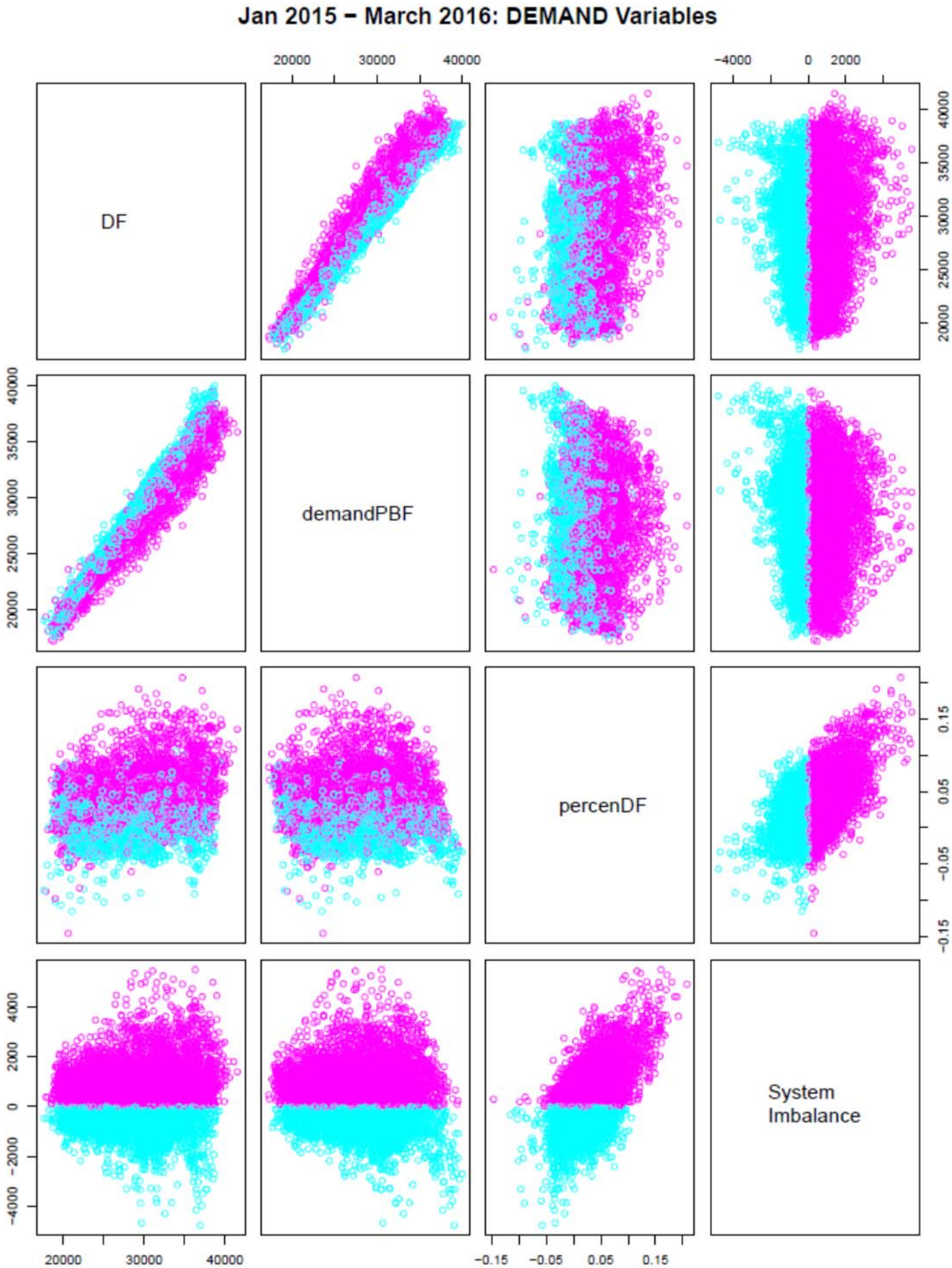
A. Cross-Correlations of all Variables

First Variable	Second Variable	Correlation
DAMP	TCPrice	0.995496533
WF	WindPBF	0.988907502
DF	DemandPBF	0.985034099
TCPrice	TCenergy	-0.702968359
WindPBF	TCPrice	-0.694363866
WF	DAMP	-0.690628723
WindPBF	DAMP	-0.690626509
WF	TCPrice	-0.690445266
DAMP	TCenergy	-0.660600232
WindPBF	TCenergy	0.452480994
DF	DAMP	0.440207768
SRBdown	SRBup	0.437909625
DemandPBF	DAMP	0.437852215
DemandPBF	TCPrice	0.422591927
DF	TCPrice	0.422269674
WF	TCenergy	0.42195802
DF	SRBdown	0.373143491
porcenDF	System Imbalance	0.37020217
DemandPBF	SRBdown	0.356358532
WF	porcenWF	0.352003922
DF	SRBup	0.347751036
DemandPBF	SRBup	0.346662585
DF	RPAS	0.299755293
DF	porcenDF	0.285290913
porcenDF	RPAS	0.28526486
DAMP	SRBup	0.263829043
TCPrice	SRBup	0.252443894
WindPBF	porcenWF	0.25238499

DemandPBF	RPAS	0.251527583
DemandPBF	TCenergy	-0.250930377
DF	TCenergy	-0.223267749
RPAS	SRBup	0.191714893
DAMP	SRBdown	0.186230683
TCPrice	SRBdown	0.176744328
porcenWF	RPAS	-0.172286062
porcenDF	SRBdown	0.170568098
porcenWF	DAMP	-0.150643933
porcenWF	TCPrice	-0.136866484
TCenergy	SRBdown	-0.133891284
porcenWF	SRBup	-0.124433054
DemandPBF	porcenDF	0.118297621
DF	porcenWF	-0.116782854
porcenDF	DAMP	0.113719702
DemandPBF	porcenWF	-0.112173738
TCenergy	SRBup	-0.107348368
porcenWF	System Imbalance	0.101509643
porcenDF	TCPrice	0.095355365
porcenDF	TCenergy	0.094810654
DAMP	RPAS	0.091960853
porcenWF	SRBdown	-0.081726348
porcenDF	SRBup	0.08167499
TCPrice	RPAS	0.079589854
RPAS	SRBdown	0.078908486
WF	SRBup	-0.075958764
System Imbalance	SRBdown	0.060403179
WindPBF	SRBup	-0.05574432
TCenergy	RPAS	0.054632141
porcenDF	porcenWF	-0.053933537
WindPBF	SRBdown	0.047211376
DF	System Imbalance	0.040038565
porcenDF	WF	-0.038361835
WindPBF	RPAS	0.032264109
System Imbalance	TCPrice	0.030559866
DemandPBF	WindPBF	0.02825709

System Imbalance	DAMP	0.027242249
DF	WindPBF	0.026346892
DemandPBF	System Imbalance	-0.02569405
System Imbalance	RPAS	0.024333969
System Imbalance	TCenergy	-0.024040173
WF	SRBdown	0.023562313
WindPBF	System Imbalance	-0.021885455
porcenDF	WindPBF	-0.013606038
WF	System Imbalance	-0.013376171
System Imbalance	SRBup	-0.009596265
DemandPBF	WF	0.003055181
DF	WF	-0.00211777
porcenWF	TCenergy	0.001946092
WF	RPAS	-0.000760361

B.1 Scatter Plot Matrix of Demand Variables



B.2 SCATTER PLOT MATRIX OF WIND VARIABLES: JAN 2016