Ph.D. Thesis dissertation

Thesis defended on the 14th of June 2018 to obtain the degree of Doctor of Engineering Science (PhD) in Electrical Engineering

by

Toubeau Jean-François

Stochastic optimization of Virtual Power Plants participating in electricity markets: from forecasting to decision-making

Supervisor: Prof. **François Vallée** (University of Mons - UMons)
Co-supervisor: Dr **Zacharie De Grève** (University of Mons - UMons)

PhD Jury

Chair: Prof. **Daniel Tuyttens** (University of Mons - UMons)
Prof. **Emmanuel De Jaeger** (Université catholique de Louvain - UCL)
Prof. **François Glineur** (Université catholique de Louvain - UCL)
Prof. **Jalal Kazempour** (Technical University of Denmark - DTU)
Prof. **Jacques Lobry** (University of Mons - UMons)
Prof. **Jean-Claude Maun** (Free University of Brussels - ULB)
Prof. **Dirk Van Hertem** (Catholic University of Leuven - KUL)

June 2018

UMONS Université de Mons

POLYTECH MONS

Content

SUMMARY

Our society is currently undergoing a major energy transition towards a more sustainable and less carbon-intensive system, characterized by an increased penetration of renewable energy sources, higher energy efficiency and reduced emissions of greenhouse gasses. In parallel, the electrification of the system (e.g. with electric vehicles and electric heat pumps) tends to make the load more dependent on weather conditions and human behaviors. Managing electricity networks under this paradigm is challenging, and the success of the energy transition thereby relies on the development of **new flexible solutions** that must **operate within a complex and uncertain environment**.

This work, which is developed in the framework of the Smartwater project, aims at evaluating the economic feasibility of the rehabilitation of old industrial infrastructures such as deep mines and quarries into medium-sized pump hydro storage stations (from one to tens of megawatt). In the context of deregulated electricity markets, the storage profitability is ensured through optimized planning strategies, which are subject to ***uncertainties*** (regarding mainly electricity prices but also load and renewable generation when the unit is integrated into a larger portfolio). This task of optimal valorization of storage units is thus articulated around **three complementary and multidisciplinary contributions**:

- **Objective 1:** Improving the current **energy models related to the operation of underground pumped-hydro energy storage (UPHES)** systems. Indeed, the operation of such technologies is significantly different from existing facilities, and is characterized by multiple nonlinear effects mainly arising from the complex geometry of the unit, and water exchanges between the porous reservoirs and their surrounding aquifers.

- **Objective 2:** Crossing the barrier between *power systems analysis* and *machine learning* (a research field specialized in learning, extracting and exploiting the complex patterns that are hidden within historical data) to provide **state-of-the-art forecasting** tools of electrical variables. Practically, this **work capitalizes on recent breakthroughs in Deep Learning** to generate more accurate multi-step ahead probabilistic forecasts, where space and time dependencies, heterogeneity and high-dimensionality are key factors.

- **Objective 3:** Integrating the output of uncertainty modeling tools into a form that is suitable to properly feed and guide the subsequent optimization. The goal is to obtain **a risk-aware integrated forecast-driven strategy** that is able to fully valorize flexible resources.

To achieve those objectives, the manuscript is divided into 7 chapters. The methodologies and main findings associated with each chapter is provided hereunder.

Chapter 1 describes the context and motivation of the work, with a particular emphasis on the **objectives and contributions**.

Chapter 2 then presents a general **overview of the current (liberalized) organization of the electricity sector**, with an emphasis on the Belgian situation. This information will serve as a basis to accurately represent the market design in the subsequent optimization models.

Chapter 3 focuses on the theoretical background behind deep learning, and we exploit this powerful framework to improve the **probabilistic forecasting of electrical variables**, i.e. the aggregated load, wind and photovoltaic power, and electricity prices. Practically, the developments are based on the use of enhanced neural networks, similar to those at the origin of breakthrough products such as Google Translate. An extensive benchmark shows the effectiveness of the method in comparison with state-of-the-art model in terms of both statistical performance and impact on the quality of decisions optimized by an agent participating in electricity markets.

Chapter 4 investigates the potential of integrating storage units into larger portfolios in order to fully leverage the available flexibility. In that respect, we show that **the portfolio effect (aggregation of technologies) results in a more efficient use of assets due to complementarities between the different resources**. In particular, it is shown that a dynamic allocation of reserves (i.e. when the contribution of each unit can vary over time) fosters the participation in ancillary services, which results in higher economic value of the portfolio.

Chapter 5 provides a framework to **integrate the nonlinear effects of an underground pumped-hydro energy storage (UPHES) system within its day-ahead scheduling in energy and reserve markets**. To that end, a hybrid approach combining an optimization tool with an advanced simulation model is developed. The results from a Belgian case study demonstrate that accurately considering these nonlinear effects is a key component to fully extract the potential of UPHES, and suggest that the proposed tool offers an effective solution for achieving this goal.

Chapter 6 leverages **surrogate-based optimization to jointly consider the tactical (week-ahead) and operational (day-ahead & real-time) decision levels**. The objective is to efficiently cope with the conflicting objectives between these time horizons (e.g. a higher profit in mid-term reduces the short-term possibilities). The procedure is applied for a typical Belgian retailer, and allows to secure feasible and efficient solutions (compared to a benchmark model) while being computationally efficient.

Finally, Chapter 7 restates the main methodological contributions, and summarizes the findings. Besides, perspectives and recommendations for future research are provided.

Overall, the **developed models have a wide scope and can be of interest for different Power Systems actors: (i) market players** who want to maximize their profit made from their generation/demand/storage portfolio, **(ii) system operators** who, as are market facilitators, aim at providing accurate and transparent (probabilistic) forecast information to the electricity sector, and **(iii) policy makers** to estimate the available flexibility and potential of new storage resources.

CHAPTER 1

INTRODUCTION – MOTIVATION, BACKGROUND, OBJECTIVES AND CONTRIBUTIONS

1.1 Context and motivation

In order to reduce significantly the European ecological footprint with a long-term perspective, the European Union (EU) leaders have established in March 2007 three ambitious objectives for 2020 that are known as the “20-20-20” targets. Setting such targets is indeed an important and functional policy mechanism for achieving specific goals. The first objective was to reduce by 20% the EU greenhouse gas (GHG) emissions (including CO_2) from the 1990 levels. This emissions reduction target is separated into two contributions: a single European-wide target for large industrial installations, and one target covering emissions from households, buildings and smaller industrial applications (such as transport, agricultures and services), which is decomposed into national targets for all 28 Member States. The other two European targets were to raise to 20% the proportion of the total energy consumption coming from renewable resources, and increasing by 20% the energy efficiency.

The European Union is well on its way to beat the 2020 target of reducing the GHG emissions, possibly even reaching -30% (with respect to 1990 levels). However, several countries seem unable to reach their objectives, not only regarding their emissions but also for the integration of renewable generation and the improvement of energy efficiency [ECR[18]]. Moreover, the current trend should be put into perspective with regard to the nuclear phase-out investigated in several countries. Indeed, as exemplified in Germany [CEW[17]], replacing this low-carbon technology leads to a transition period that can hinder the reduction of GHG emissions.

The conclusions for Belgium are equally contrasted. As represented in Figure 1.1, the Belgian target for the reduction of GHG emissions (i.e. 15% less compared to 2005 levels) was already fulfilled in 2011, and has been further improved since then. Nevertheless, the uncertain situation regarding a nuclear phase-out (representing currently 40% of the Belgium installed power capacity [Elia[17,a]]) prevents having a clear vision for the future. Moreover, the current trend indicates that the objective to cover 13% of the total consumption in 2020 with renewable energy is not expected to be reached (Figure 1.2), and the objective of 20% is even deferred to 2030. Finally, Belgium is not on track to attain the goal to reduce the total primary energy consumption to 43.7 Million tons of oil equivalent (Mtoe). It can be observed in Figure 1.3 that such energy savings are more complicated to control (more volatile) due to the strong correlation with the economic situation.

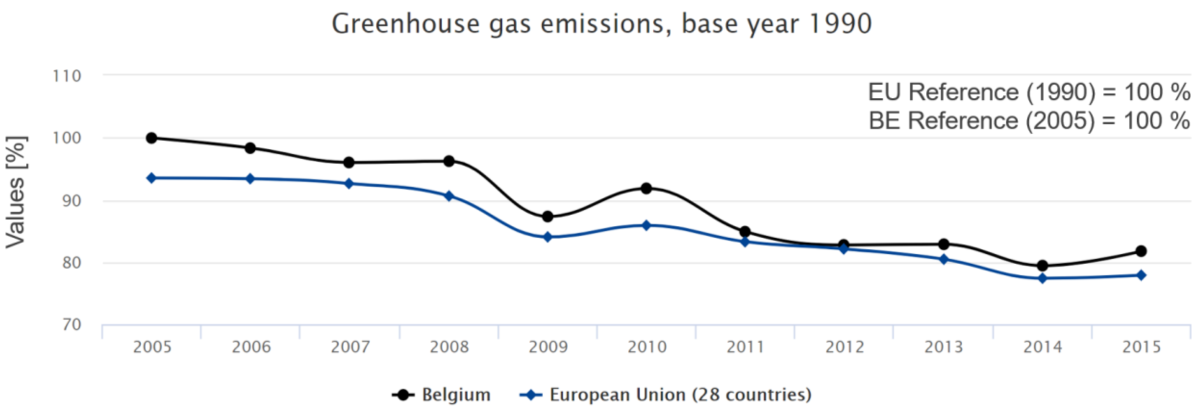


Figure 1.1 – Greenhouse gas emissions [ECR[18]].

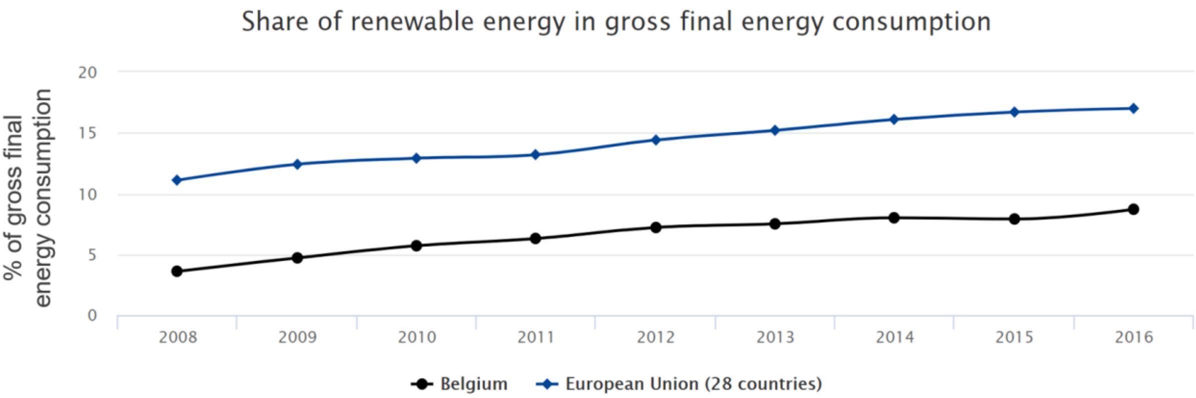


Figure 1.2 – Share of renewable energy [ECR[18]].

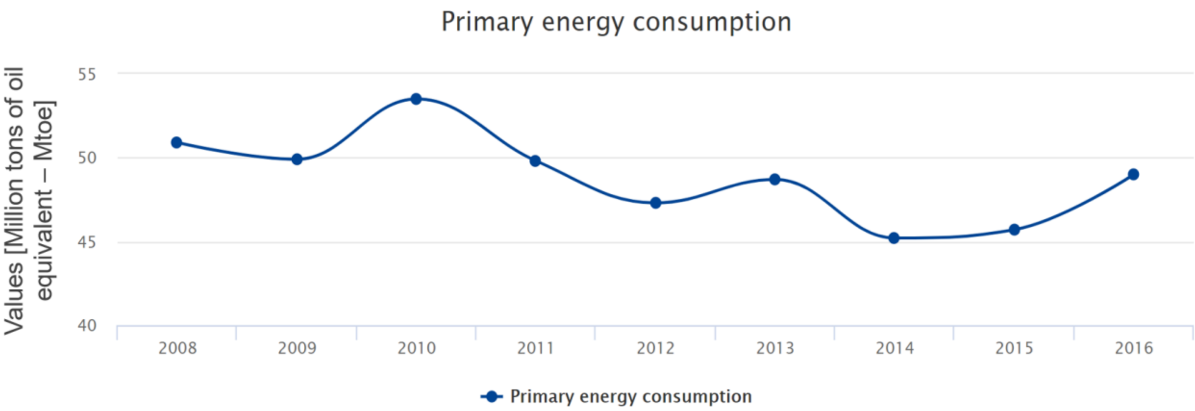


Figure 1.3 – Energy efficiency [ECR[18]].

In October 2014, EU leaders have built, on basis of 2020 objectives, the *2030 Climate and Energy Targets*. These goals can be summarized as follows [EUCO[14]]:

- Reduction of 40% of the greenhouse gas emissions (compared to 1990 levels);
- Increase the share of renewable generation to 27%;
- Increase the energy savings to 27%.

These targets are nonetheless contested and deemed as not compatible with the goals set within the Paris Agreement (12 December 2015) to limit the temperature rise to 1.5°C [United Nations[15]]. In this respect, environmental groups (Climate Action Network, Greenpeace and World Wildlife Fund) have called for stronger targets to be applied urgently, i.e. 55% reduction of GHG emissions, 45% of renewable generation and 40% of energy savings [EPRS[14]]. Overall, these considerations fall within a larger framework to ensure that **Europe can achieve a full decarbonisation by 2050**, and it is essential that the roadmap towards this central unifying objective is smartly balanced.

It should be noted that these targets encompass all energy sectors. In this way, the overall share of renewable energy in the total energy mix includes not only generation of electricity (from wind, photovoltaic, hydro, etc.) but also heat from renewable energy sources (solar and geothermal heating, etc.) and use of biofuels for the transport sector. Currently, the share of electricity in final energy consumption is around 20%, whereas 50% are associated with heat consumption and 30% comes from transportation [FROnT[18], Schäfer[05]]. However, in Belgium, the electric consumption is projected to increase by 50% with respect to 2018 levels [Elia[17,b]] due to electrification in transportation (expected development of electric vehicles after 2020) as well as in the heating and cooling sectors (expected development of electric heat pumps after 2030), which will make the load more dependent on weather conditions and human behaviors.

In this dissertation, the focus will be given to the electrical sector whose main challenge in the following years will be to efficiently accommodate renewable generation sources (to ensure the low-carbon transformation), while guaranteeing the security of supply and keeping fair and competitive prices for end-users.

1.1.1 Renewables in power systems: challenges

The introduction of this renewable generation in power systems, bolstered by their technological maturity and political willingness to promote a low carbon society, is not without consequences. In this way, the increasing contribution of intermittent power is progressively redesigning the historical structure of the electricity sector, arising stability issues for both transmission and distribution systems, which are here summarized in four aspects.

Firstly, **renewable sources, such as solar and wind, are highly volatile and partially unpredictable**. This strongly complicates both the long-term planning (e.g. grid expansion plan, determination of the optimal energy mix, etc.) and the operational control of power systems due to the nature of electrical energy. Indeed, a continuous equilibrium (balance) between the total generation and consumption needs to be respected for ensuring the frequency stability of the electrical grid (thereby preventing involuntary load shedding and blackouts). The stochastic behavior of the renewable generation (in addition to the load uncertainty) leads to growing needs of **flexibility** (i.e. ability of generation/consumption/storage resources to adjust their output power when required).

Secondly, **this energy transition takes place in a liberalized environment**, and, in order to foster the integration of renewables (which are still not competitive financially with conventional technologies), financial levers are implemented by national (political) authorities. For instance, Belgium has set up a mechanism of green certificates in order to pay the energy generated by systems emitting little or no CO_2. Such financial incentives granted to the renewable generation for ensuring their profitability has biased the standard law of supply and demand, and has progressively driven down the electricity prices to the point of dropping below the profitability threshold of some conventional plants (e.g. open cycle and combined cycle gas turbines). This situation led to the progressive mothballing or even closure (and dismantling) of these power generators, which used to assume fundamental roles for maintaining the system stability. Indeed, such large units significantly contribute to the grid inertia (frequency stability) thanks to the large rotating machines alleviating the sudden frequency deviations. Then, they also enhance the network resilience to perturbations (e.g. transient effects during the start-up of industrial processes) by increasing the short-circuit power throughout the grid. Finally, conventional units are generally highly dispatchable[1], thereby constituting the main source of operational flexibility in traditional power systems to alleviate both frequency and voltage deviations. Renewables have thus a twofold conflicting impact on the flexibility means within the system (they increase the need of adjustable resources due to their volatile nature, while reducing the existing inertia and stability resources by replacing flexible conventional generators in the energy mix). Hence, a more flexible use of renewable generation will need to be developed, which can be achieved with new control strategies of the power electronics converters located at the interface between the renewable generator and the electrical network. Up until now, these converters are mostly used to continuously extract the maximum amount of power (maximum power point tracking (MPPT) control) based on weather conditions, but these power converters can offer a larger panel of opportunities to improve the grid operation (synthetic inertia, bidirectional (upward and downward) reserves if they are operated below their maximum power point, voltage control, congestion management and power quality improvement) [Clastres[10], Renner[18], Van de Vyver[14]].

Thirdly, most renewable technologies are characterized by an installed power varying from several kilowatt (for photovoltaic residential installations) up to a few megawatt (for wind farms). It is therefore not technically suitable to connect these sources to the transmission grid (at the exception of the largest wind parks). These **renewable-based generation systems are thus erratically installed throughout the distribution network in a fully decentralized way**. Originally, distribution systems were not designed to host such an amount of generation since the power generation was then exclusively realized by large centralized power plants (e.g. nuclear and thermal units). The power was then transported along the transmission grid over the country with different connection points with the local distribution systems. The latter were thus dimensioned with the single objective of ensuring a secure and efficient supply of electricity to the different end-users. Modern distribution networks are consequently subject to new difficulties. Contrarily to the transmission systems that are equipped with metering devices and state estimation tools for identifying in real time or even anticipating the potential problems as well as different mechanisms to quickly solve such issues, the distribution systems are not equipped with such technologies. These could thus suffer to overcome voltage violations and line congestion problems (both associated with detrimental effects on the electrical equipment) that occur more frequently, although over limited periods of time.

Fourthly, **due to the intermittent nature of renewable energy sources, back-up solutions are needed to cover the energy demand at times of low availability of renewables**. The energy transition therefore necessitates to carefully design the future energy mix (along with the network infrastructure) so as to ensure both the long-term system adequacy, i.e. sufficient generation capacity to cover peak demand, and the efficient management of generation surplus (to avoid curtailment of renewable generation), all without oversizing the global system.

[1] Dispatchable power plants can adapt their output power in accordance with an external signal.

1.1.2 Solutions for a better integration of renewable energies

Overall, the energy transition is accompanied by a great need of new additional sources of flexibility, and it is indispensable to find alternatives less expensive than simply reinforcing the existing infrastructure. In this way, the efficient and reliable operation of future power systems (integrating a large share of renewable energies) can be achieved by a combination of different solutions.

The first one is to increase the interconnection capacity between countries in order to create a global (European) system more resilient and secure. In addition to reliability considerations, creating a more interconnected grid has also positive economic effects by mitigating the electricity price volatility across countries, while minimizing the need of aggregated flexibility by taking advantages of opposite effects between neighboring areas. This expansion of the cross-border capacity is actually one of the solutions chosen by Elia, the Belgian transmission system operator (TSO) through HVDC tie lines with Great-Britain (i.e. NEMO project) and Germany (i.e. ALEGRO project).

Then, it is important to diversify the types of technologies in order to minimize the positive dependency between the energy produced by the different sites. For instance, the energy generated by photovoltaic (PV) panels can adequately complement wind energy since the wind speed tends to increase in winter and during the night when there is no PV generation (negative dependency).

Another possibility is to encourage the implementation of more active networks whose principle is to adapt the consumption to the generation (i.e. demand-side management). This involves increasing the involvement of end-users, then referred as *prosumactors*[2], by exploiting their deferrable loads with the implementation of new dynamic pricing schemes that allow them (typically though cooperatives or aggregators) to adopt a behavior in line with the network needs. Such mechanisms necessitate the prior installation of metering devices recording the energy exchanges with an appropriate time resolution (such as 15 minutes) in order to adequately associate the dynamic prices to the actual period of consumption and/or generation. In this context, higher penetration of electric vehicles and electric heat pumps can potentially lead to greater demand peaks, but (with adequate incentives that do not lead to uncontrolled synchronization effects) may also provide services to the electricity grid by reasonably charging up in case of electricity surplus and feeding electricity back into the grid in case of scarcity.

Aforementioned solutions may prove to be expensive and/or slow to deploy in the current context, and overall insufficient to tackle future challenges. In that context, **the integration of electricity storage may bring an interesting contribution to compensate the lack of available flexibility**. Different technologies (pumped storage hydro, batteries, compressed air, etc.) can be investigated, varying along three dimensions, i.e. two pertaining to time (response speed and storage horizon), and the energy capacity. First, different dynamics (speed at which the stored energy is accessible) are required for solving not only transient issues resulting from the temporal discrepancies between generation and consumption, but also the more structural problems (such as electricity shortages). Then, one should carefully consider the time horizon during which the energy can be stored without significant losses (e.g. ultra-short-term storage like supercapacitors, daily storage as in batteries, or inter-seasonal storage

[2] *prosumactors* is a contraction of prosumers (end-users having their own generation asset), and actors (end-users who are actively contributing to the system through an energy management system).

as in large pumped storage hydro stations). Finally, different storage sizes are useful to adequately address the needs of the different actors of the electricity system (from small devices offering local solutions for, e.g., micro-grids or peer-to-peer energy trading[3], up to centralized storage utilities designed to efficiently participate in wholesale markets). In complement to these electricity-to-electricity solutions, promising technologies such as power-to-gas or power-to-heat also are also offering interesting alternatives.

An appropriate option regarding storage technologies is offered by **pumped storage hydropower (PSH) plants** due to their ability to quickly and cost-effectively respond to a mismatch between generation and consumption. These stations can indeed store a large amount of energy with low operating costs. Besides, recent progress in power electronics have enabled PSH units to operate with a reliable variable-speed feature in both pump and turbine modes, consequently fostering their ability to adjust their output power at the request of the plant owner. This favorable environment is currently leading to the development of new technologies such as underground PSH units [Alvarado[15], Pujades[17]], in which the lower reservoir is located into the ground taking for instance profit of end-of-life mines or quarries that are exploited as natural basins for saving civil engineering expenses. These stations have indeed very limited impacts on landscape, vegetation and wildlife, and are not limited by topography so that more sites can be exploited [Alvarado[15]]. Such a solution is therefore fully investigated in the Walloon Region through the *Smartwater* project that aims at evaluating the feasibility of the rehabilitation of old industrial infrastructures into small to medium-sized PSH stations (from one to tens of megawatt) connected to the distribution network [Smartwater[18]]. **The present PhD thesis, which has been prepared in the framework of the *Smartwater* project, contributes to provide an answer to this question, by quantifying the profitability of small-to-medium PSH stations, as it will be further detailed in the next section.**

1.1.3 Integration of Pumped Storage Hydro units: economic challenges

Currently, the main component of the profitability of PSH is linked to the valorization of flexibility, i.e. typically by participating in electricity markets related to the transmission grid. Indeed, in a liberalized environment where generation and retail are decoupled from the transmission and distribution of electricity, the system operator generally does not have its own resources. The efficient operation of the system is therefore at a large extent apportioned to market participants. In particular, the latter are financially incentivized (by the market design) to improve their ability to address the different sources of uncertainty within their portfolio for securing the success (optimal profitability) of their operational planning strategies. It is thus of general interest to improve portfolio management of electricity market participants, especially as it can also contribute to the emergence of new actors investing in renewable energies. They can indeed rely on robust tools for managing risk in electricity markets, which will overall accelerate the energy transition. In that regard, PSH units represent promising solutions for market participants to smooth the inherent uncertainties of load and renewable generation.

However, these small-to-medium PSH units are connected to distribution systems, and no electricity market designed to address the challenges arising within the distribution system is available (but seems vital for solving the issues progressively emerging in an already ageing network). But, even in the favorable case of an adequate market implementation in distribution

[3] Peer-to-peer (p2p) is a trading mechanism that allows consumers and producers to directly make deals, typically through a trading platform, on their own terms (regardless of their size), without a middleman (retailer). The users can set their own terms regarding the price, the energy source, etc.

systems, the natural location of the PSH unit conditions the contribution to local services such as congestions or voltage issues (no need of flexibility if there is no problem in the area). The location may also prevent the site to be coupled with an industrial company (that could exploit the flexibility from storage to improve its energy profile so as to be more adapted to the system needs). Overall, these potential additional revenues for ancillary services are not investigated in this thesis.

Even in the context of providing flexibility to the transmission system, operating a storage unit alone is not optimal due to their limited energy capacity, which prevents them to provide energy for a sufficiently long period of time (typically limited to a few hours) [Al-Awami[11], Archer[07], Castronuovo[04], García-González[08]]. In this way, **any flexible unit offers a real added value when it is included within an existing generation/consumption portfolio**. Within this trend, electricity markets are observing the increasing development of virtual power plants (VPPs), i.e. aggregation of assets from different technologies (gathering capacities from consumers, generation, and storage) that are jointly co-optimized as a single entity in a multi-market environment with the objective to maximize their expected profit. This solution entails several assets by taking advantage of the specificities of different units. For instance, most thermal plants often have fast-ramping capabilities, which can be useful for coordinating wind generation [García-González[08]]. However, such units are restrained by minimum up/down times and are thus exposed to the risk of operating at low-profitability or even at loss during some periods in order to take advantage of temporary high prices. It can thus turn out to be profitable to wipe out their contribution at convenient times, which can be efficiently achieved, for instance, by storage units. Moreover, the mixing of several technologies also contributes to reduce the dependence on one form of energy, which can potentially reduce the global prediction error. Then, the aggregation of several units mitigates the risk due to contingencies such as the loss of a generating unit, by decreasing the volatility of the expected profit over time. Finally, it allows VPP to benefit from a pool of flexibility for taking advantage of furtive extreme prices and participating in the potentially more lucrative ancillary services.

Overall, managing the portfolio of market players (such as virtual power plants) under the liberalized framework is hard, and gives rise to challenging optimization problems, with the following characteristics:

- ***they have a multi-stage structure*** (corresponding to the sequential clearing of electricity markets at different time horizons);
- ***they are dynamic*** (in the sense that they optimize over a given future time horizon);
- ***they have a mix of integer and continuous decision variables*** (to adequately model the technical processes of the different assets);
- ***they must be solved under uncertainties*** (regarding mainly electricity prices as well as load and renewable generation);
- ***they lie at the frontier of game theory*** (due to the competitive framework in which the actions of a player can influence the market clearing).

1.2 Objectives

In light of this exciting environment, this thesis aims at developing mathematical tools dedicated to the improvement of the scheduling strategies of market players (virtual power plants). Specifically, to adequately account for the relevant revenue streams of actors with flexible resources, the work will encompass both medium-term (week-ahead) and short-term (day-ahead) perspectives. Indeed, in the current regulatory framework, part of the flexibility is acquired by the system operator in mid-term, and this decision stage is therefore essential to ensure the optimal management of the portfolio.

Overall, our contributions relate to two main research areas:

(1) the uncertainty modeling in order to properly represent the future state of the stochastic decision environment.

The objective is to cross the barrier between *power systems analysis* and *machine learning* (a research field specialized in learning, extracting and exploiting the complex patterns that are hidden within data) so as to provide state-of-the-art predictive tools. Market players must indeed operate within a complex, uncertain environment, and consequently need to rely on accurate **multivariate and multi-step ahead probabilistic predictions** that adequately quantify the level of uncertainty associated with each variable over the prediction horizon.

Regarding the day-ahead perspective, the purpose is to directly generate forecasts with the highest degree of precision (reduction of the uncertainty space to facilitate the task of the subsequent optimization tool). However, for the medium-term horizon (week-ahead or month-ahead), the accuracy of the forecasting models can be questionable (especially for volatile variables such as renewable generation), and the objective will rather be to provide a small number of **time trajectories (scenarios), representative of the statistical behavior of the available historical dataset**.

The scientific contributions regarding this part dedicated to *data analytics* are:

- J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, "Deep Learning-based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets," in *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203-1215, March 2019.

- J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, "Improved Day-Ahead Predictions of Load and Renewable Generation by Optimally Exploiting Multi-Scale Dependencies," in IEEE Innovative Smart Grid Technologies, Auckland, New-Zealand, 2017.

- J. Bottieau, F. Vallée, Z. De Grève and J.-F. Toubeau, "Leveraging Provision of Frequency Regulation Services from Wind Generation by Improving Day-Ahead Predictions using LSTM Neural Networks," in IEEE Energycon, Limassol, Chyprus, 2018.

(2) the decision-making (optimization) procedure under uncertainty.

The objective is to capitalize on the impressive amount of work recently realized in stochastic optimization, as well as its application on bidding strategies in electricity markets [Amin Tajeddini[14], Conejo[02,b], Kazempour[15], Mashhour[11], Pandzic[13,b]]. In this way, as a first step, a **generic procedure for the stochastic day-ahead scheduling (economic valorization with an objective of profit maximization) of virtual power plants with diverse technologies** (conventional and renewable generation, storage units as well as controllable loads) is developed. The complexity of the resulting problem (joint participation in energy and ancillary services markets) necessitates to appropriately define the modeling equations (with unavoidable simplifications in the design of the complex properties related to the different utilities) in order to obtain a compact (and thus tractable) mathematical formulation that can be reliably solved.

From this starting formulation of reference, the purpose is then to **integrate all relevant nonlinear characteristics of pump-storage hydro stations with a high time resolution within a computationally efficient approach**. Indeed, the operation of these small to medium-sized units is governed by multiple nonlinearities arising from turbine and pump performance curves, head effects[4] as well as groundwater exchanges between reservoirs and their surrounding aquifers. Accurately considering these nonlinear effects is in this way a key component to extract the full economic potential of these underground stations.

Then, benefiting from the knowledge developed when implementing the short-term decision tool, the mid-term horizon is next considered. In that case, the tractability of the problem is strongly jeopardized due to the addition of a supplementary decision stage, and requires to **couple time horizons with possible conflicting objectives**. Indeed, the mid-term decisions infer constraints on the short-term management through the obligation to uphold these longer-term commitments (e.g. the reserves that were contracted in mid-term must be provided in real-time when requested by the system operator) and disregarding this dependence may lead to suboptimal or even unfeasible solutions.

The scientific contributions regarding this part dedicated to *optimization in electrical markets* are:

- J. F. Toubeau, Z. De Grève and F. Vallée, “Medium-Term Multimarket Optimization for Virtual Power Plants: a Stochastic-Based Decision Environment,” in *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1399-1410, March 2018.

- J.-F. Toubeau, S. Iassinovski, E. Jean, J.-Y. Parfait, J. Bottieau, Z. De Grève, and F. Vallée, “A Nonlinear Hybrid Approach for the Scheduling of Merchant Underground Pumped Hydro Energy Storage,” in *IET Generation, Transmission & Distribution*, in press.

- J.-F. Toubeau, Z. De Grève, P. Goderniaux, F. Vallée and K. Bruninx, "Chance-Constrained Scheduling of Underground Pumped Hydro Energy Storage in Presence of Model Uncertainties," in *IEEE Transactions on Sustainable Energy*, in press.

- J.-F. Toubeau, Z. De Grève, F. Vallée, “Technical Impacts on Distribution Systems of Medium-Sized Storage Plants Participating in Energy and Power Reserve Markets,” in 24th International Conference & Exhibition on Electricity Distribution, CIRED 2017, Glasgow, Scotland.

It should be noted that such (mid to short-term) scheduling procedures can also be used to interact with long-term studies (planning stage). In particular, once the sizing of a unit (regardless of its technology) is theoretically optimized (with the required assumptions to obtain a tractable methodology), it is interesting to take these constraints into account in a more sophisticated operational strategy. In this way, a feedback from the shorter term operation can bring valuable information (regarding realistic scheduling decisions), and the sizing of the unit

[4] The head effect is the height variation between water levels within the reservoirs, which gives rise to significant impacts on the operation of PSH stations.

can be adjusted if necessary. However, this relationship between planning and operational levels goes beyond the scope of this work.

As tacitly underlined in the description of both fields of study, the tools are developed with the willingness to exploit with a general vision the complementarities between the way the uncertainty is characterized, and how it is integrated in the stochastic decision procedure.

1.3 Outline and scientific contributions

Chapter 2 aims at providing a general overview of the current organization of the electricity sector, with an emphasis on the Belgian situation. The objective is not to give a deep description of the different mechanisms (such as the mathematical subtleties of market clearing algorithms or the enumeration of the procedures faced by large power plants when generating power), but rather to inform on the different consequences and challenges arising from the liberalized market structure. The scope of the description is voluntarily wider than the specific context associated with the market-related optimization tools developed in the following chapters. Indeed, the objective is to provide a larger vision of the current situation, by coupling the perspectives of both electrical and politico-economic worlds. The knowledge of these technical and regulation considerations is indeed indispensable to be fully aware of the issues related to the massive penetration of renewable energies or the potential repercussions of the integration of new solutions such as peer-to-peer energy trading or micro-grids.

Our main contribution with regard to this chapter is the gathering of information concerning past, current and future evolutions of the market rules, and to condense it in a more structured, simplified document.

Chapter 3 assembles our work in the context of day-ahead probabilistic forecasting in a multivariate environment with heterogeneous data of different nature. Indeed, in the current competition framework governing the electricity sector, complex dependencies exist between electrical and market data. The objective is therefore to provide forecasts under the form of intervals or densities (that can thereafter be used in stochastic optimization frameworks such as robust [Sun17], interval [Yu05] or chance-constrained [Wu14] approaches), but also under the form of time-dependent trajectories for scenario-based stochastic programming [Conejo10]. To that end, this **work capitalizes on recent breakthroughs in Deep Learning** (which are based on the use of neural networks with an improved memory management, similar to those exploited by major technology companies for products such as Google Translate or the speech recognition applications in smartphones) to generate more accurate multi-step ahead forecasts. Then, a copula-based sampling strategy (from the predictive densities) is implemented to obtain scenarios that embody both temporal information of individual variables (e.g. autocorrelation structure, regime switching, etc.) and cross-variable dependencies (statistical relationships between uncertain variables).

The main contribution of this chapter is to exploit and adapt new recurrent neural networks architectures with rich dynamics to increase the predictive capability of (both deterministic and probabilistic) forecasts. Then, two different models for characterizing the prediction uncertainty are compared, i.e. a parametric method assuming a Gaussian assumption of errors, and a non-parametric approach (that makes no assumption on the underlying probability distribution of variables). The value of the forecasts is then compared with other neural network approaches (all implemented ourselves in the same software) not only in terms of statistical performance, but also regarding the practical impact of the quality of decisions optimized within a stochastic optimization tool (dedicated to the day-ahead multi-market scheduling of electricity aggregators).

Chapter 4 presents the scenario-based stochastic optimization framework on which relies the day-ahead decision-making of a virtual power plant participating in both energy and ancillary services markets (whose structures are previously described in chapter 2). The main objective of the formulation is to properly account for all sources of uncertainties, in particular the time-varying amount of energy that will be called on for ancillary services. The purpose is indeed to overcome limitations faced by resources such as storage units and demand response strategies that cannot guarantee the provision of services for long periods due to their limited energy capacity.

The principal contribution is the joint inclusion of both technical and economic effects arising from the uncertain real-time activation of allocated reserves. Specifically, the formulation takes into account both revenues from the actual provision of reserves and the variable cost structure of all considered technologies. This allows to obtain a cost-optimal allocation of assets to the different ancillary services over the scheduling horizon.

Chapter 5 builds on the previous scenario-based stochastic formulation to **properly model the nonlinear effects associated with pumped storage hydro units (head dependencies, groundwater exchanges) that cannot be easily modeled analytically (and even less easily solved)**. This computational problem is addressed using a hybrid approach combining an optimization tool with an advanced simulation model of hydro plants so as to adequately account for intricate dependencies among hydrogeological and electrical parameters within the decision procedure. Thanks to the knowledge accumulated within the *Smartwater* project, simulation model of PSH units takes as inputs realistic models coming from the worlds of electro-mechanics (for operation of hydraulic and electrical machines) and hydro-geology (for water exchanges between reservoirs and surrounding aquifers).

Compared to the existing literature, the proposed hybrid formulation takes into full consideration all nonlinearities inherent in the operation of PSH stations within a computationally efficient environment. The principle of this hybrid approach can be extended to easily integrate other sources of nonlinearity (e.g. state-space model of the thermal behavior of buildings with heat pumps supplying operational flexibility to the grid) without significantly affecting the simulation time since different simulators can be run in parallel.

Chapter 6 focuses on the medium-term (typically one week up to one month) optimization of a portfolio manager. At this stage, the decisions infer important constraints on the short-term management through the obligation to uphold these longer-term (tactical) commitments. Up until now, the mid-term decisions of VPPs were usually taken by making simplifying assumptions concerning the short-term operation (neglect inter-temporal constraints of units, integrate a very limited number of stochastic parameters in the formulation, etc.) in order to rely on a single mathematical tool. Here, a different vision is tested. The principle is to better account for dependencies with the short-term horizon by including in the mid-term formulation a detailed description of the underlying daily decision-making procedures. The proposed formulation is modular and flexible so as to comply with any portfolio configuration and to follow evolutions of the market regulation policy.

The main asset of the method is to jointly consider tactical and operational decision levels so as to cope with the conflicting objectives between the different time horizons. This allows taking adequate mid-term decisions based on accurate feedback coming from the short-term simulation. In order to hedge against intractability of the resulting problem regarding both time and computer memory requirements, **this work proposes to firstly learn (as a pre-processing task) the intricate relationship between mid-term decisions and the resulting profit that can be generated in short-term. Practically, this relationship is established by training a surrogate model of adequate complexity**. Then, the medium-term decision process can be solved using the pre-determined model without having to simulate the optimal short-term VPP scheduling problem (surrogate-based optimization). A second contribution of this work consists in the implementation of a new method for dealing with mid-term uncertainty in order to include a large number of stochastic variables (e.g. up to ten) into the formulation, while conserving a limited but statistically representative set of scenarios. The principle is to encompass all dependencies into the same statistical model thanks to non-parametric copulas.

Finally, chapter 7 restates the main methodological contributions, and summarizes the important findings. Additionally, some perspectives and recommendations for future research are formulated.

CHAPTER 2

GENERAL FRAMEWORK OF ELECTRICITY MARKETS – APPLICATION TO THE BELGIAN CASE

2.1 Introduction

The 4th September 1882 marks an important milestone in the history of electricity. Indeed, on this date, Thomas Edison put into operation the first power plant designed for electric lighting at Pearl Street in New-York City. Whereas it could supply (in DC current) up to 600 kW (i.e. a power then equivalent to 7200 lamps), Edison only had 400 lamps to worry about that first day. The power plant consisted in 6 dynamos of a power of 100 kW each, called Jumbo dynamos (as a reference to the elephant Jumbo, the biggest elephant in captivity at that time, which became famous worldwide after having been bought 10.000 $ from the London zoo in 1882 by the American showman P.T. Barnum in order to make it the centerpiece of its circus). Each Jumbo dynamo of Edison's power plant was directly connected to a high-speed steam engine, which represents a turning point compared to traditional systems composed of ropes or belts for transmitting power between engines and dynamos.

The technique of Edison was then quickly improved so that the generation costs were significantly lessened around 1900. Consequently, the electricity prices were reduced and numerous new applications were developed. Among others, one can cite the electric iron (1893), radiators and toasters (1895), television (1926), fridges (1930) or microwaves (1960). Moreover, as a result of the success of its exhibition in 1879, where he presented with his partner J.G. Halske the world-first electric train in which power was supplied through rails, Werner von[5] Siemens built an electric tramway of 2.5 km in Berlin at his own expense in 1881. The most visionary spirits of the time had therefore the idea of combining both generation and transportation systems, creating in this way the first electrical grids.

It is worth noting that towards the end of 19th century, the **emergence of AC current** spread throughout Europe and North America. Such a technology allowed indeed to increase

[5] Born with the name of Werner Siemens, he was ennobled in 1888 following its industrial career. He became then Werner **von** Siemens.

both the quality and efficiency of generation while facilitating transmission of energy thanks to an easy transition between voltage levels[6]. In this way, the high voltage electrical network (backbone structure of the system) is useful to transport significant amount of power while limiting the power losses (Joule effect). But, it is also necessary to have multiple decreasing voltage levels so as to efficiently integrate small and medium-sized generation technologies and safely supply end-users of varying sizes (large industrial clients to households). Then, the **AC three-phase structure** allows economizing conductors (for the same mass of copper, more energy can be transported with three-phase networks), while accommodating generation technologies (the rotating magnetic field with constant direction and magnitude simplifies the design of electrical machines).

In Belgium, the first electrical power plant was constructed in 1885 (i.e. 3 years after Edison's plant) in Brussels. The first local implementations of electrification systems for transportation as well as public lighting were observed in the following years. The first clients were municipalities that progressively began to light public buildings as well as their streets with the aim of improving the security of citizens. Convincing the residential households turned out to be more complicated and the first electrical companies resorted to promotional means that look rather surprising nowadays. Indeed, potential customers were offered the installation of the electrical equipment combined with one year of free consumption. The distribution of electric power was then totally uncoordinated at the Belgian level and municipalities entrusted to private distributors the supply of electricity to end-users.

In 1956, the company Ebes is created by merging four electrical societies. In the meantime, Intercom, which was originally founded in 1901, became one of the biggest energy company in Belgium (due to several mergers). Moreover, the merging of other energy companies led to the creation of a third private group specialized in the generation and supply of energy, Unerg. The transmission system was then highly fragmented among several participants and the system regulation was far to be optimal. In order to address this issue, and with the aim of increasing the service delivery while achieving economies of scale, the general meetings of Ebes and Intercom approved the 10th July 1990 the grouping of both energy companies into one private company, to which Unerg will gather soon afterwards. On this occasion, Ebes is renamed Electrabel.

Consequently, the structure of the electricity market was boiled down to the interaction between two close actors. The first one (i.e. Electrabel after that Ebes, Intercom and Unerg have merged) was responsible of optimally operating both transmission and distribution systems, and had to that end the control of the whole generating fleet within its control area. Notwithstanding the monopoly of the wholesale electricity prices then held by Electrabel, this situation allowed an efficient communication between the different competences (e.g. maintenance and development of the grid, real-time operation …). In this way, Electrabel could operate the grid with a long-term vision and was able to invest in the best generation units with regard to technical, financial and environmental constraints, while preserving the safe and reliable operation of the grid. Then, the second actor was in charge of the supply and distribution of electricity. This was ensured locally by the intercommunal localized in the considered area. In this way, each area was characterized by a different supplier, which had a monopoly concerning the retail electricity market. The prices could thus vary among the different zones, even within the same province.

[6] Currently, high voltage direct current (HVDC) is emerging as an alternative. Such lines have lower losses but require more expensive power electronics (in substations), and it is difficult to make connections in the middle of the line [EPRS16].

Hence, in order to break down such a monopoly (vertically integrated structure where the same company produces, transmit and sells the electricity at a single imposed price), the European Union decided by means of three legislative packages (1996, 2003 and 2009) to deregulate electricity markets in order to create an unbundled structure with more competition. The main drivers[7] for this change towards a liberalized and internal energy market were originally to ensure fair prices, develop renewable energies and improve security of supply thanks to the introduction of competition at both generation and supply levels [EPRS[16]].

The European Directives were converted by the Member States in national legislation. In Belgium, this was translated into the creation of Elia, the current transmission system operator (TSO), which was assigned with the task of managing the transmission system (through a regulated monopoly). Elia was then legally forbidden to own any generation unit and the production of electricity was opened to competition. However, Electrabel had preserved its whole generating fleet and it was rather complicated at first for new participants to compete with this company deeply rooted in the Belgian electrical energy landscape.

Overall, the liberalization of electricity sector can be summarized as follows [Brijs[17]]:

- generation and retail activities are fully decoupled from the transmission and distribution of electricity;
- introduction of competition at both generation and retail levels;
- organization of energy and ancillary services markets as tools for market players and grid operators to ensure the safe and efficient operation of the grid;
- installation of regulated monopolies for the transmission and distribution of energy;
- apparition of regulators to monitor both regulated and market-related activities.

The consequences and challenges arising from the deregulation of the electrical sector, within the context of transition towards a low-carbon economy (development of renewable energy sources, improved energy efficiency and electrification of transportation and heating sectors), constitute the main subject of this chapter. In Europe, although a growing interest in harmonizing the market rules (and to obtain convergent prices across the different countries), many areas have their own (minor) subtleties and there exists almost as many regulation policies as countries [ECR[15]]. **The objective is therefore to make the link between the Belgian case and the other European regulation mechanisms**.

First, an overall overview of the market mechanism is presented in Section 2.2 in order to introduce the different actors and interactions between them. Then, the operation of the wholesale energy markets is studied in Section 2.3, whereas Section 2.4 focuses on the functioning of the retail market. The ancillary services designed to ensure the grid frequency stability while maintaining voltage at suitable levels and preventing major line congestions are described in Section 2.5. The mechanisms incentivizing market players to maintain the system balance are introduced in Section 2.6, whereas the strategic reserve to ensure capacity of supply during winter months is described in Section 2.7. The impacts issuing from the penetration of renewable-based generation in the current framework of electricity markets are presented in Section 2.8, and the relationship between electricity and other energy markets constitutes the core of Section 2.9.

[7] Beyond the techno-economic motivation, European Union believed that economic integration goes hand in hand with political integration (so as to create a strong federal Europe and avoid the destructive wars of the past). In this way, natural gas and electricity simply followed past examples of integrated European markets such as steel and coal (with European Coal and Steel Community (ECSC) 1952) or atomic energy (with Euratom in 1957).

2.2 General overview of electricity markets

With the liberalization of the electricity sector, the system is now composed of the physical infrastructure (electricity generation, transport and utilization), and of an organized electricity market. The physical grid, as represented in Figure 2.1, is commonly subdivided into the transmission system (to carry the electrical energy generated by big power plants over long distances) and distribution systems (to source the energy to residential and industrial consumers). The electrical flows in the system cannot be guided, and follow the path of least resistance (Kirchhoff's law), so that end-users are supplied with electricity from mixed sources.

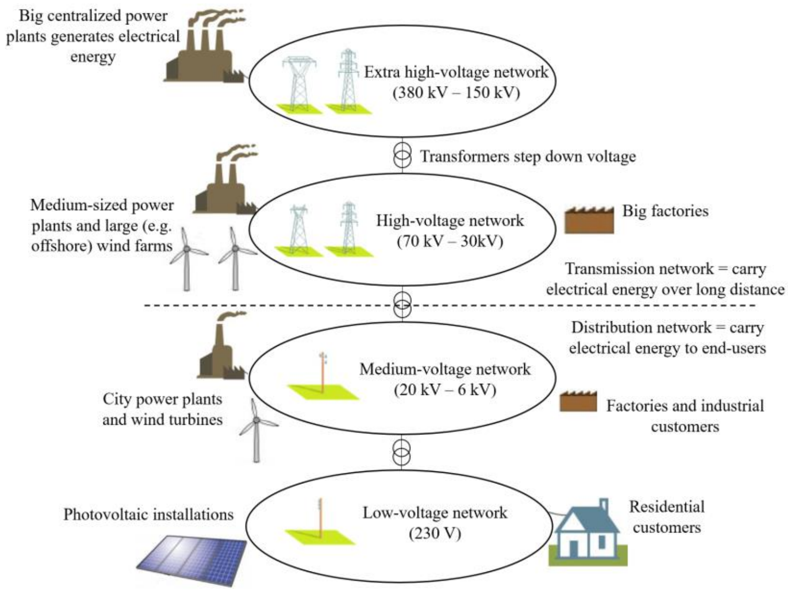


Figure 2.1 – Organization of the physical electrical network.

In order to ensure the security of supply and a stable power grid operation, a continuous balance between the total generation and consumption (including grid losses due to Joule effect) has to be maintained over the electrical system. Indeed, in case of lack of supply, the missing electricity is taken from the inertia of the rotating machines that are synchronized with the grid. These generators are then decelerating, which leads to a decrease of the network frequency, fixed to 50 Hz in Europe[8]. Likewise, a rise in frequency is observed when the generation exceeds the global demand. Such speed variations can be damaging for the machines if the rotating speed goes outside its operational limitations. The different units dispose thus of a security system, that continuously monitors specific parameters such as the frequency and the voltage level, and automatically disconnects the unit when an undesired value is detected. Consequently, if a frequency imbalance is not immediately alleviated, the electrical grid faces a domino effect of disconnections of generators, which importantly jeopardize its stability. In this way, a failure to restore the balance could ultimately lead to a blackout (system collapse), typically when the frequency drops below the critical value of 47.5 Hz.

However, due to legal obligations, the Belgian TSO (Elia) cannot possess its own generation means and is thus not able to ensure directly by itself the stability of its network. Neither can the TSO have large storage units for compensating the imbalances within its control area. For this reason, the task of maintaining the grid balance is attributed to other entities,

[8] The North-American grid as well as the southwestern part of the Japanese grid are operating at the frequency of 60 Hz. Electrical devices launched on the Japanese market can thus be switched between both 50 and 60 Hz.

commonly known as Access Responsible Parties (ARPs), which take the responsibility to compose a balanced portfolio on a quarter-hourly basis. As the balancing area of each ARP depends exclusively of its own portfolio, there is no geographical logic in the repartition of the total control area (Belgian Elia grid) among ARPs. In this way, for residential households, this role is ensured by their electricity supplier and the management of a given part of the distribution grid (such as a street) can be shared between several ARPs, each one being responsible for different clients. In this way, the total energy exchanged at the Elia grid access points (interface between transmission and distribution levels where the energy exchanges are measured) is distributed among different ARPs.

Each ARP is thus responsible of the continuous energy balance within its portfolio, which can be composed of its own generation, its own consumption, but also of the electricity traded with other ARPs. Indeed, in order to help ARPs in their balancing task, different opportunities, referred to as electrical energy markets (Section 2.3) are at their disposal for exchanging energy (at different time horizons, from years ahead up to close to real-time).

Access Responsible Parties can also import or export electricity via tie lines with neighboring countries. Practically, the auctioning of this cross-border capacity, which allows ARPs to acquire the right to import or export electrical energy aims at providing a transparent market based method for congestion management.

The electrical energy exchanged in the wholesale market has still to be supplied to end-users connected to the network. This is carried out via the retail market in which retailers, which have purchased and/or self-generated the electricity, sell the latter to their clients (Section 2.4). The final price covers not only the electrical energy actually delivered, but also grid fees (from both transmission and distribution levels) and taxes and levies (e.g. to support renewable energies or any other policy target, protect the more vulnerable consumers, etc.). The task of predicting the right amount of electrical energy that will be necessary to supply end-users for each period of the day belongs therefore to the ARP responsible for the retailer portfolio.

However, in case of real-time imbalance (after closure of energy markets) between total load and generation within a control area (imbalance resulting from the sum of the net imbalance position of all ARPs), the TSO is responsible of restoring the balance (Section 2.5). To that end, the TSO needs to call on **balancing services** that are classified into different products with regard to their response speed (Figure 2.2). First, the frequency containment reserve (FCR), or primary reserve, is automatically activated in a decentralized way to stabilize the grid frequency after a disturbance (alleviating momentary frequency deviations). In this way, when an imbalance occurs, it is thus almost mitigated through the contribution of primary reserves from the whole European interconnected transmission system. Then, the automatic frequency restoration reserve (aFRR), or secondary reserve, aims at restoring the balance in the control zone, thereby relieving the activated FCR within the system. However, if the problem persists, the system operator requests the activation of the manual frequency restoration reserve (mFRR), or tertiary reserve, which remains online until the situation is resolved.

These reserves were historically provided by conventional power plants due to their ability to efficiently modify their output power (hydropower is the most flexible technology, gas and, to a lesser extent, coal have also good ramping capabilities, while nuclear is the least flexible source). However, the current context is driving the emergence of new actors. The most popular ones are currently storage units, demand response strategies (deferrable loads), and modulation of the output power of renewable energies through power electronic devices.

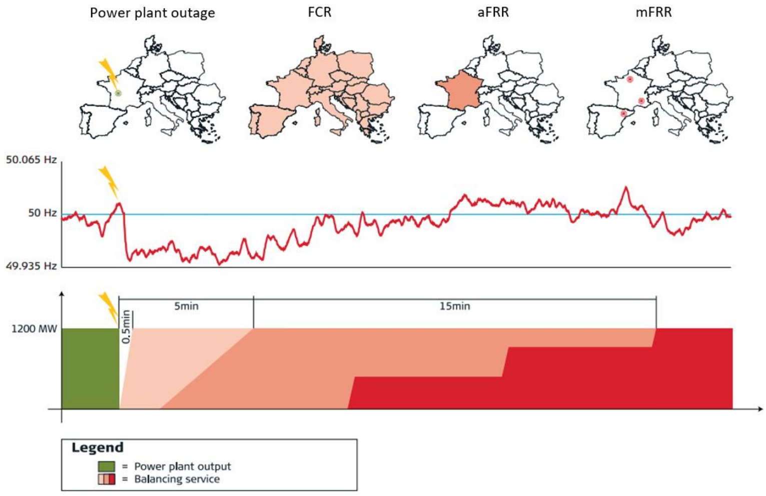


Figure 2.2 – Activation procedure of balancing reserves [Swissgrid[10]].

The procurement of balancing capacity is performed in mid-term (in Belgium, the FCR and aFRR is attributed following a weekly procurement procedure, while the remaining mFRR volume is purchased via a monthly purchasing cycle), and is remunerated at a fixed price throughout the contractual period. **The balancing energy can then be requested in real-time when necessary for facing residual grid imbalances.** The price related to this activation covers both the start-up costs (if relevant) and the energy effectively supplied. The flexibility remuneration encompasses therefore two contributions: payment for the availability and for the actual provision of the reserve. Finally, the TSO carries out a posteriori analysis to check whether the reserves are correctly activated and to evaluate the efficiency of the balancing service operation.

Whereas the fixed costs related to the availability of the reserve (capacity) are included in the grid fees (which are ultimately reflected in the electricity bill of end-users), the variable costs resulting from the real-time activation of the reserve are covered by the ARPs who were not able to fulfill their balance position. This mechanism, known as **imbalance settlement**, acts as a financial incentive for market players not to deviate from their schedule (Section 2.6).

It should be emphasized that the activation of FRR (aFRR and mFRR) is neutralized on the balancing perimeter of the ARP so that the participation to these balancing services cannot lead to portfolio imbalances (that have to be financially compensated). The contribution of FCR, however, is more complex to measure (due to its decentralized and automatic activation) and is assumed to be symmetrical in time (same amount of upward and downward reserves activated throughout the day). Consequently, the contribution of FCR is not offset from the ARP's portfolio.

The balancing services, along with the voltage control, congestion management and black start capabilities[9], are referred to as ancillary services.

For complementing the electrical energy and ancillary services markets, capacity remuneration mechanisms have been introduced in an increasing number of European countries in order to guarantee the stability of the electrical system in case of demand peaks [EPRS[16]]. In

[9] The black start services are used to re-energize the transmission system (in case of black-out) and provide start-up power to generators which cannot self-start. Black start service providers are thus generators that are able to restart without electricity.

Belgium, it materialized with the strategic reserve (Section 2.7), which was designed to cover the structural shortage[10] in electricity generation during the winter period and consists in a reserve of power coming from both off-market power plants and demand-side management offers [Elia[14]]. Overall, the strategic reserve is a *capacity-based payment* designed to minimize the interferences with energy and balancing markets. In this way, this mechanism differs from balancing reserves, which are used to offset the sum of residual imbalances of ARPs in real time. Hence, even in period of shortage when the strategic reserve is activated, the TSO can still face residual imbalances and have to resort to the balancing market. The advantage of the strategic reserve can be summarized in three contributions. Firstly, it participates to the security of supply of the country during situations of scarcity and prevents thus the extreme solution of shedding grid users (who are not remunerated during this forced outage). Secondly, this strategic reserve preserves the balancing reserves that are not intended to address such structural deficit in generation. Finally, it avoids the mothballing or dismantlement of fully functional power plants by giving them remuneration for the provision of the strategic reserve.

The balancing and strategic reserve markets are designed with the underlying objective to minimize interactions with electrical energy markets so as to avoid that market participants intentionally contributes to weaken the network stability in order to take advantage of this situation on balancing and capacity markets.

Finally, each ARP must compensate the active electrical losses (power dissipated as heat in transformers and conductors) on the federal grid related to all its network connection points. In order to distribute the contribution associated with each ARP in a transparent and non-discriminatory way, the financial compensation is expressed as a percentage of net offtakes of each ARP portfolio. From the 1st of January 2018, the applicable percentages are [Elia[18]]:

- Peak hours (weekdays from 8h00 until 20h00): 1,30%;
- Off-peak hours (weekdays from 20h00 until 8h00 and weekends): 1,20%.

The general structure of liberalized electricity markets with the different interactions among participants is summarized in Figure 2.3.

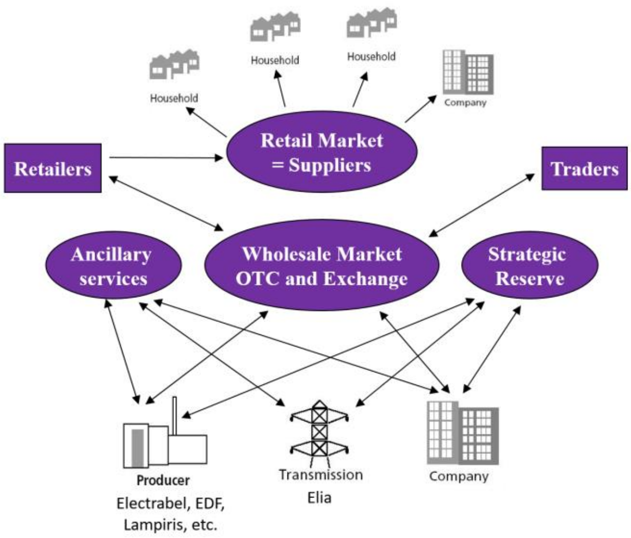


Figure 2.3 – General overview of electricity markets [Elia[13]].

[10] The structural shortage of a control area is evaluated on the basis of the statistical computation of the Loss of Load Expectation (LOLE), which reflects the number of hours during which the total generation will not be able to cover the load, taking the interconnections into account for a statistically representative year.

Transmission System operator (TSO)

The transmission system operator is responsible for managing both economic and technical aspects related to the transmission system, namely:

- Build and maintain the electrical grid;
- Assure the non-discriminatory (equal) access to all customers;
- Guarantee the security and quality of power supply;
- Control and manage the energy balance, while preventing voltage violations and line congestions;
- Promote efficient and transparent electricity markets.

In Belgium, the transmission system is managed by Elia, which is responsible for the totality of 380-220-150 kV network as well as for 94% of the high voltage network (70-30 kV). The connection to the grid depends on the installed power. Typically, if the maximum power is lower than 25 MW, the unit must be connected to the distribution network, whereas clients (generation or consumption) with an installed power equal or greater than 25 MW are connected on the Elia grid.

It should be noted that interconnections between European electricity grids enable cross-border electricity exchanges, while allowing countries to help each other in case of need. This collaboration between TSOs is managed by the European Network of Transmission System Operators for Electricity (ENTSO-E).

Distribution System Operator (DSO)

Distribution companies (such as Ores for 197 municipalities in Wallonia) are responsible for the safe and reliable operation of the distribution network. Their role is also to connect end-users, install electricity meters and communicate the metering to the suppliers. Similarly to the transmission system, the costs associated with the management of the network, known as grid fees, are passed on to the final clients (both consumers and producers).

Generation companies (producers)

In the liberalized electricity market, the objective of power producers is to maximize their profit by selling their energy to end-users. In Belgium, generation units (such as nuclear, gas-fired, coal-fired, combined heat and power units, etc.) that are directly connected to the Elia grid or with a nominal capacity higher than 25 MW must sign a CIPU (Coordination of the Injections of Production Units) contract with Elia. The purpose of such a contract is to inform Elia of the scheduling of the generation fleet. This starts from one year-ahead with information regarding the availability of power plants to the short-term (day-ahead and intraday) communication of the quarter-hourly scheduling of each generation unit. Moreover, this contract provides a legal framework so that Elia can use the capacity that is not used by generators, thereby providing additional flexibility to the power system that can be employed to either complement the FCR, aFRR and mFRR reserves, or for voltage control, congestion management or black start purposes.

End-users (households and companies)

The end-users (or consumers) are distributed throughout the system, and can be of any size from residential households (connected in the low voltage network that buy electricity in the retail market) to major industrial actors (connected to the high-voltage grid that can directly participate in the wholesale electricity markets). Up until now, the aggregated demand was typically considered as inelastic, i.e. independent to price variations, but the large-scale integration of volatile renewable generation is opening the door to demand-side management (such as deferrable loads).

Storage utilities

In Belgium, there is no legal status for storage units. These are thus alternatively considered as electricity consumer (in charging mode) and generator (in discharging mode). A significant consequence is that such utilities have to pay grid fees and taxes for both offtakes and injections towards the network, which significantly reduce their profitability.

Retailers

The retailers buy electrical energy on wholesale markets, and then sell this energy (= supplier) to end-users that are not participating to the wholesale market. All suppliers compete in the retail market to sell electricity to final consumers, who can freely choose their supplier based on the different offers (fixed-price or variable-price, deal over one year or several years, traditional or green generation mix, etc.).

The electricity market liberalization has led to the differentiation of the final price into different components: energy commodity, grid fees as well as taxes and levies. The suppliers can only compete on their price offer regarding the energy component.

Access Responsible Parties (ARPs)

All market players participating to the wholesale market (and/or ancillary services) must sign an ARP contract with Elia. All others parties (consumers and producers) must therefore be represented by an ARP that exchanges the energy in their stead. In many other countries, ARPs are often referred to as Balancing Responsible Parties (BRPs).

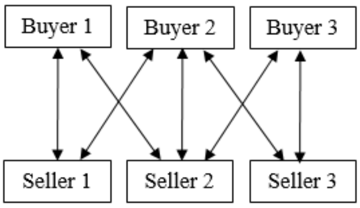
Regulator

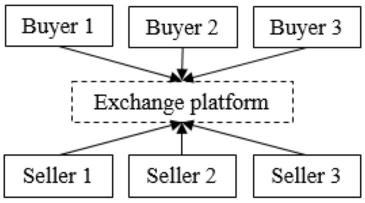
The electricity markets are controlled by an independent organization called a regulator, which is entrusted to ensure transparency and competitiveness of electricity markets, with the driving goal of serving the public interest. The regulator determines or approves the electricity market rules and monitors the operation of the market. The objective is to ensure fair prices concerning the different products and services. If needed, it also investigates the suspect cases of abuse (when market power is exercised). The (federal) Belgian organism in charge of the regulation of the Belgian transmission system is called the CREG (Commission of Regulation of Electricity and Gas), while the entity at the European level is the Agency for the Cooperation of Energy Regulators (ACER).

2.3 Wholesale energy markets

In a liberalized market, the energy can be freely traded between ARPs. Two types of electrical energy exchanges are coexisting, i.e. over-the-counter (OTC) markets and “power exchanges”. The characteristics and the different products associated with these two wholesale markets are summarized in Table 2.1. In OTC markets, the participants negotiate one with another without a central physical location. In this way, two dealers (ARPs) can directly trade electricity volumes and prices without others knowing the details of the transaction. Such exchanges have thus little transparency, and are subject to fewer regulations than traditional markets. The main advantage of such contracts is that they can be completely customized to fit a customer’s requirements, giving more flexibility to the involved parties. Then, in parallel to these decentralized markets, “power exchanges” offer (fully electronic) anonymous platforms for trading energy with higher levels of security and liquidity (due to the many participants).

Table 2.1
Main characteristics of wholesale electricity markets.

<table>
<tr><td colspan="2"></td><td>OTC market</td><td>Power Exchanges</td></tr>
<tr><td colspan="2">Characteristics</td><td>
Bilateral
Not anonymous
Standard or customized products


</td><td>
Multilateral
Anonymous
Standard products


</td></tr>
<tr><td rowspan="2">Products</td><td>Long term</td><td>Forward contracts</td><td>Futures contracts</td></tr>
<tr><td>Short term</td><td>Spot OTC contracts</td><td><u>EPEX Spot Belgium:</u>
Day-ahead market
Intraday market</td></tr>
</table>

Electrical energy markets are differentiated according to their time horizon, starting years before the actual delivery until even after the supply of energy. An overview of the temporal ordering of electricity markets, with a particular emphasis on “power exchanges”, is shown in Figure 2.4.

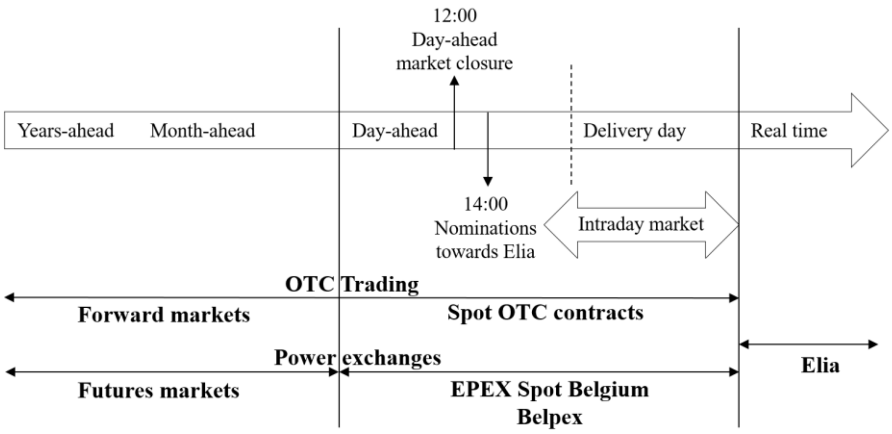


Figure 2.4 – Rationale of wholesale energy markets.

Next to forward (over-the-counter) trading, the futures market allows exchanging energy long before actual delivery, thereby ensuring to both buyer and demander to fulfill their basic need for a long period.

Then, every day at 12h00, the (spot) day-ahead wholesale market is cleared through an auction mechanism at the end of which both clearing price and volume are obtained for the 24 hours of the next day. At 14h00 the day before the physical delivery, i.e. after that the day-ahead market results are unveiled (typically around 13h05), each ARP must provide Elia with **nominations**, i.e. the balanced schedule on a quarter-hourly basis of its power injections and offtakes in order to help Elia predicting the volume and distribution of power exchanges on its network at different times. Such nominations are essential, since the market is cleared without accounting for grid constraints within the Belgian market zone[11], which can lead to voltage violations or line congestions. The nominations allow thus anticipating potential problems, and to take the appropriate measures. Indeed, the transmission capacity constraints are checked in a second stage (with the information from nominations) by the TSO, with potential re-dispatching actions to prevent future congestion. In Belgium, this re-dispatching affects currently 0.08% of the annual electricity production, and increases the operating cost by approximately 0.3%, i.e. 2.9 million Euros per year [Van den Bergh[15,a]].

At the same time (14h00 the day before delivery), the intraday market is opened up and allows market players to face forecast errors by exchanging energy until close to real-time.

2.3.1 Long-term markets

Long-term markets run from years before delivery up to the day-ahead. Such forwards (customized products traded bilaterally over-the-counter) and futures (standardized products traded on "power exchanges") are contracts to exchange a fixed amount of electricity at a certain time in the future for a price agreed when the contract is made.

Long term markets provide security for market players who can sell or buy their base load well in advance, typically from one to three years. This market is thus perfectly suited for inflexible plants with a steady output power such as nuclear plants and run-of-the-water hydro units but also for large consumers[12] eager to pay a regular price for their base consumption. Globally, the benefits of the long-term markets can be summarized in three respects [Ausubel[10]]. First, the long-term markets address risk by allowing the participants to secure long-time prices and quantities, limiting interactions with the much more volatile spot market[13]. Secondly, the participants having a more balanced position when entering the spot market will be less attempted to distort bids for ensuring the acceptance of their offers[14], which improves the market power. Thirdly, the long-term markets facilitate investments in new resources.

The available products can be divided into base, peak and off peak profiles, and allow exchanging a fixed amount of energy during the whole contractual period that can cover either a day, a week, a month, a quarter (3 months) or a year. One of the most commonly traded electricity products is the baseload future for one year, which corresponds to the delivery of a fixed amount of energy for each hour of the year. The futures prices mostly depend on fuel prices, due to their significant contribution in the marginal costs of conventional generation.

The Belgian power futures can be traded on the ICE Endex and the European Energy Exchange (EEX). As an example, the total volume of monthly futures (divided into temporal products) exchanged on the EEX platform during 2014 and begin of 2015 is illustrated in Figure 2.5.

[11] A market zone, or bidding zone, generally corresponds to a Member State, although there exists some exceptions (Germany and Austria jointly constitute only one market zone, whereas countries such as Norway and Sweden contain multiple bidding zones).
[12] Only large industrials directly connected to the transmission grid are allowed to participate to electricity markets. The purchase of electricity for other clients is usually carried out by their electricity supplier.
[13] The day-ahead market prices can go beyond 80 €/MWh in case of high demand, and drop to zero (or even below) in cases of low demand.
[14] The market participants that are under the obligation to sell (or purchase) a large quantity of energy in order to be balanced are likely to make offers at the maximum (or minimum) market price for ensuring to be cleared by the market algorithm, which can distort the traditional law of supply and demand and negatively impact prices.

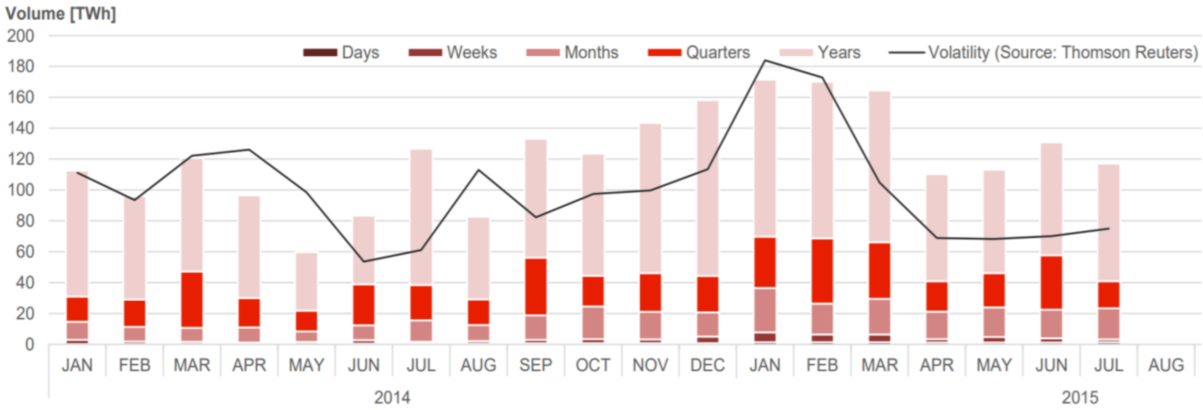


Figure 2.5 – Volumes exchanged in the EEX Power futures [EEX[15]].

The importance of such long-term markets was demonstrated during the California crisis of 2000-2001. This crisis was indeed characterized by insufficient forwards or futures contracting, and a situation of supply scarcity. Therefore, during this period, the balance between supply and demand was distorted such that prices shot up. The producers took advantage of the situation by exacerbating even more the high prices. The electricity suppliers who did not dispose of their own generation means faltered to bankruptcy, and the market ultimately collapsed.

2.3.2 Day-ahead market

On the day-ahead "power exchange", the electricity is traded at 12h00 for the 24 hours of the following day. The day-ahead market (DAM) consists in an anonymous trading platform where the cleared price (MCP) and volume (MCV) are set by the intersection of the demand curve and the supply curve as represented in Figure 2.6.

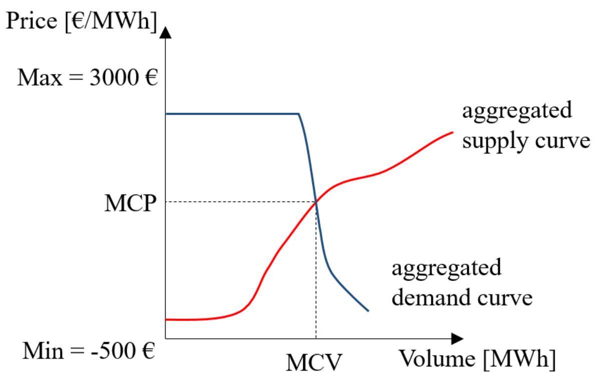


Figure 2.6 – Determination of the day-ahead market clearing price (MCP) and volume (MCV).

The demand curve is a monotonically decreasing function where demand bids are ranked with respect to their price offer from the buyer that accept to purchase at the highest price to the lowest offer. Similarly, the supply curve is a monotonically increasing function where the sale bids are classified from the most competitive price offer to the one associated with the highest price. Under normal circumstances, the contracts are published no later than 13h05 (i.e. the clearing procedure is completed 1 hour after the market closure).

Practically, the market clearing procedure aims at maximizing the social welfare of all participants by matching the offers from producers and bids from consumers so as to obtain the clearing prices and volumes for each time period of the horizon. As represented in Figure 2.7, maximizing the social welfare amounts to find the equilibrium between the maximization of producers and consumers surplus.

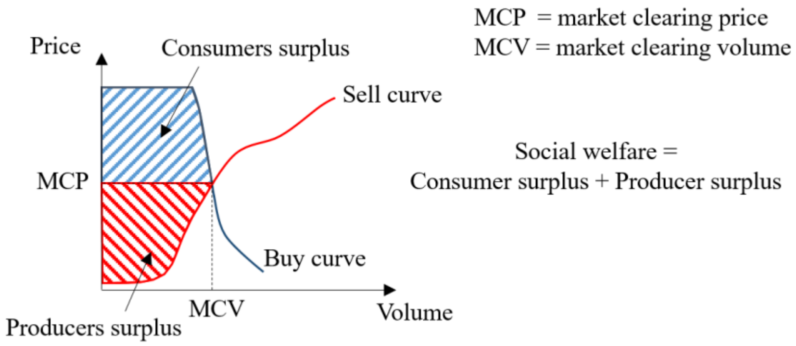


Figure 2.7 – Definition of social welfare.

In Belgium, the market operator is EPEX SPOT Belgium (or Belpex). The energy exchanged on this platform represents currently around 30 % of the total yearly consumption (~ 83 TWh), whereas it was only 20 % in 2013.

The ARPs who do not wish to participate themselves in the wholesale market may delegate this task to traders. This situation is referred to as *indirect participation* on the Belpex spot market. This is an excellent solution for ARPs with less trading experience since operational trading activities are taken care of by a service provider who trades on the ARP's behalf [Belpex[15]].

The Belgian market zone is implicitly coupled with other European market zones through a mechanism known as **market coupling**. The principle is to match the highest purchase bids with the lowest sales offers, regardless of the bidding zone in which they have been introduced, while accounting for the available cross-border capacities. The energy and interconnection capacity are thus traded together, and the market participants have automatically access to transmission capacity by submitting orders to the "power exchange". This mechanism results in a price harmonization between the coupled market zones when sufficient capacity is available. In this regard, market coupling represents a major step towards an integrated European market that increases the overall welfare for society.

The coordination between market zones is essential since electricity flows cannot be imposed by commercial trades but are subject to the law of physics (Kirchhoff's law). In this way, when France imports electrical energy from Germany, part of the exchanged power will flow across Belgium instead of following the transmission lines existing between both countries. The impact of all transactions therefore needs to be considered on the available capacity at each border.

Figure 2.8 shows the aggregated net positions for 2016 for the different market zones across Europe (from 2015 to 2016, the net positions changed slightly). The negatives numbers show the total electricity imports whereas exports are associated with positive values. It can be seen that the Belgian importations are more than 6 times higher than the exportations.

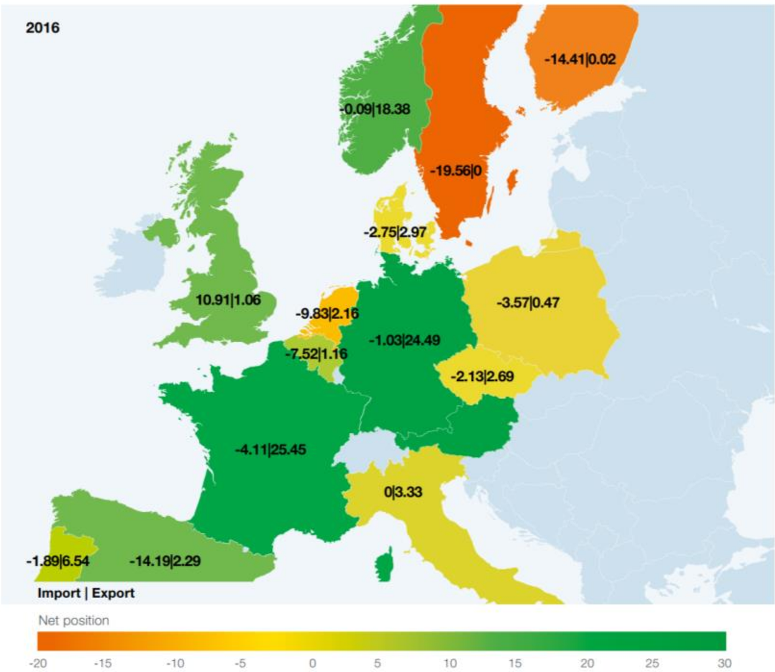


Figure 2.8 – Aggregated net positions in Central Western Europe region, with imports (negative values) and exports (positive values).

In Central Western Europe (CWE), the **Flow-based market coupling** model is used since 2015. We refer the interested reader to [Elia[15]] for more detailed explanations concerning the whole methodology.

It should nonetheless be mentioned that the market participants have access to different products to optimally manage their portfolio in day-ahead. Firstly, they can rely on **limit orders**, which are characterized by a specific volume and price that are offered for a particular one-hour segment of the day. A buy limit order is executed (matched by the market) if the final market clearing price is lower than the offered price, whereas a sell limit order is in-the-market only if its price is higher than the clearing price. Beyond this basic product, market players can also resort to more complex products, known as **block orders**. Block orders are currently of four types, namely regular orders, profile orders, linked orders and exclusive orders. The last two are often referenced to as *smart orders* due to their even more complex nature.

Regular blocks orders, as illustrated in Figure 2.9 (a), link a number of consecutive hours characterized by the same volume and the same price for each selected hour. If one of the selected hours of the order cannot be matched, then the whole block order will not be matched. Profile block orders, however, as represented in Figure 2.9 (b), allow users to submit a block order for several non-consecutive hours of a delivery day as well as to submit different volumes for each selected hour.

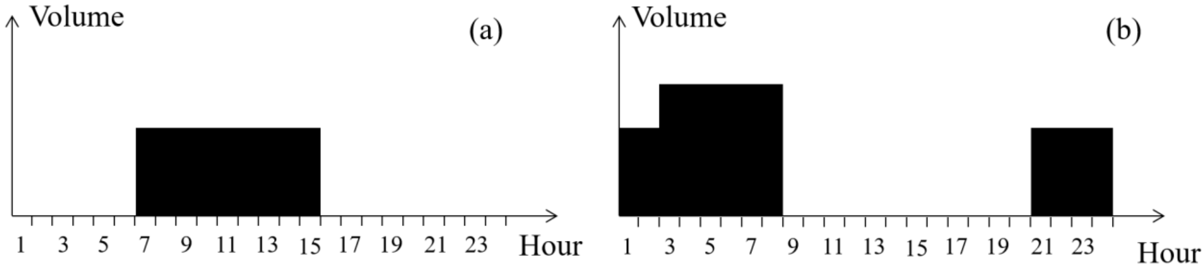


Figure 2.9 – Regular block order (a), and profile block order (b).

It is also possible to define links between block orders (linked block orders), which means that the acceptance of one block order (daughter block) depends on the acceptance of another block order (mother block). In this way, several block orders can be linked together, thereby creating a family, which allows to explicitly consider technical and economic constraints of units. For instance, as represented in Figure 2.10, a first block may include the start-up costs of a generation unit, and daughter blocks accounting for fuel costs may be linked to this first block.

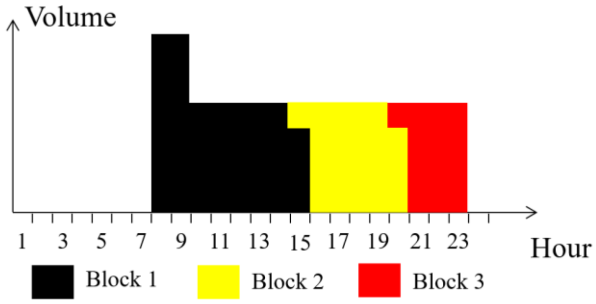


Figure 2.10 – Illustration of linked block orders.

Finally, an exclusive order is defined as a set (currently limited to maximum 24) profile block orders where at most one block among the whole set can be accepted.

All these block orders couple (link) hourly periods between them, and introduce non-convexities in the search space (market clearing procedure), which considerably increase the complexity of finding an optimal solution (which is typically obtained and revealed at 13h05).

After the clearing of the day-ahead market (and the notification to market participants of which orders are accepted and which are off-the-market), each ARP has to submit a balanced portfolio at 14h00 to the TSO (the so-called nominations). These nominations give the planned generation or consumption for every connection node to the transmission grid (which contrast with trading which is performed at the portfolio level). Moreover, the nominations have a quarter-hourly time resolution whereas energy exchanges have an hourly time step.

2.3.3 Intraday market

The intraday market opens up at 14h the day before delivery, and enables market participants to exchange energy up to five minutes before delivery. The ARPs can thus meet any unexpected changes in their electricity portfolio (due to forecast errors and unexpected events), and subsequently correct their day-ahead nominations. For instance, if an ARP faces an unexpected power plant outage after the day-ahead market closure whereas the unit was committed to online between 18h00 and 20h00, the ARP can buy energy on the Intraday market until 17h55 for the 18h-19h period, and until 18h55 for the 19h-20h period.

The Intraday market is an organized (transparent and anonymous) platform, where participants can continuously submit generation and demand bids so that the market is also cleared continuously. Practically, it consists in an open order book where the participants can see all the other (anonymized) orders. In this way, one market player can bilaterally accept the bid of another market player, which results in different prices for each trade.

The principle is that the demand bid with the highest (more competitive) price is matched with the selling offer with the lowest price, provided that the demand price exceeds the selling price. However, the market price is settled according to the order that was submitted on the market first. In this way, when a participant submits on the trading platform a demand order with a price of 50 €/MWh, and that another participant submits afterwards its selling order with an asking price of 30 €/MWh, the resulting transaction price will be 50 €/MWh. In case two selling (or buying) orders have similar prices, the first submitted order will be automatically prioritized (the volumes have no impact on the prioritization).

Within the objective to achieve a more efficient integrated European electricity market, the European Commission has established guidelines where the cross-zonal transmission capacity (still available after the day-ahead market) is implicitly allocated in the Intraday trading. This resulted in an initiative called the XBID Market Project involving the TSOs from 12 countries (including Belgium). It allows thus ARPs to benefit not only from the available Intraday liquidity[15] within their bidding zone, but also from the available liquidity in other areas, provided that there is sufficient cross-border capacity.

An overview of the main specifications associated with both day-ahead and Intraday markets is given in Table 2.2.

Table 2.2
Overview of the main product specifications [Belpex[18]].

Specifications	DAM	CIM
Instrument duration	1h	1h
Trading window	14 days before delivery until 12h the day ahead of delivery	From 14h the day ahead of delivery until 5 min before delivery
Publication time	Usually, no later than 13:05 the day ahead of delivery	Immediately
Order type	Limit orders and block orders	Limit orders and block orders
Volume bounds	[0.1, →[MWh	[0.1, →[MWh
Fixing process	Auction	Continuous trading
Price bounds	[-500, 3000] €/MWh	[-99 999.99, 99 999.99] €/MWh
Price accuracy	0.01 €/MWh	0.01 €/MWh
Elia nomination gate closure time	14:00 day-ahead	14:00 day after delivery

The residual real-time portfolio imbalances are taken care of by the TSO in the balancing market (Section 2.5), but the costs associated with this reserve activation are financially impacted on the unbalanced ARP (Section 2.6).

2.4 Retail energy markets

The structure of the retail market is designed such as to enable to the different customers to take advantage of attractive prices thanks to the competition among different suppliers. The simplified framework is represented in Figure 2.11. Firstly, a new customer needs a connection contract with the Distribution System Operator responsible of the area (e.g. Ores for Mons). Then, he has to choose his electricity supplier among the private companies operating in his area. In order to practice as an energy supplier in a given area, the interested company needs to

[15] The market liquidity represents the volume of activity within the market, i.e. the extent to which the commodity can be exchanged with stable prices.

sign beforehand an access contract with the DSO. It is indeed by means of this contract that the DSO passes on the transmission and distribution grid fees to the suppliers.

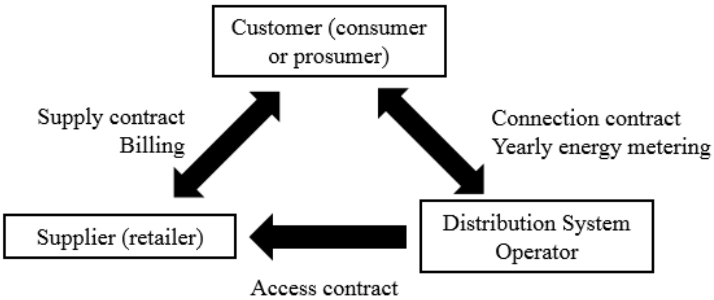


Figure 2.11 – Operation of the electricity market at the distribution level (retail market).

Currently, the billing between the supplier and its customers depends on the yearly consumption. Hence, for avoiding any conflict of interest, the yearly electricity meter reading is not carried out by the supplier but by the DSO (neutral party regarding the energy prices). However, the billing mechanism is expected to change in the future with the rolling out of smart metering devices that will allow the implementation of time-variant pricing to better align consumption with renewable generation.

2.5 Ancillary services

With the separation between generation and transmission tasks (issuing the electricity market liberalization), the TSO does not have the resources to alter injections or offtakes, and is therefore forced to buy such flexibility capacities from market participants (generators and consumers). These flexible resources, or reserves, are necessary to ensure the continuous balance of the system (for the frequency stability) while maintaining voltage at suitable levels, preventing line congestions, and guaranteeing the grid recovery in case of major incident by means of black-start resources. The services are commonly known as **ancillary services**.

2.5.1 Balancing market

The Belgian high-voltage grid is part of a larger interconnected system. The coordination of the operation and the development of this European system is handled by the European Network of Transmission System Operators for Electricity (ENTSOE). The members of ENTSOE are shown in Figure 2.12.

ENTSO-E requires that each TSO allows a maximum deviation of the frequency level of ± 20 mHz within its control area. Elia is even more demanding and aims at maintaining the frequency within a range from 49.99 Hz to 50.01 Hz. Indeed, uncontrolled frequency deviations quickly generate grid instability, which may eventually degenerate into a blackout. For preventing such an extreme scenario, each TSO disposes of balancing reserves that are made up of several products and enable the TSO to restore the grid frequency in case of imbalances. These reserves are classified according to their response speed into the following categories:

- Frequency Containment Reserve (FCR), or Primary reserve (R1);
- International Grid Control Cooperation (IGCC);
- automatic Frequency Restoration Reserve (aFFR), or Secondary reserve (R2);

- manual Frequency Restoration Reserve (mFFR), or Tertiary reserve (R3);
- mFRR non-reserved power;
- Inter-TSO exchanges.

Throughout the dissertation, the terms '**frequency services'**, '**reserves**' and '**operational flexibility**' are used interchangeably to refer to '**balancing services'**.

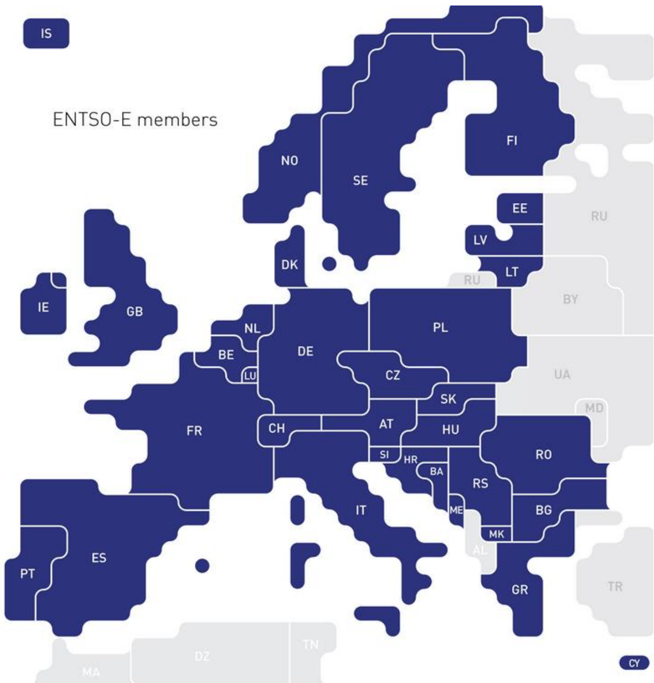


Figure 2.12– Map of ENTSO-E members [ENTSOE[18]].

The volume of the FCR (fixed to 3000 MW on ENTSO-E level) is distributed every year between the different control zones according to their weight in the synchronous area of Continental Europe. The sizing of the FRR capacity is then delegated to the national TSOs, but ENTSO-E nonetheless sets boundary conditions for the sizing procedure to guarantee the safe and reliable operation of its transmission grid. In this way, the FRR capacity in both in upward and downward directions cannot be smaller than their respective highest system imbalances. In Belgium, the required total volumes for 2018 are 81 MW for (symmetrical, i.e. both upward and downward) FCR, 139 MW for (symmetrical) aFRR and 830 MW for (upward) mFRR.

As previously mentioned, the volumes of FCR and aFRR reserves are auctioned in week-ahead, and the mFRR is contracted on a monthly basis. It should be noted that market players can bilaterally transfer reserve (FCR, aFRR and mFRR) obligations to each other via a secondary market.

Currently, the reserve procurement is mainly organized at the national level. However, one of the main target of the European Commission regarding the electricity markets is to harmonize the balancing mechanism across countries so as to achieve a more efficient procurement of reserves, while reducing the risk of shortage in reserve supply for TSOs, thereby increasing the overall system security. Obtaining such an **integrated European balancing market** is a complex task due to the varying market designs across countries [ENTSOE[15]]. Market zones are indeed characterized by different clearing mechanisms ("pay-as-bid", "pay-as-cleared", etc.)[16] with different activation rules (sequential activation of offers based on a merit order, or parallel activation of all participants on a prorate basis), have different times at which the procurement of both capacity and energy is carried out (month-ahead, week-ahead, day-ahead, etc.), propose different types of products (symmetrical, asymmetrical, etc.) for each service with different time and size (in MW) resolutions, etc.

However, as a first step, the IGCC was introduced in 2012 in order to enable TSOs to exchange opposing imbalances (thereby reducing the activation of opposite reserves in different zones). Then, since the first week of August 2016, a cross border FCR auction is carried out for German, Dutch, Swiss, Austrian, Danish and Belgian market zones, introducing competition between foreign flexible resources (and shifting FCR obligations between countries). Another important guideline that dictates the evolution of the balancing market design is the switch from technology-oriented products (development of new products to foster participation of a particular technology) to technology neutral products (so as to ensure a level playing field for all technologies). In Belgium, it is translated into the progressive opening of the market to the flexibility distributed in all layers of the grid (both transmission and distribution grids) and an anticipated simplification of the proposed products.

Then, the sizing of the reserve capacity is expected to shift from a static approach (on the current annual basis) towards a dynamic reserve sizing dynamic, i.e. over smaller time periods, e.g. on a monthly, weekly, daily or hourly basis [van Stiphout[17]].

Frequency Containment Reserves (FCR)

The frequency containment reserve (FCR) consists in an automatic activation of power reserves in case of a frequency deviation. The participating units detect automatically frequency fluctuations and adjust their output power in a very short time for a period up to 15 minutes. The participation requires the installation of specific equipment able to continuously measure the grid frequency and to adapt their profile to reach the half of contractual primary reserve within 15 seconds and the whole reserved power over a timeframe up to 30 seconds. The reserve has to be available during each contracting period and the amount of activated power is linearly dependent of the frequency deviation Δf. Moreover, ENTSO-E imposes a symmetrical and linear activation of R1 with a total activation at a Δf_{max} equal to ±200 mHz. All these technical requirements are shown in Figure 2.13.

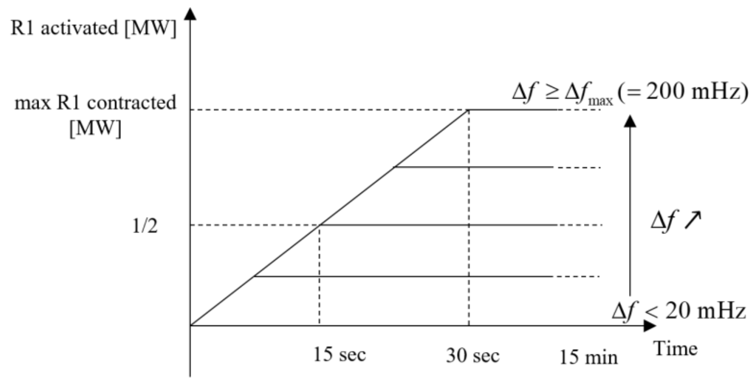


Figure 2.13 – Technical requirements for a local activation of primary reserve.

16 In a pay-as-bid system (such as in Belgium for the procurement of both reserve capacity and reserve energy), the market participants who are in-the-market (i.e. for which the offer has been accepted by the market) receive the price at which they bid in the market, resulting in a potentially different price for each player. It contrasts thus with the pay-as-cleared system (such as the day-ahead energy market) where market participants all receive the uniform market clearing price.

International Grid Control Cooperation (IGCC)

The International Grid Control Cooperation (IGCC) aims at taking advantage of opposite imbalances between neighbouring system operators. Practically, the IGCC consists in the pooling of the imbalances of each participating market zone, which allows avoiding the activation of balancing reserves in opposite directions and reduces thus the total volume of activated regulation reserves. This operation is carried out continuously with a 5 seconds refreshing rate of the imbalance signal and is limited by the available transmission capacity on borders as well as the amount of reserved aFRR of each zone[17]. The residual imbalance within a given control area must be addressed by the concerned TSO.

Automatic Frequency Restauration Reserves (aFRR)

The symmetric aFRR reserves are controlled automatically by the TSO and activated centrally based on a set point that is sent continuously to the reserve provider (Figure 2.14). The full activation of the reserve in one direction (upward or downward) must be performed in 7.5 minutes, and remains active for the time needed.

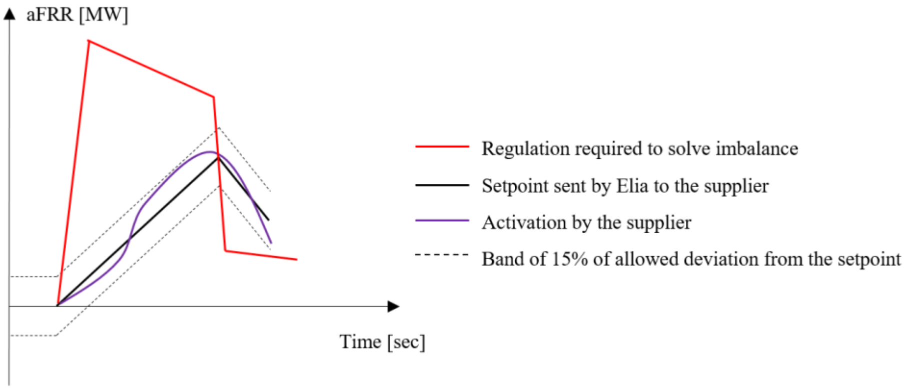


Figure 2.14 – Activation of the automatic frequency restauration reserve.

Manual Frequency Restauration Reserves (mFRR)

The mFRR enables the TSO to both alleviate the aFRR in case of significant imbalance in the control area (so as to offset frequency variations) and to cope with major congestion problems. Unlike the FCR and aFRR that are automatically activated, the mFRR is activated manually upon a specific request from the TSO. Any grid user (generation or consumption) whose resources comply with certain technical requirements (e.g. the power must be fully delivered within 15 minutes after request) can sign a contract with the TSO to take part in the service.

mFRR non-reserved power

In order to create the possibility to offer bids on the balancing market from flexibility coming from all grid users, aggregators and smaller production or storage units, Elia launched in July 2017 the “Tertiary Control Non-Reserved Power” (formerly called “bidladder”). The offered volumes are submitted until 45 minutes before real-time through the BMAP (Balancing Market Platform), and reserve providers are subsequently activated by Elia in case of need. These are only remunerated for the provided energy (and not for their availability). When activated by Elia, the reserve provider should activate the offered volume as soon as possible within a margin of 15 minutes.

[17] The imbalance of a control area that is pooled in the IGCC procedure cannot exceed the total volume of secondary reserve contracted by the concerned TSO.

Since it is desired that grid customers can valorize their flexibility independently from their basic requirements (e.g. non-shiftable consumption), they can offer their flexibility to new players, known as Balancing Service Providers (BSPs). If the BSP differs from the ARP of the same access/delivery point, a procedure (referred to as transfer of energy), has to be carried out to fairly distribute the energy flows among actors.

Inter-TSO support

Finally, as a last resort, available capacity from emergency support with neighboring TSOs can be used. The availability is non-firm and non-guaranteed.

2.5.2 Voltage control

As above-mentioned, generators with output power exceeding 25 MVA are linked to Elia through the Coordination of the Injection of the Production Units (CIPU) contract. In this respect, they have to contribute to the control (i.e. generation and absorption) of reactive power so as to maintain the voltage plan within acceptable limits (automatic regulation). The units receive a fixed remuneration to cover one-time expenses (IT communication with Elia, technical adaptations for expanding the technical band of the unit), and an activation price covering the reactive energy actually produced or absorbed.

Since the voltage control is a local issue (the voltage levels vary across the grid based on the distribution of power generation/consumption), Elia selects the participating units on the basis of their location. Elia currently contracts some 6300 MVAR of reactive generation capacity and 3200 MVAR of absorption capacity. The contracts have a one year duration.

If the voltage level is too low (due to a high load), Elia asks to generate additional MVARs, whereas it resorts to increasing the local consumption of reactive power when the voltage level is exceeding the upper limit. Penalties are applied in case the automatic control is not well executed.

2.5.3 Black start

The TSO has to make sure that it can restore its grid in the event of a blackout, by relying on generation units that can start up without an external electricity supply (e.g. hydro generation). The service provider receives a fixed payment, regardless of whether it is activated or not. The participating units must also be able to operate smoothly at any time and have to meet certain technical criteria such as a minimum power level (100 MW) and a grid restoration time. The units are selected on the basis of their costs and location.

2.6 Imbalance settlement

The **system imbalance (SI)** is the imbalance that would be faced by the system (Belgian control area) without activation of the power reserves. There is no direct measurement of this imbalance and it is therefore necessary to introduce the fundamental notion of **Area Control Error (ACE)**. The ACE is defined as the unintentional residual imbalance between supply and demand after the activation of reserve (due to the imperfection inherent to the balancing market). The ACE can be determined by taking the difference between the scheduled and measured values from the exchanges at the borders of the TSO control zone, taking into account the effect of frequency bias (i.e. linear approximation of the FCR contribution (in MW) within the control area).

Illustration of the concept of Area Control Error (ACE):

As an illustrative example, let us consider the case of an interconnected grid composed of two control areas A and B. If entity A has a disturbance in its balancing area, such as the loss of a generator of 500 MW, the FCR response in both areas will automatically increase the generation (or decrease the consumption).

Consequently to the loss of 500 MW in entity A, the energy flows from area B to area A will increase, and the ACE of entity B will therefore show a positive value (indicating a surplus of generation). In this way, if the frequency bias is not taken into account in the ACE equation, the balancing mechanism of entity B would react by decreasing the total generation (so as to restore the energy balance). The frequency bias term allows thus the FCR response of areas adjacent to an unbalanced zone to continue supporting the interconnection frequency.

Concerning the area in shortage, if the frequency bias is not considered, the ACE would be higher than -500 MW due to the automatic activation of FCR within its control area (e.g. ACE equal to -400 MW if 400 MW of importation for 100 MW coming from its own FCR). However, it is essential to alleviate as soon as possible the automatic primary control. Therefore, the frequency bias adjustment brings back the ACE value to a quantity as close as possible to the deficit caused by the disturbance, namely around -500 MW.

The purpose of the TSO is to minimize the ACE in order to maintain the grid frequency while keeping the cross border exchanges as scheduled. When the ACE is unbalanced, Elia has to activate FRR in compensation. This activation has a price, which is directly reflected by the imbalance tariff (single pricing scheme[18]). The imbalance tariffs depend mainly on two important concepts, namely the net regulation volume (NRV) and the marginal price.

The **net regulation volume (NRV)** is calculated for each quarter of an hour by taking the difference between all the activated volumes for upward regulation (GUV = Gross Upward regulation Volume) and the activated volumes for downward regulation (GDV = Gross Downward regulation Volume) requested by the TSO (including the exchanges carried out via the International Grid Control Cooperation). A positive value of NRV correspond thus to a negative imbalance of the system that had to be counterbalanced by the activation of upward reserves, and vice-versa. It can therefore be concluded that:

$$SI = ACE - NRV \tag{2.1}$$

The second indicator is the marginal price which is defined by two different regimes, depending on the type of activated reserve. In this way, for a particular quarter of an hour, the **marginal incremental price (MIP)** corresponds to the highest price paid by the TSO for upward regulation, whereas the **marginal decremental price (MDP)** is defined by the lowest price received by the TSO for downward regulation. It should be noted that the MIP will always be positive (i.e. the TSO will always pay a producer for extra power in situation of scarcity) while the MDP can be negative when the global generation is much higher than consumption.

[18] The single pricing means that ARPs with a negative imbalance are faced with the same price as the ARPs with a generation surplus, resulting in only one imbalance price for each player. This contrasts with dual pricing where separate prices are determined for positive and negative imbalances.

Indeed in such a case, the TSO is forced to activate reserve far in the merit order, where the prices are likely to be negative. The money flows are then reversed and the market players with a negative imbalance are paid for their excess consumption.

Table 2.3 summarizes the imbalance settlement procedure. The imbalance tariff applied to each ARP depends on both its own imbalance and the imbalance of the control area.

The imbalance of a given ARP is the quarter-hourly difference between its total injections and its total offtakes within its balance perimeter. In this way, a positive imbalance is characterized by a surplus of generation, while a negative imbalance consists in a lack of generation to cover the total demand. It is nonetheless worthwhile to recall that the activation of ancillary services, at the exception of the FCR (considered as symmetrical), is neutralized in the portfolio of ARPs, and cannot lead to imbalances.

The choice of the penalty price (MDP or MIP) is defined by the NRV sign within the control area (for the considered quarter-of-an-hour). If the value of the NRV is positive, the MIP is applied in the tariff for balancing energy and conversely, the MDP is applied in case of a negative NRV.

Table 2.3
Imbalance settlement.

P = Production C = Consumption		Situation in the TSO control area	
		Excess (P>C) NRV < 0	Shortage (P<C) NRV > 0
ARP area	Excess (P>C) Positive imbalance	MDP – α TSO pays ARP	MIP TSO pays ARP
	Shortage (P<C) Negative imbalance	MDP ARP pays TSO	MIP + α ARP pays TSO

The term α is an additional financial incentive applied to the regulation costs in case of important imbalance of the system.

It can be seen that four possible situations can be encountered:

- **There is a surplus in the TSO area and the ARP is in surplus as well:**

The imbalance of the ARP amplifies the severity of the global situation in the control area. If the situation is not critical enough to lead to negative prices (Section 2.8.1), the ARP is still paid for the surplus of production but at a price lower than it could have expected in the electrical energy market.

- **There is a surplus in the TSO area while the ARP area is in shortage:**

The imbalance of the ARP is helping the global situation in the TSO control area. The ARP is thus paying at a reasonable price for its negative imbalance and can even be paid in the extreme case of a negative MDP.

- **There is a shortage in the TSO area while the ARP is in surplus:**

The positive imbalance of the ARP is helping the global situation of scarcity within the TSO area and the ARP is therefore paid accordingly. The amount corresponds to the MIP, which is determined by the highest price paid by the TSO for activating the upward reserve. Consequently, the remuneration received by the ARP increases with the severity of the imbalance within the TSO control area.

- **There is a shortage in the TSO area and the ARP is in shortage as well:**

The negative imbalance of the ARP worsens the situation of scarcity of the TSO area. In such a case, the ARP can face major penalty if the situation necessitates activating power reserve located far in the merit order.

Overall, with an imbalance settlement characterized by a single pricing mechanism, the ARPs that degrade the system balance are penalized, while those that help maintaining the system balance are rewarded.

2.7 Strategic reserve

Each year, the Belgian Federal Minister for Energy may give to Elia the task of constituting a strategic reserve (SR) and sets the required reserve volume in MW. The strategic reserve may change over years depending on the requirements concerning the reliable and efficient security of supply. The strategic reserve is awarded to market participants through a competitive tendering procedure at the end of which the selected tenderers sign a contract with Elia for the winter period lasting from 1 November to 31 March.

The strategic reserve is activated once a risk of an energy shortage has been detected. To that end, two indicators, which are respectively referred to as economic and technical triggers, are used. The economic trigger is based on an automatic detection of shortage risk based on the day-ahead market clearing results. The technical trigger, however, is activated either in day-ahead or in the course of a day, based on forecasts of total generation and consumption.

2.7.1 Economic trigger

The first way of detecting a potential shortage is carried out at the end of the day-ahead market clearing process, when the market clearing volume (MCV) is not sufficient to cover the bids offered at the maximum price of the market (i.e. 3000 €/MWh). This situation is represented in Figure 2.15. If the available volume of strategic reserve (SR) contracted by Elia is not sufficient to cover all the 3000 €/MWh demand, this SR volume is prorated on this demand for ensuring the maximum welfare among market participants (the rest of the load needs however to be shed).

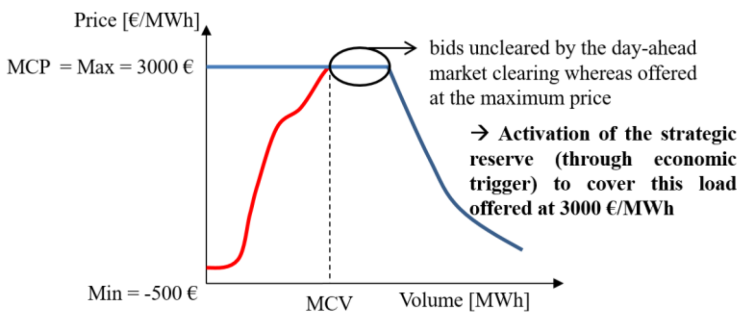


Figure 2.15 – Activation of the economic trigger of strategic reserve.

2.7.2 Technical trigger

The second trigger for activating the strategic reserve is based on the continuous monitoring of the grid situation carried out by Elia. The first evaluation occurs in day-ahead no earlier than 18:00, when Elia disposes of the relevant information to perform a reliable analysis and the technical trigger can be activated until 4 hours[19] before real-time. However, Elia can exceptionally resort to using strategic reserve closer to real time, taking into account the activation time of the units participating in the strategic reserve, if this activation allows avoiding the load shedding of end-users who are not paid during this outage.

Contrary to the activation by economic trigger that results from an automatic process, the activation by technical trigger is the outcome of a whole decisional process in which the human contribution is important.

2.8 Impacts of renewable generation

A mechanism of Green Certificates (GC) was introduced in Belgium in order to promote the electricity generation from renewable sources, namely solar, wind, hydraulic, biomass and cogeneration. The purpose of the GC mechanism is to increase the share of renewable energy in the total electricity generation. Indeed, currently, despite the absence of fuel costs, green energy remains more expensive than conventional generation (fossil and nuclear). This can be explained by the intrinsic costs related to the technology as well as their decentralized nature potentially requiring expensive grid connection costs. However, such technologies are generating no or very few CO_2 emissions and do not require handling toxic waste.

The principle of the support mechanism is to provide an additional income to green electricity producers through Green Certificates for each MWh generated. The financial value of these certificates is not fixed and is subject to its own tailored market. Indeed, the electricity consumers (by the intermediary of their supplier) have to provide a number of GC proportional to their consumption. Hence, if they do not have their own renewable generation, they have to buy these GC to green electricity producers. In case of lack of GC in the market, their price will rise and ultimately foster investment in renewable generation. Since the percentage of GC that consumers have to provide is fixed by the authorities, the mechanism provides a regulatory framework allowing to progressively converge towards this given percentage of green electricity generation at the scale of the considered area. Moreover, with this system, the extra costs related to renewable generation are distributed among all consumers by virtue of the fact that their emergence is beneficial for everyone, in terms of both environment preservation and long-term security of supply.

The value of the GC is not fixed and fluctuates with respect to supply and demand. It is actually the demand that is boosting supply since the quota of GC is increasing each year, which constitutes the main asset of the system. However, the Walloon government has set up a complementary mechanism designed to ensure a minimum price for Green Certificates. The local transmission system operator Elia, as part of its public service task, has indeed the obligation to purchase each extra GC at a minimum price of 65 €.

[19] This duration is the reasonable period for estimating accurately the risk of structural shortage taking into account of the time necessary for market players to submit nominations in Intraday.

The Green Certificates are awarded to certified green electricity producers in proportion of the amount of generated energy. However, the number of received GC also depends on two criteria, i.e., the estimated additional costs of the technology as well as its environmental performance. In this way, the green electricity production is compared to the generation of a combined cycle gas turbine (CCGT) plant, which specifically emits 456 kg of CO_2 for each MWh of electricity. Concerning out-of-the-water hydroelectricity and wind turbines, the generation of 1 MWh of electricity entitles receiving 1 Green Certificate. However, the amount of CO_2 emissions generated during the preparation of the biomass reduces the number of perceived GC, whereas the heat production allows cogeneration to receive more than 1 GC for each electrical MWh.

For enhancing attractiveness of residential photovoltaic (PV), the Walloon government had originally (in the 2000s) applied a multiplying (>1) coefficient to the number of GC for each generated MWh. This coefficient first allowed convincing the most skeptical[20] to invest in the technology but overall annihilated the main asset of the whole support mechanism by considerably inflating the supply of GC, thereby preventing the system from converging towards the targeted percentage of renewables. In order to solve this issue, since the 1st March 2014, the photovoltaic installations in Wallonia of a power ≤ 10 kVA are subject to the Qualiwatt plan and can no more pretend to Green Certificates. With the Qualiwatt plan, the new residential PV installations are profiting of an annual premium during 5 years paid by the local DSO. The amount of the premium is fixed so that a return on investment can be achieved in 8 years.

2.8.1 Impact on market prices

The financial incentives granted to the renewable generation for ensuring their profitability has biased the standard law of supply and demand. Indeed, due to the GC mechanism that remunerates renewables for each MWh generated, it is more profitable for such units to sell the electricity at low prices than to be off-the-market (i.e. curtailed). The increase of renewables has therefore progressively driven down the electricity prices on electrical energy markets to the point of dropping below the profitability threshold of some conventional plants. For instance, in the wholesale markets, the price of electricity is around 40 €/MWh whereas the fuel cost for producing the same energy for gas-fired units is worth 50 €. As represented in Figure 2.16 for the day-ahead market, this may even result in negative prices at certain times of low demand.

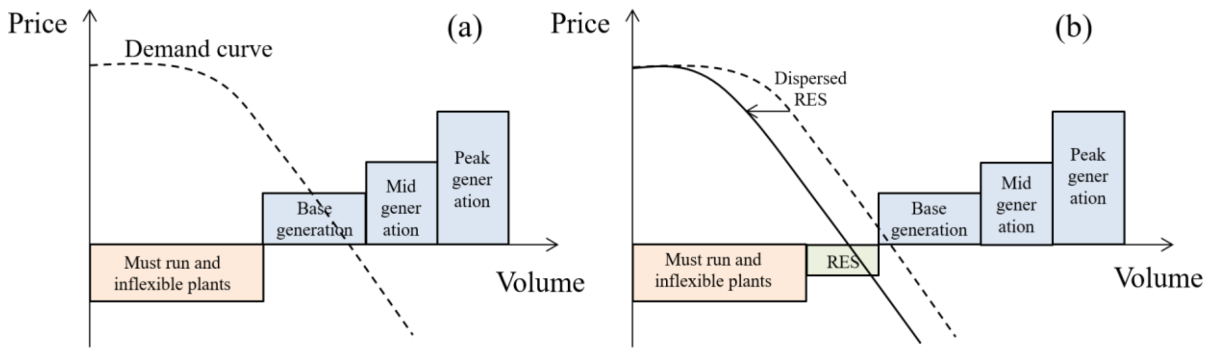


Figure 2.16 – Day-ahead market clearing without (a) and with renewable energy sources (b) in case of low load.

[20] The Green Certificates, added to the regional bonus for reducing installation costs, the reduction of the electricity bill and the absence of taxes for the injections towards the grid, made the investment highly (too much?) lucrative.

2.8.2 Impact on balancing services

With the growing share of stochastic (and difficult to predict) renewable generation, the energy exchanges are carried out closer to real-time, which is reflected by the increased importance of the day-ahead market. In this respect, the European TSOs expect a growing need of balancing reserves (as well as the creation of an inertia service with a response time much lower than FCR) to alleviate real-time imbalances.

If these renewable sources progressively replace conventional units, less reserve capacity will be available. In this context, solutions such as (big centralized and small-scale decentralized) storage utilities as well as demand response strategies are envisaged to alleviate this issue. However, renewables could also be part of the solution and provide cost-effective ancillary services by relying on an adequate control of the power electronics devices that are located at the interface between the installation and the electrical grid. It has indeed been demonstrated that wind turbines are perfectly adapted to quickly modulate their output power, and even deliver bidirectional (upward and downward) reserves if they are operated below their maximum power point [Martinez[12]].

The procurement of frequency control by wind turbines is however very limited or even non-existing in most countries, mainly due to two reasons. The first one originates from the financial incentives such as Green Certificates that are remunerating renewables for each MWh generated. Indeed, this external revenue combined with the profit realized by selling energy to consumers currently exceeds what renewable technologies could obtain from delivering ancillary services. Then, the contribution of renewable generation in frequency regulation is hampered by the regulation framework. Indeed, the time resolution for the reservation of balancing services is typically of one day and flexible resources have to be fully available during this 24 hours period, which often turns out to be prohibitive for power production with high volatility. Moreover, in most countries, the ancillary services reservation currently takes place in mid-term (typically in week- or month-ahead), which prevents them from efficiently contributing since they have limited knowledge on the daily energy that will be available with sufficient reliability (due to their poor predictability over such a long horizon).

To overcome these barriers and to foster participation of energy-constrained technologies (storage, demand response and modulation of RES), major stakeholders related to European electricity markets (e.g. energy regulators and political representatives of each country) gathered in June 2015 in Florence and have identified important guidelines for the future operation of electricity markets.

Specifically, it was agreed that the reservation of balancing capacity should shift towards shorter time horizons, from mid-term to day-ahead [ECR[17]].

Then, in order to bring costs further down, innovative sizing and allocation strategies for balancing services such as dynamic (hourly) reserve sizing (e.g. high forecasted RES output power involves higher upward reserve requirements) need to be encouraged.

Moreover, it was emphasized that renewable power producers should follow the same regulation than conventional generation and compete with them in the current liberalized environment without any external support [ECR[15]].

2.8.3 Impact on energy market design

In order to reach the European climate and energy targets, it is necessary to change the current electricity generation mix (to reduce pollutant emissions). However, the current market design might not be tailored to properly facilitate this transition.

Currently, most electricity markets (including Belgium) are organized as energy-only markets, where generators are only remunerated for the energy actually sold to the market (expressed in €/MWh during a specific period). Assuming perfect competition (no player exercises market power to deviate the prices for its own individual benefit), the electricity price is fixed by the equilibrium of supply and demand, which is equal to the variable generation cost of the most expensive units cleared by the market. In this way, power plants with relatively low variable costs can realize an inframarginal rent (Figure 2.17) whereas a peak load unit with higher variable costs (that determines the market price) only cover their marginal costs[21]. The number of (profitable) operating hours of these plants is not sufficient to cover their investment and maintenance costs, resulting in their shut-down (which is associated with a dangerous decrease of the total generation adequacy). In Belgium, as early as winter 2012-2013, the total import capacity of 3500 MW was at certain periods fully used to cover shortages.

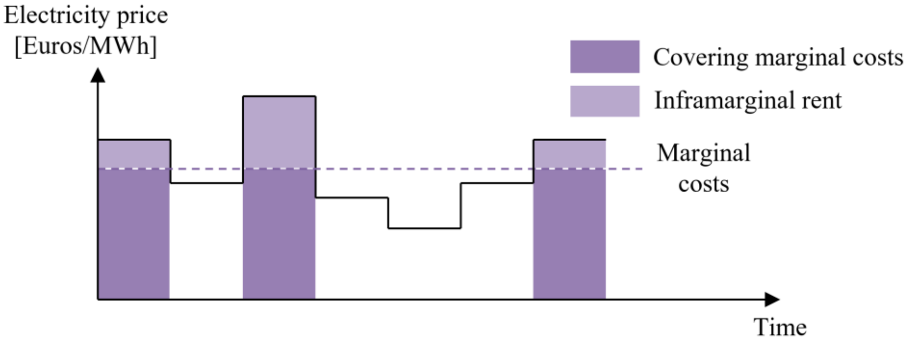


Figure 2.17 – Return on investment of conventional generation.

This issue resulting from the energy-only market structure can be addressed through adequate capacity mechanisms. The principle is to generate an additional source of revenue, which values the installed generation capacity, in addition to the profit realized in the energy market (inframarginal rent). The objective is to obtain a generation mix that allows compensating the unpredicted fluctuations of the renewable-based generation in the short-term while ensuring the system adequacy in the long-term.

At a longer time horizon, in a market with only renewables, the clearing prices can no longer be determined by the marginal cost of renewable power plants (which is close to zero) since it would lead to capital cost recovery problems for renewable producers. A new price-setting mechanism has thus to be designed and implemented to ensure stability of the system.

2.8.4 Impact on network operation – what markets cannot see…

The large-scale integration of renewable generation raises questions about the operation of the electric power grid in terms of stability regarding both frequency and voltage controls.

[21] Inframarginal rents are needed for power plants to recover from fixed generation costs.

In traditional (directly coupled synchronous generator based) power systems, oscillations of the frequency are, in a very short time frame (less than 1 second), directly counteracted by the inertial response of the directly-connected synchronous generators (by using their kinetic energy). In the absence of such units, the transient stability may thus be jeopardized. This issue has been studied and solutions have been proposed (for wind power-electronic connected units). Indeed, in such systems, the inertial behavior can be emulated through an appropriate control scheme using the rotating inertia of the turbine, i.e., by the so-called synthetic inertia.

In contrast with the network frequency, the voltage is a vector quantity in the sense that it varies across the different nodes of the grid. Maintaining an acceptable voltage plan is an important task since both under- and over-voltages can lead to operational issues and even be destructive for the electrical equipment. In this way, a variable-speed industrial motor can be cut out for voltage variation around ± 15 %. In case of loss of conventional capacity, this voltage control needs to be provided (cost-effectively) by other resources. In order to avoid installing costly devices such as flexible alternating current transmission systems (FACTS), the voltage control can be performed by the injection of reactive power from renewable technologies, which has an effect on the dimensioning of both the components and the grid connection.

Finally, since the conventional synchronous generators are connected in parallel to the power system, they allow decreasing the resulting short-circuit impedance Z_{sc} throughout the grid. Hence, with the loss of such generators, this short-circuit impedance Z_{sc} will increase, thereby increasing the short-circuit power P_{sc} (inverse of short-circuit impedance Z_{sc}) across the network. This value of P_{sc} is an image of the grid sensitivity to perturbation (higher values of P_{sc} are associated with a more robust and insensitive electrical network). One should thus remain careful when integrating renewables, and should above all strive to guarantee a smart and controlled energy transition.

2.8.5 Peer-to-peer energy trading

The roll-out of renewable generation distributed throughout the system has raised the opportunity for all end-users (regardless of their size) to directly trade energy between themselves, by allowing consumers to take advantage of the unused energy generated by other users. This mechanism, known as peer-to-peer energy trading, is further boosted by the increased accessibility of small-sized storage applications (due to significant costs reductions), and enable to bypass retailers (and the costs associated with the margin of profit of these middlemen).

The main benefit of peer-to-peer exchanges is the improved differentiation among the different products. In this way, the players can set their own terms (regarding price), and have the freedom to trade energy with friends or family, to choose a particular technology (such as solar or wind) and to prioritize local generation.

This paradigm can be even pushed further with a cooperation between these actors with the aim of intelligently manage all small and decentralized systems to relieve network issues and ensure the energy balance.

2.8.6 Microgrid paradigm

In the same line as the operational issues associated with the integration of renewable generation, the development of microgrids[22] needs to be carefully studied.

The main driver of such microgrids is not environmental (e.g. for consuming the energy more locally since the Kirchhoff's law governing the electrical flows were already ensuring that), but is purely economic. The main business model is to offset the load and generation locally so that the electrical bill is only related to the residual load of the microgrid. This enables the actors to considerably decrease their invoice by bypassing part of the costs related to grid fees and taxes. Indeed, the total grid fees and taxes (at the country level) are distributed among the different client connected to the grid, in proportion to their consumption/generation. By netting the profiles of a group of clients, the aggregated energy exchanges lead to grid fees and taxes significantly lower than the sum of individual contributions. The problem is that the earnings realized by the microgrid have to be re-distributed among the rest of the grid customers, ultimately benefiting the players able to constitute a microgrid (at the expense of small actors with limited financial resources).

Overall, it is important that the way towards improved (techno-economic) efficiency in power systems do not prevail over the concept of solidarity.

2.9 Interactions of electricity markets with other energy markets

The variable costs of conventional power plants depend on both fuel costs and the need to purchase CO_2 emission allowances for their CO_2 emissions. The amount of CO_2 generated during combustion is a function of the carbon content of the fuel. In this way, natural gas is emitting around 502 kg CO_2 for each MWh of electricity, compared with 987 kg for hard coal and 1,170 kg for brown coal [EIA[18]]. The CO_2 emission allowances are traded on international exchanges (centralized market with transparent prices).

Hard coal, natural gas and crude oil are traded on global markets and have thus a transparent price. Lignite or uranium on the other hand are not traded on global markets, which makes their prices much less transparent. This originates from the transportation costs of lignite that are so high that the lignite power plants are usually located closely to the lignite pits. For uranium, legal conditions restrict mining and trading.

The market prices for natural gas, hard coal and CO_2 emissions for 2015-2016 are illustrated in Figure 2.18. It can be seen that the market related to CO_2 emissions allowances is completely malfunctioning, with very low prices that do not properly penalize coal power plants, allowing countries such as Germany to offer very competitive prices (but with deleterious environmental effects). It can nonetheless be observed that the CO_2 price was more volatile in 2016 than in 2015, which can be attributed to the uncertainty associated with European political decisions in the light of the Paris agreement.

[22] A microgrid is a localized energy system composed of loads, electrical generation and flexibility resources (such as storage) that operates as a single actor connected to the main grid. It can even be disconnected from the backbone system (islanded mode), and operates autonomously (very rare in practice due to the technical challenges).

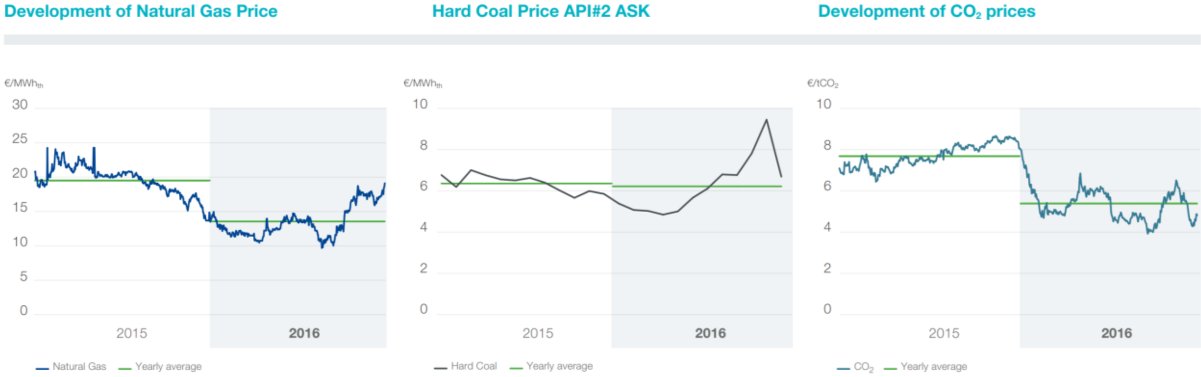


Figure 2.18 – Daily day-ahead gas prices from EEX, monthly hard coal prices from API#2 ASK(CIF ARA), and daily CO2 futures prices, all traded through 2015/2016 [TenneT[17]].

2.10 Conclusions and perspectives

This chapter was devoted to describe the general situation (particularized to the Belgian case) resulting from the liberalization of electricity markets. In this way, after the construction of the first power plants at the end of the 19th century and the subsequent implementation of localized small-scaled electrical systems, the grid progressively became more robust and interconnected with a centralized, vertically integrated structure. Then, in the second half of the 1990s, the European Union has decided to deregulate electricity markets with the aim of ensuring fair prices and improve security of supply, by introducing competition at both generation and retail levels. Such a regulation, combined with a strong will to move towards decarbonisation of the overall energy system, has facilitated the development of renewable-based generation, typically from wind and photovoltaic sources. However, their stochastic behavior is inferring an increased need of flexibility in the system, need that has triggered technological advances, most noteworthy regarding storage utilities and electric vehicles. Consequently, numerous actors of varying sizes (from smart buildings to big industrial estates) have seen an opportunity to become more resilient with respect to energy shortages during critical periods, while bypassing grids fees and related taxes by developing autonomous areas with a self-centered energy management. The general structure of electricity markets can be summarized as represented in Figure 2.19, with the co-existence of energy, ancillary services and capacity mechanisms to ensure stability of the system from a long-term perspective to real-time considerations.

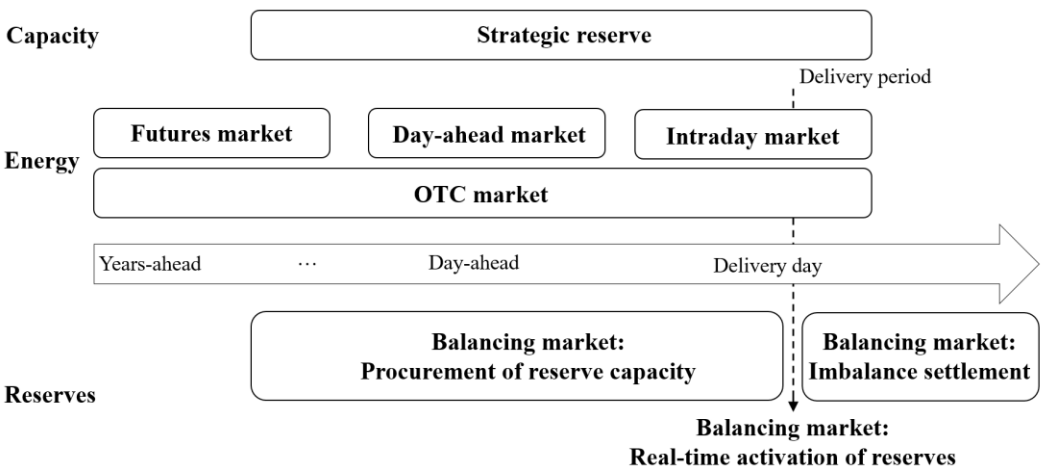


Figure 2.19 – General structure of electricity markets.

In the following years, the main challenge for the power industry will be to cleverly harmonize and accommodate the current regulation policies so as to favor the generalized implementation of smart solutions for integrating the renewable-based generation and promoting flexibility within the grid, while preserving the legacy infrastructure (so as to safeguard security of supply). The latter constitutes indeed not only a resilient backbone structure guaranteeing solidarity among interconnected areas, but also a reliable system allowing energy exchanges between remote locations, and, as a result, global energy efficiency at a large-scale. The regulation should therefore create a favorable environment ensuring a level playing field among all technologies that enable all actors to take advantage of their resources.

However, this development of active customers at all layers of the system leads to the question of how will the levels interact. More particularly, how to foster an efficient energy systems integration with adequate roles for each actor is a key issue to address in the following years. Indeed, the end-user (residential client) is not, in essence, a genuine economic stakeholder of the energy sector and it is important to protect him from professional actors while increasing its awareness of energy challenges to improve its daily behavior and stimulate him to contribute at a wider scale (e.g. through cooperatives, aggregators, etc.).

The market design should moreover remain flexible enough to leave the door open to new technical breakthroughs. The energy sector is indeed constantly changing and the whole system must be tailored and prepared to adapt to both expected and unexpected evolutions. In this regard, the further digitalization of the energy system will break the barriers hindering the development of active consumers by facilitating information exchanges.

Globally, underestimating the importance of the regulatory framework may potentially lead to adverse situations. Firstly, the energy transition should be carefully monitored with regard to current need and sufficient investments in conventional power plants constituting the basis for stability and inertia of the grid should be maintained as long as necessary. Secondly, a significant development of energy-autonomous areas may jeopardize the viability of the current infrastructure and eventually cause its progressive dismantlement if poor incentives for global ancillary services are provided. Finally, the way towards improved efficiency in power systems should not prevail over the concept of solidarity nor negatively impact the bill of end-users. Indeed, although it is important that motivated and contributing (load-responsive) actors are rewarded, there is a strong responsibility of decision-making authority to protect the most exposed part of the population. One should not forget the original purpose of electrical grids, namely **provide energy to all end-users at a competitive price**.

This chapter aimed thus at giving an overview of the liberalized context currently governing the electricity sector in Europe (with a particular focus on the Belgian specificities). This knowledge is indeed essential to take full advantage of the economic potential of a portfolio participating in electricity markets (chapters 4 to 6). However, it is also necessary to properly account for the numerous uncertainties associated with the optimization procedure (so as to ensure robustness of the decisions), and the prediction of such fluctuating and uncertain variables is addressed in the next chapter.

CHAPTER 3

SHORT-TERM MULTIVARIATE PROBABILISTIC FORECASTING

3.1 Introduction

Our society is currently undergoing a major energy transition, mainly driven in the electricity sector by an increased penetration of renewable energy sources, an improvement of the energy, and reduced emissions of greenhouse gas. In parallel, in order ensure the success and foster this energy transition, European Union has decided to set up a competitive environment (liberalization of the electricity sector) with underlying mechanisms encouraging the actors to invest and act towards decarbonization. As a consequence, intricate dependencies have been developed between electrical and market data, and this relationship strengthens over time.

This context gives rise to complex stochastic optimization problems, not only for system operators but also for all the other players that have indeed to operate within a complex (multiple platform markets) and uncertain (renewables sources, load, market data) environment. Hence, the success of their planning strategies, and as a corollary of an affordable energy transition, strongly relies on the knowledge of the system state (through adequate prediction tools) at time horizons which go from day-ahead to quasi real-time [Pinson[07]- Thatte[13]].

However, due to the complex nature of the signals of interest (arising from their non-stationarity and nonlinear nature), predictions over such time horizons are ineluctably vitiated by errors. The uncertainty mainly originates from noise in the explanatory variables (e.g. due to the chaotic nature of the weather system) as well as model misspecifications. Hence, traditional point (deterministic) forecasts that only predict the conditional mean of the signal are providing very limited information to decision-makers. Indeed, in order to ensure decisions that are robust with regard to forecast errors and unexpected events, it is also necessary to quantify the level of uncertainty associated with predictions.

In this way, different approaches for obtaining such uncertainty regions (or by extrapolation densities) can be found in the literature for respectively wind power [Felder[10], Kavousi-Fard[16], Pinson[10], Quan[14], Wan[14], Zhang[14]], PV generation [Wan[17]], load [Khosravi[10], Quan[14]] and electricity prices [Shrivastava[15], Zhao[08]]. Then, techniques such as robust, interval and chance-constrained optimization were developed to hedge against uncertainties, by relying on probabilistic forecasts. However, these optimization techniques have two main drawbacks. Firstly, robust and interval techniques are known to yield conservative (and thus sub-optimal) solutions since these are intrinsically designed to be optimal with regard to extreme scenarios [Bruninx[16]]. Chance-constrained optimization offers a less conservative and more practical approach by considering a probability for satisfying each constraint, but such a formulation is very difficult to solve in practice (due to the non-convexity of the resulting problem). Secondly, since the uncertainty characterization (under the form of intervals) provide little information on the intertemporal relationship between time steps (i.e. actual time variations, or ramps), the quality of subsequent decisions may be affected [Pinson[08]].

Consequently, for time-dependent decision problems that have to be carried out on a regular basis (such as the daily participation in electricity markets), scenario-based stochastic optimization provides a practical framework that yields efficient (less conservatives) outcomes in general [Morales-Espana[14]]. But this technique, which optimizes the expectation of some loss function (e.g. profit of an electricity retailer) under a forecast distribution, can be associated with tractability issues, depending on the number of scenarios (time trajectories) used to represent uncertainties. In this respect, implementing a methodology able to provide a limited set of representative scenarios is highly valuable [Dupacová[03], Gröwe-Kuska[03]].

This problem is tackled in [Morales[10,a]], where an autoregressive moving average (ARMA) model is developed for individual wind sites, and a stationary variance-covariance matrix is thereafter used for integrating the spatial correlation among series. However, this approach does not allow to properly take into consideration the uncertainty associated with particular conditions (e.g. higher uncertainty during strong winds). In [Ma[13], Papaefthymiou[08], Pinson[08]], the scenarios are constructed by computing the *complex* covariance matrix based on a multivariate Gaussian distribution assumption. In [Quan[15]], a nonparametric neural network dedicated to the quantification of prediction intervals (PIs) is firstly implemented. Then, these PIs are used to estimate an empirical cumulative distribution function, from which scenarios are generated. But, due to the independent nature of the sampling methodology, these scenarios do not account for the time-varying structure of forecasts errors.

In this work, this issue is overcome using a copula-based strategy to sample the multivariate distribution originating from probabilistic forecasts. This allows to generate scenarios that comply with both the predicted distributions and interdependence structure of variables. Overall, the main contributions of the chapter can be summarized as follows.

Firstly, **in order to increase the predictive capability of (both point and probabilistic) forecasts, the work capitalizes on recent breakthroughs in Deep Learning (with the use of neural networks with improved memory) to generate more accurate multi-step ahead forecasts.** Specifically, we make use of deep Bidirectional Long Short Term Memory (BLSTM) neural networks, a particular type of recurrent architecture with rich dynamics, designed to automatically select and propagate through time the most relevant contextual information. The results demonstrate that this approach infers lower forecast errors with regard to traditional techniques and are well suited for time series forecasting, which allows to reduce the uncertainty space of the subsequent decision making problem.

Secondly, **concerning probabilistic forecast, two different models for characterizing the uncertainty are compared. In this way, the BLSTM network is trained**

to either generate a Gaussian [Flunkert17] or a non-parametric predictive distribution of the dependent variables [Bremnes04, Nielsen06, Wen17]. It enables to confront the Gaussian assumption of prediction errors with an empirical approach (that makes no assumption on the underlying probability distribution of variables).

Thirdly, **although the method can provide prediction intervals and densities, it is here extended with the aim to provide predictive scenarios.** Practically, the tool relies on a copula-based sampling of the multivariate forecasted distribution so as to generate time trajectories (sample paths) that mimic actual time and cross-variable dependencies. In this way, whereas most of the literature focuses on individual variables, the proposed approach attempts to exploit information in a multi-dimensional context with heterogeneous data from different natures. Indeed, in the competitive framework governing the current electricity sector, complex dependencies between electrical and market data are taking shape, and it is thus important to implement a strategy that is able to capture this information.

Fourthly, **the value of the methodology is compared with other approaches not only in terms of statistical performance, but also regarding the practical impact of the quality of scenarios on the decisions optimized within a scenario-based stochastic optimization tool.** Here, the day-ahead scheduling of electricity aggregators (such as energy retailers or generation companies) in a multi-market environment is used as a case study.

Moreover, thanks to the self-learning nature of the proposed methodology, minimal manual engineering or data pre-processing is needed. Then, within the objective of quickly and efficiently integrating the new information that is revealed each day, the method is developed in such a way that the models can be dynamically adapted using exclusively new data. This step circumvents the need of retraining the global architecture from scratch with the whole set of historical data.

The rest of the chapter is structured as follows. Firstly, an introduction to neural networks, with an emphasis on recurrent architectures, is provided. Such mathematical models will indeed be used for the short-term (day-ahead) predictions of electrical and market data. Specifically, a deterministic (point) forecasting tool will firstly be developed, and the procedure will then be extended to generate probabilistic forecasts (under the form of densities). Then, the subsequent sampling policy implemented to generate predictive multivariate scenarios is thoroughly explained and motivated. Finally, the results illustrating the benefits of the proposed approach (with regard to traditional methods such as the multilayer perceptron) in terms of both statistical and impact on the quality of decisions optimized within a dedicated stochastic optimization tool are discussed.

3.2 Neural networks

It has been observed in the past decade that, due to improvement in computer capabilities, machine learning methods outperform the best physical models when a sufficient amount of historical data is available [LeCun15]. Generally, these methods are divided into three categories, depending on the desired objective. The first one is referred to as supervised learning, and consists in determining the hidden relationship between a set of inputs-outputs pairs, typically for regression or classification tasks. The objective is to bypass the (possibly complex) physical modeling of the underlying phenomena, by directly exploiting the information contained in the data. The second category is the reinforcement learning, where the purpose is to optimize the strategy of an agent in a given environment, which is achieved by assigning a positive or negative reward for each decision during training (based on its impact on the objective function). These two techniques differ thus from the unsupervised learning, where no task-specific target is defined. The algorithm rather attempts to infer the hidden structure of data, typically for clustering purposes.

For the prediction task in which we are interested, this work therefore relies on supervised learning, and more specifically on neural networks. Neural networks are indeed universal, i.e. theoretically able to model any complex function between pairs of inputs-outputs by capturing and replicating their hidden underpinning mechanisms [Hornik[89]].

3.2.1 Multilayer perceptron

The artificial neural networks (ANNs) were initially developed to mimic the processing properties of the human brain [McCulloch[88], Rosenblatt[63], Rumelhart[86]]. In this way, their structure is composed of several processing units (or neurons), connected to each other by weighted connection (or synapses) so as to mathematically represent any relationship between inputs (explanatory variables) and outputs (dependent variables of interest).

Different architectures of ANNs have been developed over the years, with the most important distinction associated with the cyclic nature of connections. ANNs without cycles (acyclic connections between neurons) are known as feedforward neural networks, among which the most popular are the radial basis function networks [Broomhead[88]] but also the multilayer perceptron [Rumelhart[86]], which will be further described in this section. ANNs whose connections create cycles are referred to as recurrent neural networks, and will constitute the main focus of Section 3.2.2.

It is important to clearly differentiate the two main stages when making use of neural networks, i.e. its *online utilization* (section 3.2.1.1), and the preceding *training phase* (Section 3.2.1.2). Neural networks are trained by using the historical datasets (of both explanatory and dependent variables) within a supervised learning strategy that adjusts the model parameters (weights between neurons) with respect to the desired task (e.g. maximize the prediction accuracy of the network). This learning procedure generally requires significant (time and space) computational resources. Then, when the network is trained, it can be stored so as to be used online for real-time applications (with very low calculation times).

3.2.1.1 Network online utilization

As represented in Figure 3.1, the neural processing units in a multilayer perceptron (MLP) are organized in layers, with forward connections from one layer to the next one. The inputs (explanatory variables forming the input layer of the network) are provided to the first hidden layer of the network. These neurons are activated and propagate this activation through all the hidden layers along the synapses to the output layer (*forward pass* of the network). By modifying the connection weights, a single neural network architecture is therefore capable of modeling different functions.

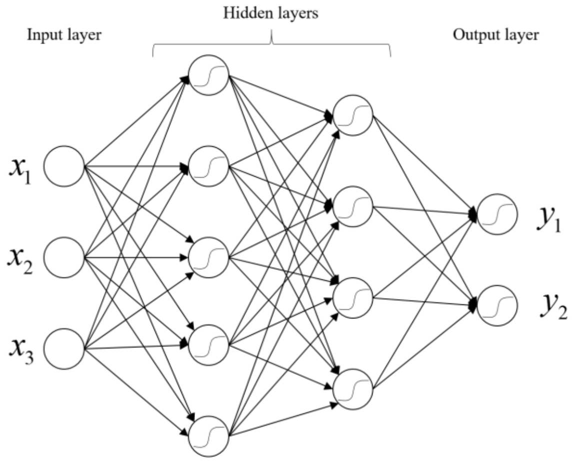


Figure 3.1 – General structure of a multilayer perceptron (MLP), in which the sigmoid-shaped curves in the neural processing units indicate the application of the sigmoidal activation function (but any other function can be employed). In practice, different functions are used for hidden and output layers.

The general form for the activation of neurons is represented in Figure 3.2.

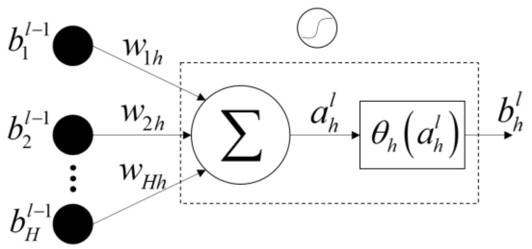


Figure 3.2 – Activation of a neural processing unit.

Let us consider a MLP hidden layer $l \in [1, L]$, activated by a vector $\mathbf{b}^{l-1}$ composed of the H outputs of the previous layer. For the first hidden layer $l = 1$, the bottom layer corresponds to the network input vector $\mathbf{x}$. The final activation b_h^l of each unit h of the layer l is determined by the following equations:

$$a_h^l = \sum_{h' \in H_{l-1}} w_{h'h} b_{h'}^{l-1} \tag{3.1}$$

$$b_h^l = \theta_h \left(a_h^l \right) \tag{3.2}$$

where $w_{h'h}$ is the weight from unit h' to unit h, and θ_h the activation function, for which the most common options are the hyperbolic tangent and the logistic sigmoid. These functions, which are represented in Figure 3.3, are indeed characterized by two interesting features.

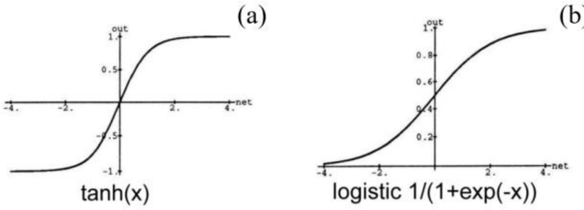


Figure 3.3 – Traditional activation functions of multilayer perceptron [Graves[12]]: hyperbolic tangent (a), and logistic sigmoid (b).

First, these functions are nonlinear. Indeed, any combination of linear operations remains linear, which contrasts with nonlinear combinations that can significantly improve the processing capabilities of the neural network (which is achieved by using successive hidden layers of nonlinear operators so as to re-represent the data [Hinton[06]]). Then, these functions are differentiable, and the network can thus be trained using traditional gradient descent algorithms (see Section 3.2.1.2).

The output vector $\mathbf{y} = [y_1, \ldots, y_K]$ of the MLP is obtained following the activation of the K neurons constitutive of the output layer. Its activation function generally differs from the one associated with hidden layers since the output layer is closely linked to the task for which the network is applied to. In this way, the standard configuration for binary classification tasks is a single output unit with a logistic sigmoid activation (for which the range is included within the [0, 1] interval), and the output can consequently be viewed as the probability that the input pertains to the first (of both) considered class. Then, for regression tasks (such as predictions where the objective is to estimate the relationship between the dependent variables at future time steps with the available information), a linear activation function (3.3) is often privileged.

$$y_k = a_k = \sum_{h \in H_L} w_{hk} b_h^L \tag{3.3}$$

where y_k is the k^{th} output of interest, and b_h^L is the output of unit h of the last hidden layer H_L.

3.2.1.2 Neural network training

The objective when training neural networks is to find the optimal weight values (connections) between neurons that minimize the error on the output (e.g. for deterministic forecasts, minimize the difference between the predicted values and the actual observations). Since MLPs are composed of differentiable operating units, they can be trained to optimize any (also differentiable) loss function using gradient descent.

After selecting a loss function Λ suitable for the considered task, the principle of gradient descent is to determine the derivatives of this loss function with respect to each weights of the network, and to subsequently adjust the weights in the direction of the negative slope (that minimizes the loss function).

In this respect, error backpropagation is to this day one of the most important achievements for training neural networks [Rumelhart[86]]. The methodology can be summarized by the four following steps:
Step 1 - *Forward pass*: Given inputs and current weights values, the outputs are computed by propagating activation of units throughout the network;
Step 2 - *Loss function*: The outputs are compared with actual observations using a pre-defined error function Λ;
Step 3 - *Backward pass*: The partial derivatives of the loss function with respect to each of network weights are computed;
Step 4 - *Weights update*: The weights (network parameters) are adjusted with the standard equation for gradient descent.

The backward pass (Step 3) is computed by a repeated application of the chain rule for partial derivatives. To that end, the first step is to calculate the derivatives of the loss function with respect to the output units. Then, we recursively apply the chain rule, working backwards through the hidden layers. In this way, for the connections between the last hidden layer and the output units:

$$\frac{\partial L}{\partial w_{hk}} = \underbrace{\frac{\partial L}{\partial y_k}\frac{\partial y_k}{\partial a_k}}_{\delta_k}\underbrace{\frac{\partial a_k}{\partial b_h^L}}_{w_{hk}} \tag{3.4}$$

At this stage, it is therefore convenient to introduce the following notation:

$$\delta_h^l = \frac{\partial L}{\partial a_h^l} \tag{3.5}$$

Concerning the weights between the penultimate hidden layer L-1 and the last one L, the partial derivatives are obtained as follows:

$$\frac{\partial L}{\partial w_{h'h}} = \left(\sum_{k=1}^{K}\underbrace{\frac{\partial L}{\partial y_k}\frac{\partial y_k}{\partial a_k}}_{\delta_k}\underbrace{\frac{\partial a_k}{\partial b_h^L}}_{w_{hk}}\right)\underbrace{\frac{\partial b_h^L}{\partial a_h^L}}_{\theta'\left(a_h^L\right)}\underbrace{\frac{\partial a_h^L}{\partial w_{h'h}}}_{b_{h'}^{L-1}} \tag{3.6}$$

The partial derivatives between layers l and l+1 can then be calculated recursively:

$$\frac{\partial L}{\partial w_{hh'}} = \left(\sum_{h'\in H_{l+1}}\delta_{h'}^{l+1}w_{hh'}\right)\theta'\left(a_h^{l+1}\right)b_h^l \tag{3.7}$$

Once all the partial derivatives are computed, the weight update procedure (Step 4) is carried out. This consists in adjusting the weights values through a small step in the direction of the negative error gradient of the loss function:

$$w_{ij}^{it+1} = w_{ij}^{it} - \alpha\frac{\partial L}{\partial w_{ij}^{it+1}} \tag{3.8}$$

where w_{ij}^{it} is the weight between neurons i (at layer l-1) and j (at layer l) at iteration it, whereas the learning rate $\alpha \in [0, 1]$.

The whole procedure (steps 1-4) is repeated until some stopping criterion (e.g. failure to decrease the loss function for a given number of iterations) is reached.

3.2.2 Recurrent neural networks

In the latter section, networks whose neural connections did not form cycles were considered. However, the dynamic nature of some electrical quantities (photovoltaic generation, etc.) cannot be optimally modeled by such static neural networks that do not capture the influence of previous states. A solution consists therefore in collecting the inputs into overlapping time-windows (i.e. relevant past information is incorporated as additional inputs of the network), and treating this task of capturing time dependencies as spatial (data from all time steps are concatenated to form a single input vector to the neural network).

This procedure can be performed by *time delay neural networks* (TDNNs), where the output vector **y** in time *t* is based on the inputs in times (t-1), (t-2), ..., (t-n_x).

$$\mathbf{y}_t = \mathrm{f}\left(\mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-n_x}\right) \tag{3.9}$$

Another similar approach can be viewed in the *nonlinear autoregressive exogenous* (NARX) model, where the output vector **y** in time *t* is based on both inputs and outputs at previous moments.

$$\mathbf{y}_t = \mathrm{f}\left(\mathbf{y}_{t-1}, ..., \mathbf{y}_{t-n_y}, \mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-n_x}\right) \tag{3.10}$$

Both these architectures necessitate to find an optimal time-window size, which is task-dependent and results from a complex trade-off between integrating sufficient temporal information while avoiding irrelevant data (too small window size and the network will neglect important explanatory information, too large and it will reduce the ability of the network to discern data noise from the relevant dependencies).

However, if the non-cyclical condition of feedforward networks is relaxed so that connections between neurons within a given layer are allowed, we obtain recurrent neural networks (RNNs), in which a sequential representation of data is then intrinsically embodied (Figure 3.4). Indeed, recurrent connections allow a natural memory of previous inputs that can thus propagate through time within the internal state $\mathbf{b}_\bullet$ of the network, resulting in a more compact and robust representation [Vermaak[98]].

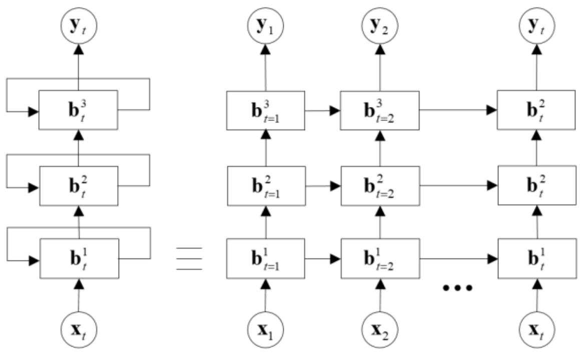


Figure 3.4 – Recurrent neural network architecture.

Such architectures are more representative of biological brains and represent an efficient generic tool, integrating both a powerful dynamical memory and computational abilities that can be very easily adapted to the complexity of the modeling task. Recurrent networks have shown high potential in processing sequential data by allowing past information to persist [LeCun[15]], and, in this regard, they constitute the natural class of neural networks for time series prediction.

The forward pass of recurrent networks is similar to the one of a MLP, with the exception that the activation of hidden layer arises not only from the output from the layer below but also from the hidden layer activation from the previous time step (thanks to recurrent connections). This enables past information to be propagated across time through the hidden layer. Let us consider an input sequence $\mathbf{x}(I \text{ x } T) = [\mathbf{x}_1,\ldots, \mathbf{x}_T]$ composed of I input units varying over T time steps, so that $x_{i,t}$ is the value of input i at time t. This sequence is presented to a RNN made up of L hidden layers (each one containing respectively H_l hidden units) and K output units, so that $y_{k,t}$ is the output k at time t.

For the first hidden layer $l = 1$, the input of unit h at time t is thus computed as follows:

$$a_{h,t}^{l} = \sum_{i=1}^{I} w_{ih} x_{i,t} + \sum_{h' \in H_1} w_{h'h} b_{h',t-1}^{1} \tag{3.11}$$

where $b_{h',t-1}^{1}$ is the activation of unit h' (pertaining to the first layer) at time t-1, whereas w_{ih} is the weight between input i and unit h, and $w_{h'h}$ is the weight between units h' and h (from the same hidden layer). Analogously, we have for deeper layers:

$$a_{h,t}^{l} = \sum_{x \in H_{l-1}} w_{xh} b_{x,t}^{l-1} + \sum_{h' \in H_l} w_{h'h} b_{h',t-1}^{l} \tag{3.12}$$

Similarly to MLP, a nonlinear (and preferably differentiable) activation function θ_h is then applied to obtain the activation state of the hidden units of the network:

$$b_{h,t}^{l} = \theta_h \left(a_{h,t}^{l} \right) \tag{3.13}$$

The output vector sequence $(\mathbf{y}_1,\ldots, \mathbf{y}_T)$ is then obtained by sequentially applying the following equation (which considers a linear activation function of the output layer) from time steps $t = 1$ to T:

$$y_{k,t} = \sum_{h \in H_L} w_{\mathrm{hk}} b_{h,t}^{L} \tag{3.14}$$

It should be mentioned that the procedure necessitates to initialize the network states before it receives information from the input sequence. In this work, b values are set to zero, but other nonzero initial values can also be used. Overall, the complete output sequence (forward pass) can thus be obtained by following this procedure:

For each time step $t = 1$ to T, knowing that $b_{h,\mathrm{t}=0}^{l} = 0 \ \forall h, l$

Compute the activation state of units of the first hidden layer by applying (3.11)-(3.13)
Compute, sequentially for each hidden layer $l > 1$, the activation state of units by applying (3.12)-(3.13)
Compute the network outputs by applying (3.14)

Whereas the additional cyclic connections of recurrent networks do not strongly impact the complexity of the forward pass, it is different for the backward pass, and RNNs are consequently more difficult to train with gradient descent. Two main algorithms have been developed to compute the desired partial derivatives of the loss function with respect to the network weights: real time recurrent learning [Robinson[87]], and backpropagation through time [Werbos[90], Williams[95]]. However, the second method emerged within the machine learning community since it is both simpler to implement and more effective in terms of computational time [Graves[12]].

Backpropagation through time (BPTT) applies the same methodology as standard backpropagation (i.e. recursive application of the chain rule). However, the activation of internal units (neurons) within layer *l* influences not only the next layer *l*+1, but also the same layer *l* at the next time step, which involves that:

$$\begin{aligned} \delta^l_{h,t} &= \frac{\partial L}{\partial a^l_{h,t}} \\ &= \left(\sum_{h' \in H_{l+1}} \delta^{l+1}_{h'} w_{hh'} + \sum_{h'' \in H_l} \delta^l_{h'',t+1} w_{hh''} \right) \theta'\left(a^{l+1}_{h,t}\right) \end{aligned} \tag{3.15}$$

where $\delta^l_{h,t=T+1} = 0 \; \forall \; l \in L, h \in H_l$, since no error originates from beyond the sequence.

Hence, the complete sequence of δ terms (backward pass) can be computed by applying (3.15) recursively from $t = T$ to 0.

Knowing that the same weights (connections between units) are used for each time step, the partial derivatives are summed over the whole sequence. For inter-layer weights (between layers *l* and *l*+1), we have:

$$\frac{\partial L}{\partial w_{hh'}} = \sum_{t=1}^{T} \frac{\partial L}{\partial a^l_{h',t}} \frac{\partial a^l_{h',t}}{\partial w_{hh'}} = \sum_{t=1}^{T} \delta^{l+1}_{h',t} b^l_{h,t} \tag{3.16}$$

For intra-layer weights (connections within the same layer *l*), the partial derivatives of the cumulative sequence error are obtained as follows:

$$\frac{\partial L}{\partial w_{hh''}} = \sum_{t=1}^{T} \delta^l_{h'',t} b^l_{h,t-1} \tag{3.17}$$

After presentation of the sequence and determination of the partial derivatives, the weights are updated using the standard equation for gradient descent (3.8), and the procedure (forward pass, backward pass, weight update) is iterated over the whole historical dataset until convergence on results is achieved.

Similarly to feedforward networks, different varieties of RNNs have been developed (i.e. different functions θ_h were proposed), such as *Elman* [Elman[90]] and *Jordan* [Jordan[90]] networks. Then, an alternative trend to train and use recurrent network has been introduced with reservoir computing, which are divided into *echo state networks* [Jaeger[01]] and *liquid state machines* [Maass[02]], depending on the way the hidden neurons are activated. These architectures are nevertheless prone to overfitting (see Section 3.2.3.2) and necessitate thus a complex optimization of their topological structure (complexity) [Lukosevicius[12]]. Moreover, their inner dynamics is limited to a narrow frequency band, which prevents them from adequately representing multiple periodicities in the data.

Overall, such traditional RNNs are characterized by two main limitations.

The first problem, widely known as the vanishing gradient problem, is that the back-propagated errors δ_h during training either fades or blows up over time due to the multiple gradient calculations associated with the backpropagation algorithm. In this way, sigmoidal activation functions (Figure 3.3 (b)) map its input into the small [0, 1] range, and any variation in the input will generate a very small change in the output (i.e. a small gradient). This phenomenon is exacerbated for recurrent networks that aims to propagate the information through time steps. Indeed, at the first time step, the hidden layer will map its input into a small region, the output of which will itself be mapped to a smaller region during the next time step, and so on. It has been shown that this process ultimately prevents the model from reliably accessing time dependencies more than a few time steps long [Hochreiter[01]].

Secondly, standard RNNs process inputs in temporal order and ignore the information contained in the future context. This results in an inadequate modeling of backwards dependencies, preventing the network from fully exploiting contextual information. This can be easily understood for speech recognition where the understanding of a word (or a phoneme) is improved after the whole sentence has been heard, and can also be extended to time series where a known future event may contain valuable information to explain previous values.

3.2.2.1 Long Short Term Memory (LSTM) networks

The first issue (vanishing gradient problem) is here tackled by using an alternative (more complex) neural architecture, referred to as Long Short-Term Memory (LSTM), which better controls the flow of information through the hidden layer by means of memory cells [Hochreiter[97], Gers[02]]. In this way, a LSTM layer $l \in L$ is made up of N_H recurrently connected blocks, known as memory blocks (or neurons). As represented in Figure 3.5, each block has three multiplicative units, known as input, output and forget gates, which can be seen as modules for respectively writing, reading and resetting information. The inputs of each layer l at time t are composed of the outputs of the same layer at the previous time step $\mathbf{b}_{t-1}^{l}$ as well as the outputs of the layer below $\mathbf{b}_{t}^{l-1}$. For the first hidden layer l = 1, the bottom layer $\mathbf{b}_{t}^{l-1}$ corresponds to the network input variables $\mathbf{x}_t$.

The activation function θ_h associated with the LSTM architecture consists in the following composite equations:

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_{i\iota}\mathbf{b}_t^{(l-1)} + \mathbf{W}_{h\iota}\mathbf{b}_{t-1} + \mathbf{W}_{c\iota}\mathbf{c}_{t-1}\right) \tag{3.18}$$

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_{i\phi}\mathbf{b}_t^{(l-1)} + \mathbf{W}_{h\phi}\mathbf{b}_{t-1} + \mathbf{W}_{c\phi}\mathbf{c}_{t-1}\right) \tag{3.19}$$

$$\mathbf{c}_t = \mathbf{f}_t\,\mathbf{c}_{t-1} + \mathbf{i}_t\,\tanh\left(\mathbf{W}_{i\gamma}\mathbf{b}_t^{(l-1)} + \mathbf{W}_{h\gamma}\mathbf{b}_{t-1}\right) \tag{3.20}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_{io}\mathbf{b}_t^{(l-1)} + \mathbf{W}_{ho}\mathbf{b}_{t-1} + \mathbf{W}_{co}\mathbf{c}_t\right) \tag{3.21}$$

$$\mathbf{h}_t = \mathbf{o}_t\,\tanh\left(\mathbf{c}_t\right) \tag{3.22}$$

where σ is the logistic sigmoid function, and $\mathbf{i}_t$, $\mathbf{f}_t$ and $\mathbf{o}_t$ are the activation vectors of the input gate, forget gate and output gate respectively, whereas $\mathbf{c}_t$ stands for the cell activation vector. All these vectors are of similar size, equal to the one of the hidden vector $\mathbf{b}_{t-1}^{l}$ (i.e. output vector of the hidden layer l).

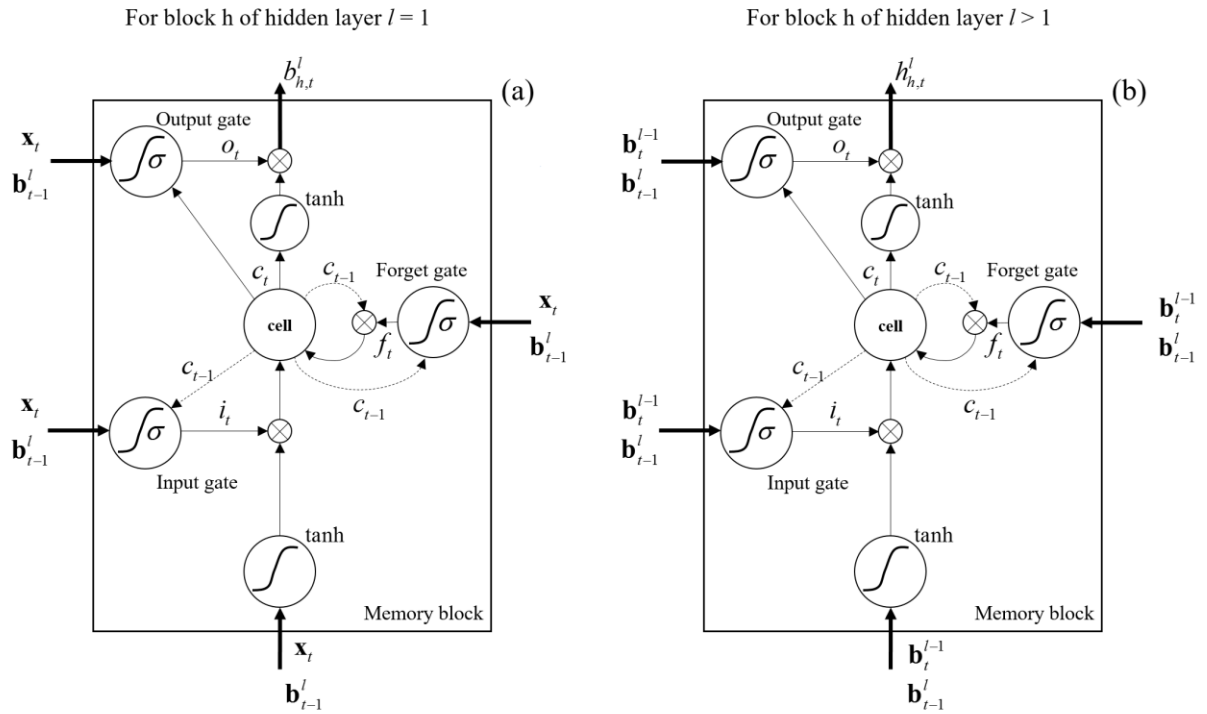


Figure 3.5 – Single-cell LSTM memory block (cell *h* of hidden layer *l* at time *t*) used in this work.

Theoretically, LSTM networks can capture both long and short periodicities of time series. Indeed, each LSTM memory block is able to adaptively memorize, forget and expose its memory content. In this way, if the current information stored in the memory cell is identified as important by the network, the forget gate will ensure that it is propagated over time, which amounts to model a long-term dependency. Consequently, periodicities can be adequately modeled by exposing the memory content when a relevant input feature is observed. Contrariwise, irrelevant information (whose effect has completely faded over time) can be reset by opening the forget gate. Since both operating modes (propagate or eliminate past information) can simultaneously occur among the different LSTM blocks within each hidden layer, the LSTM network is potentially able to model any complex signals with multiple time scales.

However, it has been shown that modeling intricate time series with multiple periodicities may prove to be more difficult in practice, even with LSTM networks [Sugiartawan[17]]. In [Jaeger[07]], a hierarchical architecture is implemented, in which the raw signal is decomposed into different dynamical features, each one with a specific frequency band. Each of the resulting components is then processed by a particular layer that is composed of a tailored echo state network. Likewise, several papers have studied the combination of the wavelet transform (to decompose the system dynamics into different timescales [Soltani[02]]) with simple models such as autoregressive moving average (ARMA) models or traditional neural networks [Kaur[15], Peng[14]].

A novel RNN architecture, referred to as gated-feedback RNN (GF-RNN), was presented in [Chung[15]] in order to address this issue of learning multiple adaptive timescales within a single procedure. This GF-RNN relies on deep architectures, consisting of stacking multiple layers on top of each other, in which new connections among recurrent layers are considered so that these layers are fully connected between each other. The global architecture of a gated-feedback recurrent neural network is sketched in Figure 3.6.

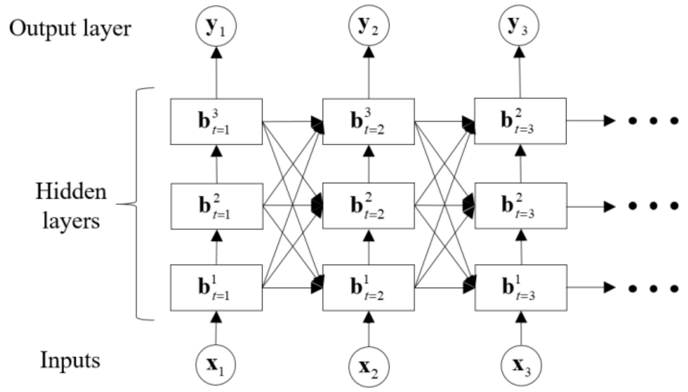


Figure 3.6 – Gated feedback recurrent neural network.

Based on the input vector $[\mathbf{x}_1,\ldots,\mathbf{x}_T]$, the proposed GF-RNN adaptively controls the flow of information between recurrent connections so as to drive each hidden layer to process different timescales of the studied signals. This is realized by using a global gating unit for each pair of layers. The global reset gate between layers l' and l is computed as

$$g_t^{l'\to l} = \sigma\left(\mathbf{W}_{xg}^{l'\to l}\mathbf{b}_t^{l-1} + \mathbf{W}_{sg}^{l'\to l}\mathbf{b}_{t-1}^{*}\right) \tag{3.23}$$

where $\mathbf{b}_{t-1}^{*}$ is the concatenation of states of all hidden layers from the previous time step. Hence, the signal between layer l' at time t-1 and layer l at time t depends on a scalar defined by the current input and all the previous hidden states. The new memory content of the GF-LSTM is then expressed as:

$$\mathbf{c}_t^l = \mathbf{f}_t^l\mathbf{c}_{t-1}^l + \mathbf{c}_t^l \tanh\left(\mathbf{W}_{i\gamma}\mathbf{b}_t^{l-1} + \sum_{l'=1}^{L} g_t^{l'\to l}\mathbf{W}_{h\gamma}^{l'\to l}\mathbf{b}_{t-1}^{l'}\right) \tag{3.24}$$

Therefore, the single GF-LSTM memory cell (i.e. neuron architecture pertaining to hidden layers) can be represented as in Figure 3.7.

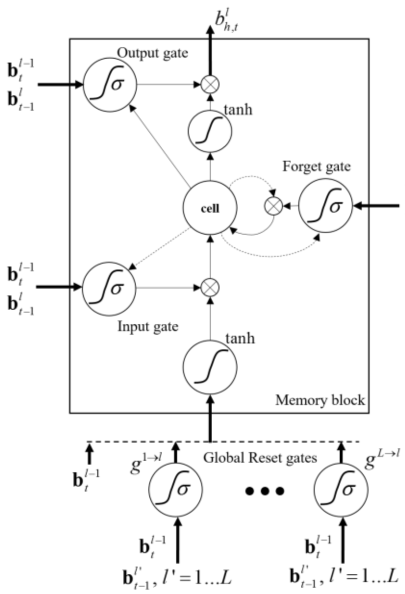


Figure 3.7 – Single-cell GF-LSTM memory block (cell *h* of hidden layer *l* at time *t*).

3.2.2.2 Bidirectional recurrent neural networks

A simple solution for the second issue of RNNs (i.e. suboptimal exploitation of the contextual information) is to add a time-window of future context, and to use the window as additional input features. In this way, the input vector $\mathbf{x}_t$ at time t include also information from the future time steps. Another possibility is presented in Figure 3.8 and consists in introducing a delay in the neural architecture between inputs and outputs corresponding to the same time step so as to give to the network information about future context. However, both these approaches do not usually make full use of backwards dependencies and require the adequate range of future context, which is usually unknown and varies over time (from segment to segment).

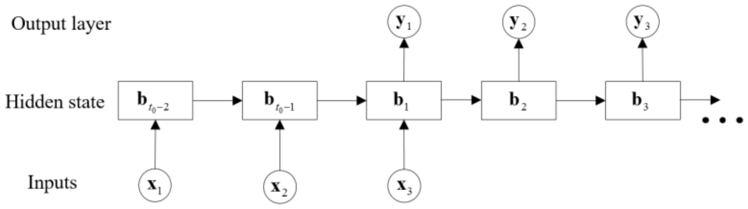


Figure 3.8 – Unidirectional RNN with delay between inputs and targets.

A more efficient solution is provided by the bidirectional topology, which harnesses at each time step t the complete information about the whole temporal horizon (before and after t). As illustrated in Figure 3.9, the principle of such bidirectional RNN is to process the training sequence forwards and backwards by two different recurrent networks, both of which being connected to the same output vector [Schuster[97]]. In this way, for every point of the input sequence, the network has a comprehensive information about past and future points. Such architectures have demonstrated to give state-of-the-art results on speech recognition [Graves[13]].

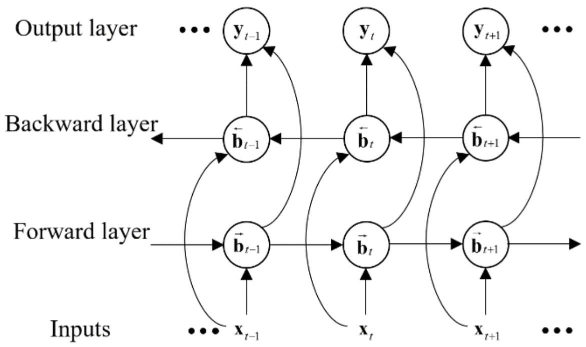


Figure 3.9 – Single layer bidirectional recurrent neural network.

At first sight, this approach is counter-intuitive for prediction tasks as it seems to violate causality. However, for offline multi-step ahead predictions that do not require to generate an output at each time step (i.e. predictions for which outputs are needed simultaneously at the end of the input segment), there is no reason to disregard reliable future information as it is likely to generate improved performance. Such a situation is often encountered in the context of operational planning (Figure 3.10) where decisions have to be simultaneously optimized over the whole scheduling horizon or, more generally, for multi-period optimization (such as the participation in the day-ahead electricity market where the optimal bidding strategy has to be determined at once for the 24 hours of the following day).

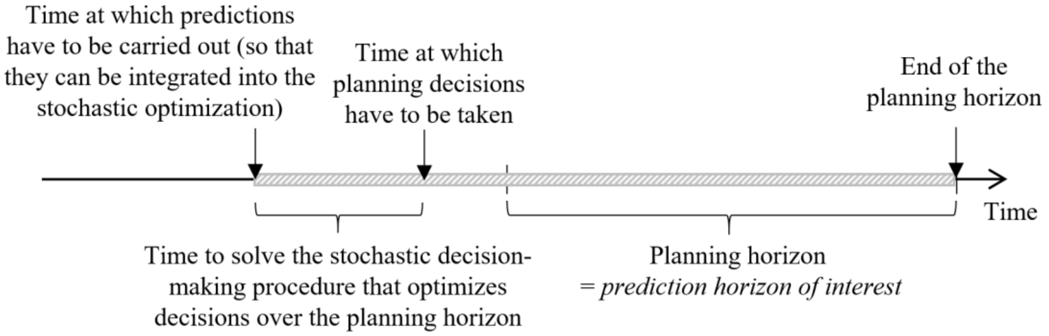


Figure 3.10 – Time line representing the sequential decision-making process for a planning optimization under uncertainty.

Moreover, bidirectional networks are faster to train, and are more robust to model uncertainties and biased inputs. Indeed, in contrast with unidirectional RNNs, they do not rely on a recursive strategy that iteratively fed back previous predictions as inputs for the next time step, which is shown to lead to error accumulation [Bengio[15], Lamb[16]].

Combing bidirectional RNNs with LSTM gives Bidirectional LSTM (BLSTM), which has the benefits of both long-range memory and bidirectional processing [Graves[05,a]]. Furthermore, it is possible to take advantage of deep architectures, which are able to build up progressively higher level representations of data, by piling up RNN layers on top of each other (the output sequence of one layer forming the inputs for the next).

To summarize, we have seen that the concept of neural networks has emerged with feedforward architectures, which were developed as generic tools able to model any nonlinear relationship between a set of explanatory variables (input vector) and the dependent variables of interest (output vector). However, such acyclic topologies do not efficiently capture temporal dependencies, and recurrent neural networks (characterized by cyclical connections) have thus appeared. Recently, the long-short term memory (LSTM) recurrent neural network architecture has become increasingly important (e.g. in the Google Translate algorithm) due to its superior modeling capabilities. However, even this advanced architecture presents limitations, and numerous works have thus been realized to devise architectural improvements, most notably the gated-feedback topology. In parallel, bidirectional recurrent architectures were introduced to better exploit contextual information, and both concepts (LSTM networks and the bidirectional data processing) have been successfully merged.

In this dissertation, the different architectures (feedforward networks, LSTM, gated-feedback LSTM and bidirectional LSTM) are implemented and compared for the task of forecasting electrical quantities (load, renewable generation and electricity prices) for the 24 hours of the following day.

3.2.3 Neural Network training

When training neural networks with gradient descent, several issues have to be addressed for ensuring that the training phase is effective (convergence towards global optimum), fast (limited number of iterations to achieve results convergence), and that the performance of the network is preserved when confronted with new (unseen) data.

3.2.3.1 Gradient descent algorithm

Different methods can be used to follow the error gradient. The simplest (traditional) method is given by (3.8), which consists in taking a small, fixed-length step Δw (whose magnitude is defined by the learning rate α) in the direction that minimizes the gradient of the loss function with respect to the network parameters.

However, due to the nature of gradient descent algorithms that only explore locally the shape of the objective function to decide on the optimal direction, they may get stuck in local minima. This can be alleviated by adding a momentum term [Plaut[86]] in the algorithm search through the weight space as follows:

$$\Delta w_{ij}^{it+1} = m\Delta w_{ij}^{it} - \alpha \frac{\partial L}{\partial w_{ij}^{it+1}} \tag{3.25}$$

where $m \in [0, 1]$ is the momentum parameter. This principle to add inertia (typically $m = 0.9$) in the search procedure has shown to speed up the convergence and help escaping local minima.

The behavior of gradient descent algorithm not only depends on the optimal search direction [Akaike[59]], but also on the choice of a step size. In this regard, a technique that aims at optimally adapting the value of the learning rate at each iteration of the gradient descent algorithm was presented in [Barzilai[88]]. This procedure was implemented in this work, but was unsuccessful for the day-ahead prediction task of electric variables (led to poorer performances than the one using a fixed α value during the whole learning phase).

Then, two different approaches can be considered for updating the weights, which are respectively referred to as **batch learning** and **online learning**. Concerning batch learning, the gradients are computed for each sample of the historical dataset but the weights are only updated once at the end of the training epoch[23]. This procedure contrasts with online learning, or stochastic gradient descent, where the model is updated for each sample in the training set. A compromise between both approaches (split the training set into small batches grouping several samples so that the model is updated at the end of each batch) can also be envisaged.

It should be noted that, due to the principle of the search procedure (small changes in the direction of the optimum) of both batch and online learning, the training procedure necessitates several epochs, i.e. to pass several times throughout the whole historical dataset in order to achieve convergence.

Traditional gradient descent is very efficient in the context of online learning since the stochasticity of the procedure can help escaping from local minima. Indeed, since the shape of the loss function (with respect to model parameters) slightly varies between training samples, the algorithm tends to avoid poor local minima [LeCun[98]]. The stochasticity can be further enhanced by randomizing the order in which the samples of the training set are processed within each epoch (pass through the training set) of the global learning procedure.

Finally, it has been shown that retraining the network (from the last optimal solution) may increase the final performance [Beringer[04], Graves[05,b]], most likely by escaping the local minima in which gradient descent algorithms may potentially get trapped.

[23] An epoch refers to one cycle throughout the entire training dataset.

All simulations carried out in Section 3.5 were carried out using online steepest descent with learning rate of 10^{-4} and a momentum of 0.9. The weights are initialized with a random distribution in the range [−0.1, 0.1].

3.2.3.2 Checking the implementation of the backpropagation with the numerical gradient

When coding backpropagation (and, by extension, backpropagation through time) from scratch, it is strongly recommended to check numerically that the procedure is correctly implemented. This can be achieved by using the **symmetrical finite differences** technique:

$$\frac{\partial L}{\partial w_{h'h}} = \frac{L\left(w_{h'h}+\varepsilon\right)-L\left(w_{h'h}-\varepsilon\right)}{2\varepsilon}+O\left(\varepsilon^{2}\right) \tag{3.26}$$

where ε is a small perturbation (typically 10^{-5}) in the weights. As illustrated in Figure 3.11, this equation (3.26) gives an estimate of the gradient of the loss function.

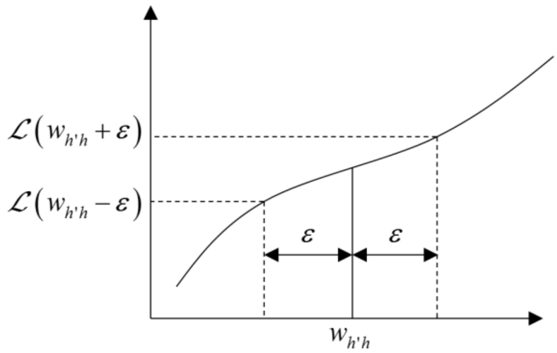


Figure 3.11 – Numerical gradient computation.

One may wonder why we do not use this simple procedure to directly compute the derivatives (instead of the more complex backpropagation procedure that may yield implementation errors). It originates from the fact that calculating the full gradient using (3.26) requires O(P^2) time, whereas backpropagation only requires O(P) time, where P is the number of model parameters (neurons), which makes numerical differentiation impractical for network training. In this way, when checking the code of the backpropagation algorithm with the numerical gradient, it is recommended to use a small-sized network architecture so as to limit the computation time of the check procedure.

3.2.3.3 Regularization

The overarching goal when training a predictive model is to maximize its **generalization** capability to unseen data. To that end, the model should be able to extract the fundamental underpinning properties of the training data while ignoring the irrelevant information included in the noise [Verstraeten[10]].

In other words, the objective is to implement a sufficiently complex model for capturing all hidden characteristics of historical data but not too complex such as to avoid **overfitting**. The overfitting is a modeling error that arises when the model is too closely adapted to a limited set of data points, which substantially reduces its predictive capacity. In this way, overfitting is more likely to occur in two situations. The first one is characterized by a learning process carried out with a small training dataset. In this case, the procedure does not dispose of enough information to discern the noise in the data from the relevant underlying patterns governing the studied process, which results in a model that fits well the training data but that does not efficiently generalize to new observations. The second major cause of overfitting is an important number of model parameters (i.e. high model complexity). In such a situation, there is an increased risk that the weights of some parameters are excessively large in order to accommodate data, resulting in model outputs very sensitive to fluctuations regarding the exact learning conditions. This phenomenon is illustrated in Figure 3.12 where polynomials of different degrees are used to model a particular system based on actual measurements. In the left graph, the first degree model is too simple to adequately represent the underlying system. In the middle plot, the model complexity is well suited, and, in the right graph, the fifth degree polynomial is overly complex and is describing the noise instead of the actual system.

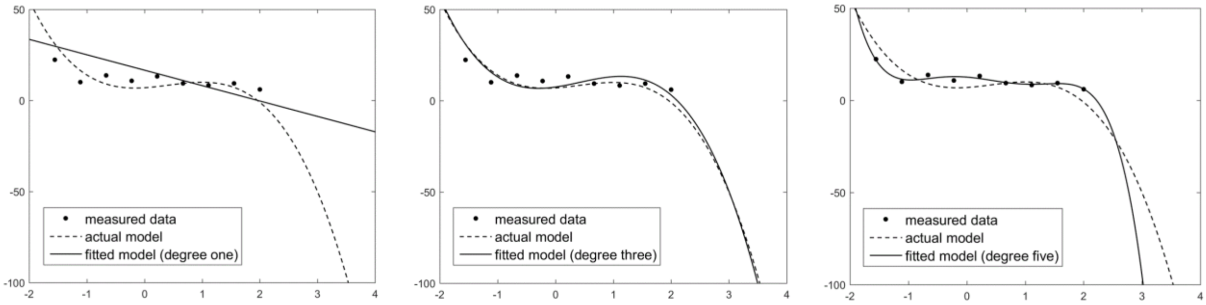


Figure 3.12 – Illustration of the dependence between model complexity and overfitting.

Early stopping

The most common way to avoid overfitting is to divide the historical dataset into two separate sets: a training set and a validation set. The principle is to train the model using only the training data, and to evaluate the performance of the model at regular intervals on the independent validation set (no gradient calculations or weight updates are performed during this test). The training is stopped when the error on the validation set is minimized. Indeed, as long as the network learns the structure of the data, the performance on the validation set will increase. Then, when the network stops deciphering the actual relationship between inputs and outputs and begins to learn the noise within the training sample, the error will stop decreasing (and will even start rising) on the validation set, while continuing to drop on the training set (Figure 3.13).

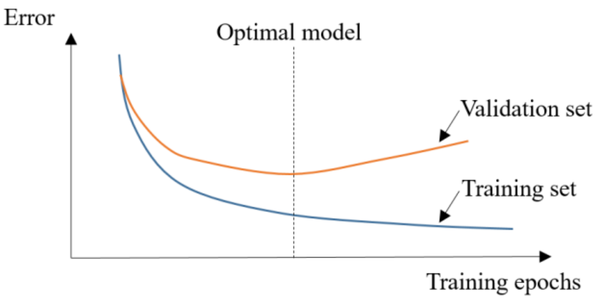


Figure 3.13 – Overfitting on training data.

The main drawback of this method is that part of the data are lost for the validation set, which can be problematical if the historical dataset is small. Moreover, there is no methodology to know the optimal division of data (or, in other words, how big the validation set should be to be sufficiently representative without sacrificing too much valuable information).

In parallel to early stopping, other **regularization** techniques are introduced to prevent from overfitting (and ensure that the training set performance carries over to the test set). The principle is introduce additional information so as to reach a trade-off between model complexity (ability to capture properties of the data) and noise robustness (ability to avoid modeling irrelevant dependencies). The main regularizers are here briefly introduced.

L1 regularization

The first technique is the LASSO (Least Absolute Shrinkage and Selection Operator) regression, which adds a penalty term, equal to the sum of the absolute value of the model coefficients (weights), into the loss function so as to avoid high values of these parameters. The loss function to be minimized is therefore given by:

$$L + \lambda \sum_{p=1}^{P} \left| w_p \right| \tag{3.27}$$

where P is the number of parameters (weights) of the model, and λ is a parameter that provides a trade-off between the original loss function and the magnitude of coefficients. In this way, values of λ that are too low will not be able to solve overfitting issues, whereas large values will result in coefficients with values close to zero (leading to model underfitting).

L2 regularization

A variant to L1 regularization, which is based on the same principle is given by ridge regression, which adds a penalty term equal to the sum of the squares of the model coefficients in the objective function:

$$L + \lambda \sum_{p=1}^{P} \left(w_p \right)^2 \tag{3.28}$$

Elastic net

Elastic net is a regularized regression method that combines the L1 and L2 penalties.

$$L + \lambda_1 \sum_{p=1}^{P} \left| w_p \right| + \lambda_2 \sum_{p=1}^{P} \left(w_p \right)^2 \tag{3.29}$$

Input noise

An alternative strategy is to add zero-mean (with a fixed variance) Gaussian noise to the inputs of the network during the training stage [Koistinen[91]]. This allows to generate more training examples by deforming the existing ones, which artificially enhances the number of data and improve the generalization.

However, such input perturbations should reflect the actual variations that can occur in the data. In this way, contrary to meteorological data, categorical information (such as day of the week) should not be altered.

Weight noise

A similar technique consists in adding zero-mean (with a fixed variance) Gaussian noise to the network weights [Murray[94]]. Since the noise acts on the internal state of the network (rather than on its inputs), it can be employed regardless of the type of data. However, the noise variance has also to be carefully selected, and the method can overall hamper the convergence of the learning procedure.

Dropout

More recently, a new technique known as dropout was presented and successfully applied to several tasks [Srivastava[14]]. At each stage of the training, each individual unit (neurons) can be either dropped out of the network (with probability 1-*u*) or kept with probability *u*. It is shown that the reduced network forces to learn more salient features. Practically, dropout tends to double the number of iterations to reach the convergence of the gradient descent algorithm, but the training time associated with each epoch is decreased. This technique has proved to give good results, especially when it is coupled with batch normalization (Section 3.2.3.5).

In this work, early stopping is combined with weight noise during the training phase for enhancing the generalization capabilities of the neural network-based prediction models.

3.2.3.4 Hyperparameters optimization

Beyond the quality of the training algorithm, the final performance of the neural network depends on two important conditions. The first one is the selection of the appropriate set of explanatory variables (inputs selection). This task is essential since any missing information will inevitably deteriorate the model ability to provide accurate outcomes whereas irrelevant input data will, for their part, lead to additional unnecessary noise (that may disorient the learning algorithm). Secondly, even though a well-thought training strategy allows to reduce under- and over-fitting issues, it is important to properly define the complexity (number of parameters to optimize) of the neural network. The latter can be tuned along two dimensions (also referred to as hyperparameters): the number of hidden layers within the network architecture and the number of neurons within each hidden layer. Finding the optimal architecture (in terms of both inputs selection and model complexity) is task-dependent, and is achieved in this thesis using to the two-nested loops approach [Toubeau[17]] presented in Figure 3.14. It should be noted that other hyperparameters (such as the learning rate of the gradient descent learning procedure, or other parameters associated with regularization techniques used during training) have to be optimized together with the complexity of the network architecture for the optimal model selection.

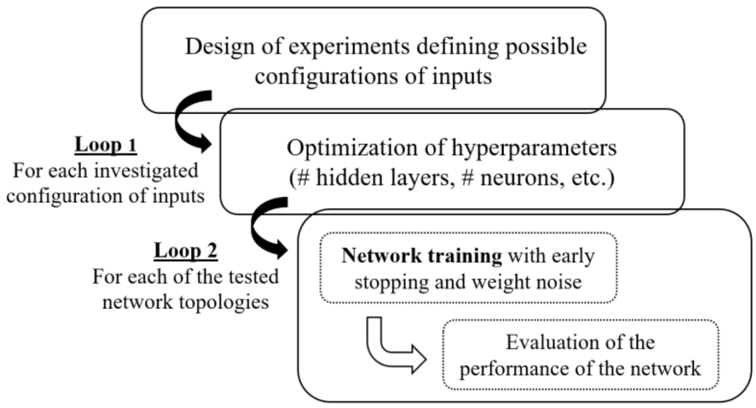


Figure 3.14 – Two-nested loops procedure designed for the optimal model selection.

Finally, at the end of the two-nested loop procedure, the different models (differentiated regarding both their inputs and hyperparameters) can be ranked with respect to their statistical score on the validation set (unseen data), and the best model is then used for practical application.

However, in order to compare the different topologies, it is necessary to define an appropriate error metric with respect to the considered task (e.g. sum of squared errors for deterministic predictions). The accuracy measure obtained at the end of the learning process (with the validation set) is slightly biased (smaller than the true error rate) since the training is stopped at the optimal time with respect to the validation set. The performance depends thus on the decomposition of the data. To solve this issue, another part of the dataset can be used as a test set. Hence, the training is carried out on the training set with regular evaluations of the model performance on the validation set. Once the learning is achieved, the final evaluation is done on the test set.

However, such a procedure further reduces the number of samples to train the model, and the final accuracy measure still depends on the particular random division of historical data. A solution to this problem is given by **cross-validation**, in which the historical data are divided into κ distinct subsets (the process is then called κ-fold cross-validation). The methodology can be summarized as follows:

For each of the κ separate folds

    This fold is used as a validation set to stop the training at the optimal time

    The model is then trained using the κ-1 remaining folds as training information

The final performance measure of the procedure is the average of the κ performance values computed in the loop

---

3.2.3.5 Input representation

The choice of a suitable representation of input data is an important pre-processing task when dealing with neural networks. This consists in adapting the input values to the operating range of neural processing units for avoiding to systematically end up in the saturation zone of neurons, which would lead to poor generalization (Figure 3.15). This procedure does not degrade the information of the explanatory variables, and improves the performance of neural networks by putting the input values in a range more suitable for the standard activation functions [LeCun[98]].

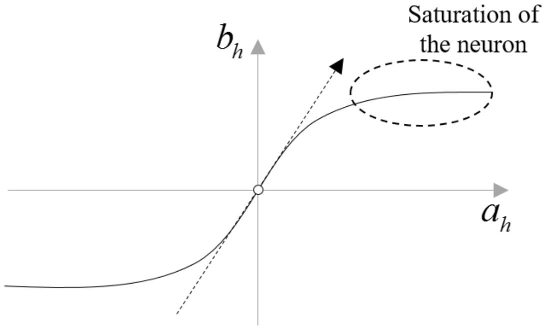


Figure 3.15 – Sigmoidal activation function of processing unit.

Typically, for hyperbolic tangent units, the components of the input vectors are standardized so as to have a zero mean and standard deviation equal to 1 over the whole training set. It should be emphasized that the validation set (as well as the test set) has to be standardized with the same parameters than those used in the training set.

Instead of normalizing uniquely the network inputs, a more complex policy, known as **batch normalization**, has been lately introduced [Ioffe[15]], where the inputs to all layers within the neural network are also normalized, by applying a transformation that maintains the mean of the activated state close to 0 and its standard deviation around 1. Such networks tend to train faster, reduce the sensitivity to the network initialization (starting weights), and allow to better extract the full potential of deeper networks (with regularization properties).

To summarize, this thesis aims to compare (in Section 3.5) four different neural network architectures (multilayer perceptron, LSTM, gated-feedback LSTM and bidirectional LSTM), trained with online gradient descent with momentum (with a combination of early stopping and weight noise to avoid the model overfitting), with the prior standardization of both input and output variables.

3.3 From deterministic to probabilistic predictions

Generally, neural networks can be used (with similar success) for three main fields of application:

- **Classification** when the output function is discrete (e.g. predict the failure status of industrial equipment in the context of predictive maintenance).
- **Regression** when the output function is continuous (e.g. deterministic forecasting of electric quantities).
- **Probability estimation** when the output function is a probability distribution (e.g. probabilistic forecasting of electric quantities).

In this thesis, only the last two categories will be investigated.

3.3.1 Point forecasting

In the context of deterministic predictions, the objective of the neural network training (for both feedforward and recurrent architectures) is to use the historical datasets (of explanatory and dependent variables) within a supervised learning strategy that adjusts the model parameters in order to maximize the predictive capability of the tool. Practically, this consists in finding the optimal weights between neurons so as to determine the conditional mean $E(y_{k,t_0+t}|y_{k,\to t_0})$ of outputs $y_{k,t}$ for each time t of the prediction horizon (which starts at t_0) for each variable of interest k.

The typical loss function for this task is the sum of the squared errors (SSE), which aims at minimizing the squared deviations between predictions and actual observations. This function, which is easily differentiable, is presented in (3.30).

$$L = \sum_{t=1}^{n} \left(y_t - d_t \right)^2 \tag{3.30}$$

where n is the number of time steps T of the sequence of interest (e.g. for hourly day-ahead prediction T = 24) , y_t the output of the prediction model (MLP, LSTM, BLSTM, etc.) and d_t the actual measured value.

However, for multi-step predictions, the stability of the predicted signal may be problematic (i.e. large deviations at the end of the prediction horizon) and is not fully captured by the traditional SSE metric since the last prediction steps are averaged with the first ones [Wyffels[13]]. To overcome such an issue, we can use a weighted error metric as loss function (during gradient descent training) that progressively weights up errors along the prediction horizon so as to ensure stability of the prediction tool. The purpose is to encourage the training to achieve a good accuracy throughout the prediction horizon rather than excellent precision for first time steps with significant deterioration over time.

$$L = \sum_{t=1}^{n} \gamma_t \left(y_t - d_t \right)^2 \tag{3.31}$$

where γ_t is a monotonically increasing function.

Here, the neural networks were thus firstly trained with a weighted Sum of Squared Error (wSSE), and the optimal solution is then used as a starting point for another learning phase (retraining) with classic SSE.

3.3.2 Probabilistic forecasting

Many decision making procedures such as the optimal bidding strategy in electricity markets, require richer information than point forecasts since deterministic optimization show very poor robustness regarding forecast errors. This additional information can be efficiently provided by a probabilistic forecast that yields the full conditional distribution $p(y_{k,t_0+t}|y_{k,\rightarrow t_0})$.

3.2.2.1 Parametric model of prediction errors

In order to obtain this predictive probability distribution of outputs, the first investigated procedure is to define a statistical (parametric) model of forecast errors y_t - d_t (e.g. Gaussian model) and to use the neural network for predicting the parameters of the specified distribution (e.g. mean and variance of the Gaussian model) using the **maximum likelihood estimation**. The neural network is thus trained so that its outputs (parameters θ of the specified distribution) are maximizing the likelihood function $L(\theta)$. In other words, the procedure aims at finding the parameters of a specified distribution that maximizes the "probability" of observing the available historical dataset.

However, maximizing the likelihood requires to compute the partial derivatives of the function $L(\theta)$ with respect to its parameters θ, and it is therefore more convenient to use the log-likelihood (which can be indifferently applied since the logarithm function is monotonically increasing). Hence, the maximum likelihood estimation is equivalent to minimizing the negative log-likelihood, and this loss function E_L can be expressed as follows:

$$E_L = -\sum_{t=t_0}^{T} \ln L\left(y_{k,t} \middle| \theta\left(\mathbf{b}_t^L \right) \right) \tag{3.32}$$

where $\mathbf{b}_t^L$ is the internal state of the last hidden layer of the neural network.

Here, the Gaussian likelihood L_G is employed, which is parametrized using the mean and standard deviation of past observations $\theta = (\mu, \sigma)$:

$$L_G = \left(2\pi\sigma^2\right)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{3.33}$$

The log-likelihood l_G is then:

$$\boldsymbol{l}_G = \frac{-1}{2}\left(\ln(2\pi) + \ln\left(\sigma^2\right) + \frac{1}{\sigma^2}(x-\mu)^2\right) \tag{3.34}$$

where x is the actual measurement of the dependent variable. The mean μ of the distribution is given by an affine function of the network output (3.35), whereas the standard deviation σ is determined by applying sequentially a softplus activation after the affine transformation, in order to ensure that its value remains strictly positive (3.36):

$$\mu\left(\mathbf{b}_t^L\right) = \mathbf{w}_\mu \mathbf{b}_t^L \tag{3.35}$$

$$\sigma\left(\mathbf{b}_t^L\right) = \ln\left(1 + \exp\left(\mathbf{w}_\sigma \mathbf{b}_t^L\right)\right) \tag{3.36}$$

where $\mathbf{w}_\mu$ and $\mathbf{w}_\sigma$ represent the output weight vectors associated respectively with the mean and standard deviation. The gradients can be computed as follows:

$$\delta_\mu^t = \frac{\partial E_{L_G}}{\partial \mu} = \frac{1}{\sigma^2}(x-\mu), \quad \frac{\partial \mu}{\partial \mathbf{b}_t^L} = \mathbf{w}_\mu \tag{3.37}$$

$$\delta_\sigma^t = \frac{\partial E_{L_G}}{\partial \sigma} = \frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}, \quad \frac{\partial \sigma}{\partial \mathbf{b}_t^L} = \frac{\mathbf{w}_\sigma \exp\left(\mathbf{w}_\sigma \mathbf{b}_t^L\right)}{1 + \exp\left(\mathbf{w}_\sigma \mathbf{b}_t^L\right)} \tag{3.38}$$

It should nonetheless be emphasized that other likelihood models can be employed, provided that the function derivatives with respect to their parameters θ can be obtained.

3.2.2.2 Non-parametric model of prediction errors

In real-life applications, it may be difficult to know the exact theoretical distribution of the uncertainty at hand. In this context, methods that do not rely on a pre-defined distributional assumption are likely to be more robust compared to other parametric methods. A solution consists therefore in using quantile regression [Koenker[78]], for which the objective is to directly predict the specified quantiles $q \in Q$ of the target distribution:

$$q = P\left(y_{t_0+t} \leq y_{t_0+t}^{(q)} \,\middle|\, y_{\to t_0}\right) \tag{3.39}$$

In this framework, models are trained to minimize the quantile loss (or pinball loss) since it has been proved in [Takeuchi[06]] that minimizing this pinball loss E_q yields the optimal quantiles. The total loss is therefore the result of the sum over all specified quantiles of interest:

$$E_q = \sum_{q \in Q} q \max\left(0, d - y^{(q)}\right) + (1-q)\max\left(0, y^{(q)} - d\right) \tag{3.40}$$

where the quantiles $y^{(q)}$ are given by an affine function of the network outputs. It is interesting to notice that, when q = 0.5, we get an estimate of the conditional median of the output distribution.

Moreover, similarly to the parametric model previously described, a great asset of the methodology is that the loss function is differentiable, so that the neural network can be trained using gradient descent. This learning procedure (contrary to metaheuristics such as genetic algorithm or particle swarm optimization) allows the network to be systematically retrained each day using only the new information that has been revealed, so that the computational burden of this retraining task is very limited.

3.4 From multi-step ahead probabilistic predictions to time-dependent scenarios

Once the predictive distributions (either Gaussian or under the form of empirical quantiles) at each time step $t \in T$ for each output variable $k \in K$ are obtained, the objective is to obtain samples $\mathbf{y}_{i,t_0:T}^{s}$ from the D-dimensional distribution ($D = \#T\#K$), where # stands for the cardinality of the associated set. The generated scenarios therefore contains the global dependence structure of variables

$$\mathbf{y}_{k,t_0:T}^{s} \sim H\left(\mathbf{y}_{k,t_0:T} \middle| \mathbf{y}_{k,\to t_0-1}, \mathbf{x}_{k,t_0:T}\right) \tag{3.41}$$

where t_0 stands for the start of the prediction horizon of interest (data before t_0 are therefore assumed to be known for the prediction phase).

However, the task of generating random vectors from a high dimensional distribution is really complex, even when the marginal distributions of each dimension are known [Law00]. The only exception is when the variables are independent since the multivariate distribution $F_{\mathbf{Y}}$ can be simply decomposed as the product of its marginal (= univariate) distributions:

$$F_{\mathbf{Y}}\left(\mathbf{y}_D\right) = F_{Y_1}\left(y_1\right) F_{Y_2}\left(y_2\right) \ldots F_{Y_D}\left(y_D\right) \tag{3.42}$$

where $F_{Y_k}(y)$ is the distribution function of the variable Y_k.

Otherwise, it is necessary to remember that a D-dimensional distribution function can be decomposed into:

$$F_{\mathbf{Y}}\left(\mathbf{y}_D\right) = F_{Y_1}\left(y_1\right) F_{Y_2|Y_1}\left(y_2 \middle| y_1\right) \ldots F_{Y_D|\mathbf{Y}_{D-1}}\left(y_D \middle| y_1, \ldots, y_{D-1}\right) \tag{3.43}$$

The conditional probability function of Y_d given the values $\mathbf{Y}_{d-1}$ can be expressed as:

$$f_{Y_d|\mathbf{Y}_{d-1}}\left(y_d \middle| y_1, \ldots, y_{d-1}\right) = \frac{f_{\mathbf{Y}_d}\left(y_1, \ldots, y_d\right)}{f_{\mathbf{Y}_{d-1}}\left(y_1, \ldots, y_{d-1}\right)} \tag{3.44}$$

The joint conditional cumulative distribution function (CDF) is then computed according to:

$$F_{Y_d|\mathbf{Y}_{d-1}}\left(y_d \middle| y_1, \ldots, y_{d-1}\right) = \int_{y=0}^{y_d} \frac{f_{\mathbf{Y}_d}\left(y_1, \ldots, y_{d-1}, y\right)}{f_{\mathbf{Y}_{D-1}}\left(y_1, \ldots, y_{d-1}\right)} dy \tag{3.45}$$

From equation (3.45), it results that the computation of the *D*-1 conditional distribution functions $F_{Y_d|\mathbf{Y}_{d-1}}$ of a *D*-dimensional distribution theoretically requires knowing the analytical function of the marginal distributions $f_{\mathbf{Y}_d}(y_1, \dots, y_d)$, which are unknown.

This problem is bypassed using an original solution [Toubeau[18]] that relies on a copula model, which represents an attractive alternative to compute multivariate distributions. As more thoroughly explained in Annex B, such models integrate the whole dependence structure of variables (independently from the constitutive univariate marginal distributions [Sklar[59]]).

The novel procedure to generate dependent samples from the multivariate multi-step ahead probabilistic forecasts is represented in Figure 3.16. It is decomposed into two parts. Firstly, the copula model is trained using historical observations of the dependent variables (Annex B). To that end, the univariate distributions (each one corresponding to a particular variable *k* at one specific time step *t*) are empirically constructed (phase I-A). The probability integral transformation is then used to convert these variables into uniform variables (phase I-B), for which the multivariate distribution (copula model) can be easily computed (phase I-C). Secondly, once the copula model is obtained, it can be used to generate uniformly distributed numbers $\mathbf{u} = (u_1, \dots, u_D) \in [0, 1]^D$ with the dependence structure of the original data. Using the marginal predictive distributions obtained with the probabilistic forecasting tool, these uniform numbers can then be converted into the original dimensions thanks to the inverse transform sampling so as to obtain the scenarios $\mathbf{y}^{s}_{i,t_0:T}$ (phase II). The sampled scenarios thus encompass both time and inter-variable dependencies.

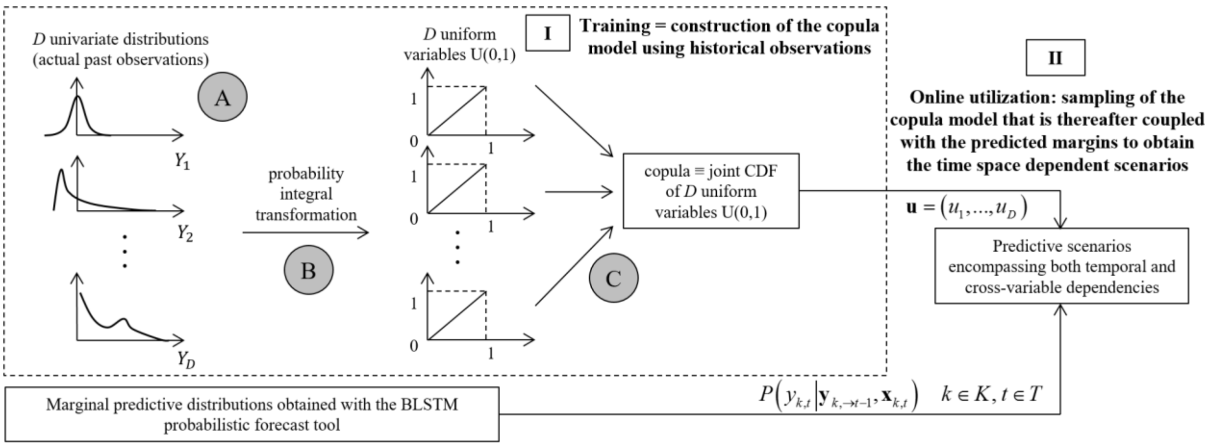


Figure 3.16 – Generation of predictive scenarios from multivariate distributions.

3.5 Results

In the current context of liberalized electricity markets, energy aggregators, need to define each day (typically at 12h00) their optimal bidding strategy for the 24 hours of the following day, and must therefore have accurate predictions of the stochastic variables influencing the decision procedure over this scheduling horizon of 24 hours. As represented in Figure 3.17, the prediction horizon of interest for the day-ahead stochastic decision-making problem spans thus from 12 to 36 hours in the future.

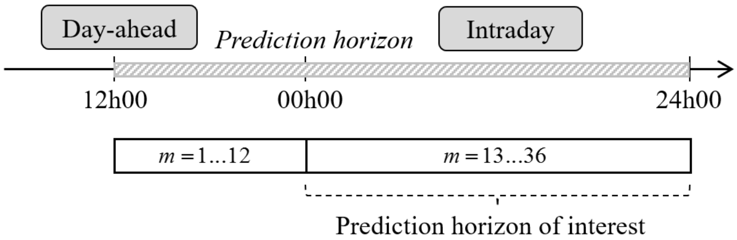


Figure 3.17 – Representation of the prediction horizon.

The predictions focus on the aggregated load and renewable generation (both onshore wind and photovoltaic generation) in Belgium in order to confront such predictions with the day-ahead forecasts published by the system operator (at 12h in day-ahead). Indeed, the latter publishes each day its forecasts for the purpose of promoting a transparent and more competitive market. The electricity prices related to the day-ahead market (DA prices) are also forecasted as they constitute highly relevant information for the day-ahead scheduling of market participants. Practically, the following neural network architectures are compared (thereafter referred by their abbreviations in brackets):

- Multilayer perceptron (MLP), i.e. traditional static feedforward network, in which outputs at every time steps are simultaneously predicted so as to avoid accumulation of errors.
- Unidirectional LSTM (LSTM)
- Gated-feedback LSTM (GF-LSTM), an improved variant of the LSTM neural architecture, developed to optimally process signals with different timescales.
- Bidirectional LSTM (BLSTM), the reference architecture to capture time dependencies.

In order to compare the different variants on a fair basis, the same amount of effort was given in the determination of the optimal topology (same number of investigated configurations in the two-nested loops procedure). Moreover, all architectures are implemented and tested using the same simulation environment (Matlab).

The prediction models were trained using hourly historical data from 2012 until 2017. The performance of the three compared neural networks (final architectures at the end of the optimal model selection) is evaluated on the month of January 2017 (test set composed of data that are not included within the learning phase). In order to increase the network robustness regarding unseen data, two (complementary) regularization techniques are jointly used during the learning phase, namely early stopping and the addition of weight noise so as to ensure that the network ignores the irrelevant information contained in the data.

It should be noted that each variable is predicted independently (with a tailored neural network), since it has surprisingly been observed that the joint prediction of several variables (single prediction tool with several outputs to better capture interdependencies) does not improve the prediction accuracy.

The selection of adequate inputs is driven by both an intuitive determination of the influencing factors and a numerical comparison of the selected input configurations.

All studied variables are characterized by both daily and yearly cycles (i.e. temporal profiles hold essentially the same shape from one day to the next and from year to year). Additionally, load and electricity prices are also strongly related to human activity, which results in a weekly periodicity. For example, the frequency content of the load signal can be visualized though its spectral density estimation (SDE) in Figure 3.18. Such a method allows describing how the power of a signal is distributed over the frequency domain and, since the period is the inverse of frequency, the frequencies that carry most of the energy correspond to the prevalent periods. Here, a non-parametric approach, referred to as Welch's method, is used for adequately representing the periodogram of the load.

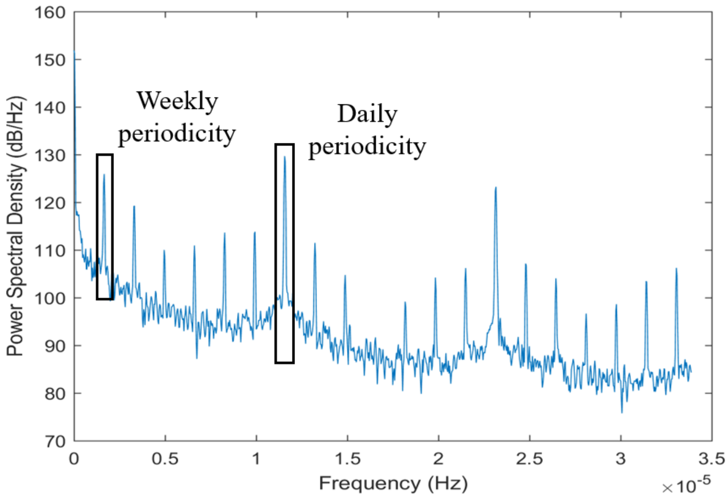


Figure 3.18 – Welch's periodogram for aggregated Belgian load.

Two main peaks can be observed, one at the frequency of 1.66 μHz, which corresponds to a weekly periodicity and the other at 11.56 μHz, which highlights the strong daily cycle of the load. In this way, the load temporal profile holds essentially the same shape from one day to the next one and from week to week. It should be noted that remaining peaks of lesser importance are only integer multiples of the aforementioned harmonics.

For temporal information, the input selection does not only concerns the determination of adequate explanatory variables but also the way the information has to be included into the network. Indeed, the relative importance of these time data is not easily quantified by a numerical value. For instance, the second hour of the day is not 2 times more important than the first one. In this context, a binary representation may provide a more natural way of expressing such data, but at the expense of an increased dimensionality of the network input space. Here, different inputs combinations were therefore tested. Hence, for describing the hourly variation within the day, the different options were:

- Incremental indexing: a single input in the form of a continuous value within the range [0.1, 2.4].
- Incremental binary representation: 5 inputs representing a binary Gray coding (from '00001' to '10100'). In contrast with the traditional binary representation, the Gray coding is a binary numerical system in which two successive values differ in only one bit, which allows smoother transitions between time steps.
- Mutually exclusive binary representation: 24 binary inputs, one for each hour of the day. With such an input representation, when one input is equal to 1, all others are set to 0.

For all temporal information (hours of the day, day of the week and month of the year), the best performance was obtained with the mutually exclusive binary representation, provided that the network complexity was sufficiently important (with a sufficient number of hidden units with each layer). Furthermore, an additional day index is introduced to efficiently represent the occasional events such as public holidays.

Then, weather conditions have also a significant influence on the studied variables. Here, the day-ahead predicted features (temperature, wind speed, cloud cover and solar radiation) provided by numerical weather predictions (NWP) at a single location in Belgium are thus integrated as network inputs.

Finally, historical information is also provided so that the neural network can exploit the recent trend of the variable.

To summarize, the inputs (explanatory variables) of the neural networks are:
- Recent historical observations (last measured values before the prediction);
- Temporal information (month of the year, day of the week and hour of the day) expressed with the mutually exclusive binary representation;
- Discrimination of public holidays with a binary variable;
- Forecasted weather data (such as temperature, cloud cover, etc.) coming from meteorological models (given by the Royal Meteorological Institute) at one location of the country.

3.5.1 Performances of point forecasts

The statistical quality of point forecasts from the different neural network topologies, which focuses on the degree of correspondence between the predictions and the actual observations, is estimated. For these deterministic forecasts, the root mean square error (RMSE) is used as error metric:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - d_t\right)^2} \tag{3.46}$$

where n is the number of sampled data (number of time-steps T predicted each day multiplied by the number of simulated days), y_t the output of the prediction model and d_t the actual measured value. The results are presented in Table 3.1.

Table 3.1
Comparison of tested architectures in terms of RMSE.

Network	Wind	PV	Load	DA prices
BLSTM	101 MW	53 MW	236 MW	17 €
GF-LSTM	105 MW	65 MW	250 MW	20 €
LSTM	108 MW	72 MW	242 MW	20 €
MLP	113 MW	78 MW	282 MW	22 €
System Operator	109 MW	67 MW	391 MW	NA

The results demonstrate that recurrent neural networks show higher accuracy than both the multilayer perceptron and the tool used by the system operator. It should nonetheless be noted that we have little knowledge regarding the information used by the system operator. On the one hand, we have assumed that the information at 11h is available when conducting the predictions at 12h, which may not be the case in practice. But on the other hand, the meteorological information at our disposal is very limited (only one location at the center of the country) compared to what is actually available (to companies that have the financial resources to obtain the data).

The best performance is given by the bidirectional architecture, which emphasizes the importance of accurately accounting for time dependencies in the context of multi-step ahead forecasting. Then, the GF-LSTM architecture is more prone to overfitting than simpler models. In this way, the GF-LSTM outperforms the traditional LSTM network for predicting the wind and the PV generation but yields lower performance for the load. This can be explained by the strong deterministic behavior of the load curve that does not necessitate a complex model to catch important features.

Finally, it is also interesting to emphasize that the optimal results are obtained with deep architectures (with several hidden layers). In this way, the best topology for predicting the wind generation, the total load and the day-ahead prices is obtained with a 2-layers BLSTM, whereas the best prediction model for the aggregated photovoltaic generation is a BLSTM network with 3 hidden layers. Furthermore, the optimal size (capacity) of the neural networks is relatively low (10 to 20 neurons within each hidden layers, which can be explained by the necessity to avoid overfitting with the small number of historical data available. In this way, over the years, the accuracy of the proposed self-learning approach is expected to grow thanks to the amount of data that will allow to progressively increase the network capacity (complexity).

3.5.2 Performances of probabilistic forecasts

The second objective is to compare the Gaussian assumption of prediction errors with a non-parametric approach. The BLSTM architecture is used as a reference to evaluate these (parametric and empirical) methods. The statistical accuracy of the conditional quantiles provided by both methods is computed using the total quantile loss, with q = 1, 5, 10, 25, 50, 75, 90, 95 and 99 %, averaged over the 24 hourly time steps of the prediction horizon.

$$E_q = \frac{1}{n}\sum_{t=1}^{n}\sum_{q\in Q} q\max\left(0, d_t - y_t^{(q)}\right) + (1-q)\max\left(0, y_t^{(q)} - d_t\right) \tag{3.47}$$

This error metric (3.47) is evaluated on the same test set (January of 2017), and the outcomes are presented in Table 3.2. It should be noted that, since the model associated with the quantile loss is trained on the same error metric than the one used for evaluating its post-hoc accuracy, the results may be slightly biased. However, this measure is the reference in both statistical and machine learning communities [Steinwart[11]], and is therefore used nonetheless.

Table 3.2
Comparison of parametric and non-parametric quantiles.

Topology	Wind	PV	Load	DA prices
BLSTM + Gaussian	171 MW	42 MW	422 MW	34 €
BLSTM + quantile	147 MW	41 MW	389 MW	28 €

Overall, the non-parametric model slightly outperforms the outcomes obtained with the Gaussian error assumption (i.e. the quantiles encapsulate more accurately the actual observations, and are characterized by more tight intervals). These results tend thus to support that the empirical model should be privileged, but, for the studied variables, assuming a Gaussian distribution of prediction errors does not lead to significant modeling errors (especially for PV generation).

For illustrating the quality of the results obtained using the BLSTM network with the (non-parametric) quantile loss function, the probabilistic forecasts associated with the four studied outputs are shown in Figure 3.19. Specifically, the concatenation of day-ahead predictions (at 12h in day-ahead for the 24 hours of the next day) carried out during 7 consecutive days (from Monday to Sunday) is presented.

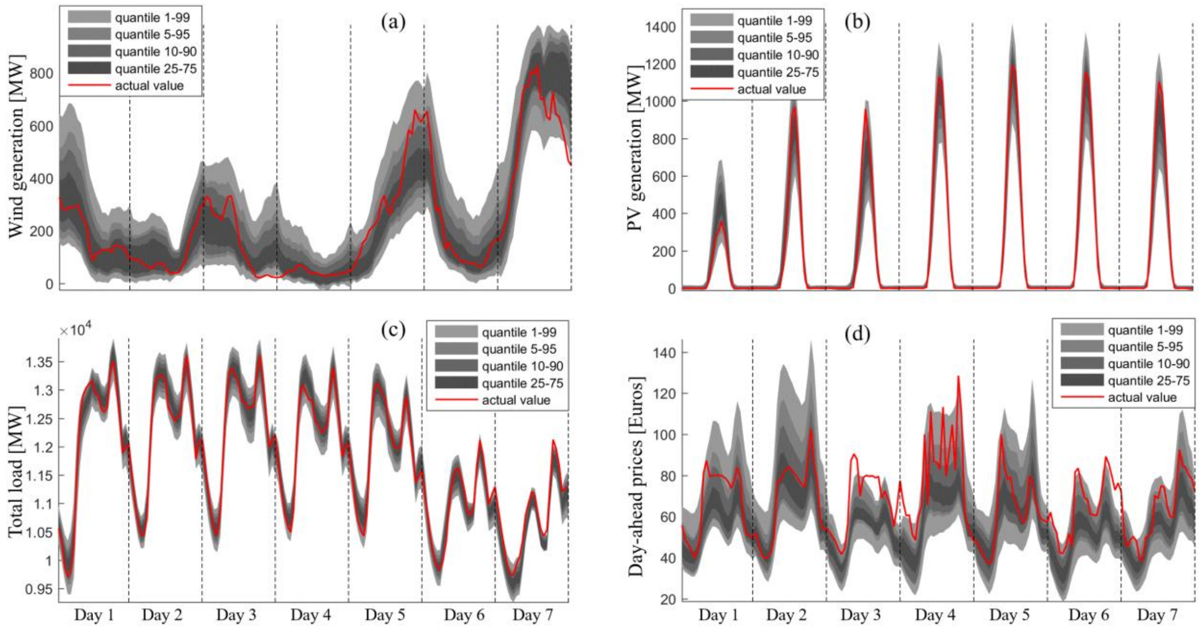


Figure 3.19 – Probabilistic forecasts (with the quantile loss as loss function) performed during 7 consecutive days for wind (a), PV generation (b), total load (c), and day-ahead electricity prices (d).

One can see that the predicted intervals properly encompass the actual realizations of uncertainties (the volatility of the studied variables is well captured). However, we observe that the quantiles are more tightened for the aggregated load, which indicates that the amount of uncertainty associated with this variable is much lower than, for instance, wind generation. Moreover, it should be mentioned that the simulated month was characterized by a high demand (and very low renewable generation) throughout Western Europe, which has considerably increased the price uncertainty (volatility) during this period.

3.5.3 Quality of stochastic scenarios

Once these probabilistic forecasts are obtained, the distributions are sampled to obtain the time trajectories that can thereafter be used in a stochastic optimization tool (Chapter 4). The quality of scenarios is compared regarding both the tool used for probabilistic forecasts (i.e. the MLP and BLSTM tools) and the sampling policy (i.e. independent [Quan[15]] and copula-based sampling methods). The results are illustrated for two representative days (days 3 and 7 of the week represented in Figure 3.19) of wind generation.

The wind power scenarios generated from the BLSTM tool for the third and seventh days are respectively represented in Figure 3.20 and Figure 3.21. Likewise, the scenarios issuing from the feedforward multilayer perceptron for the same days are shown in Figure 3.22 and Figure 3.23.

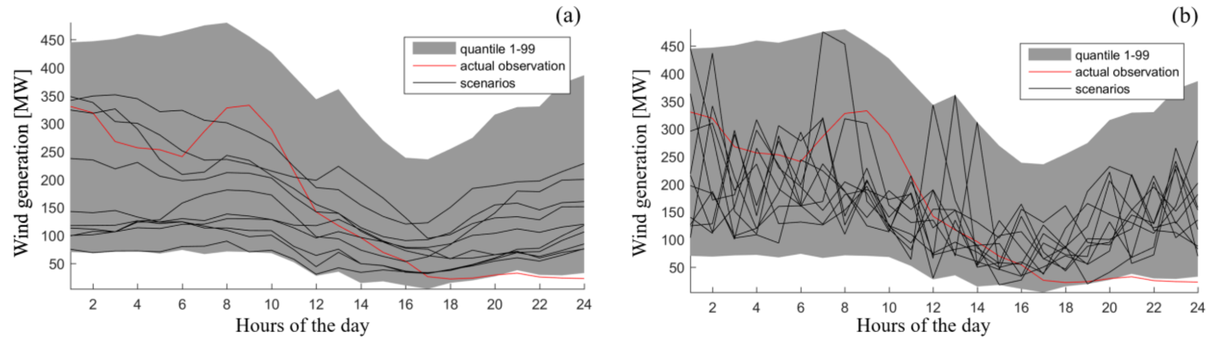


Figure 3.20 – Scenarios of wind generation for the third day of the considered week obtained using the copula-based sampling (a), and independent sampling (b) methods from the BLSTM prediction tool.

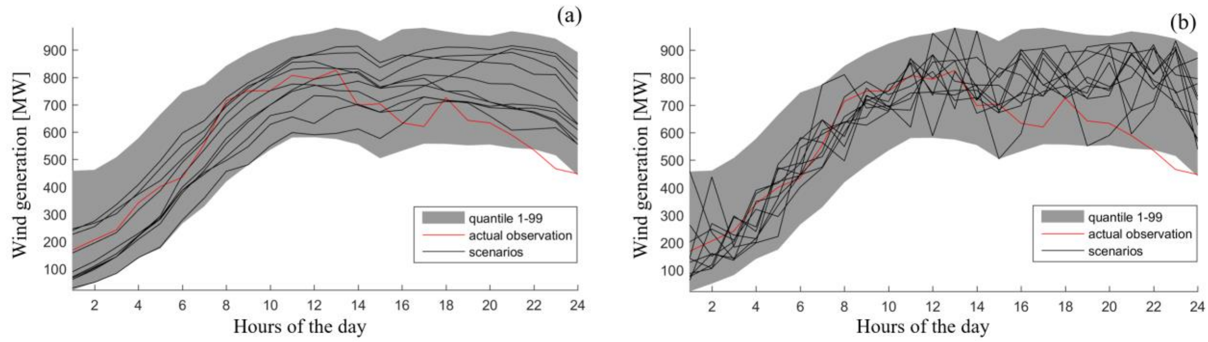


Figure 3.21 – Scenarios of wind generation for the seventh day of the considered week obtained using the copula-based sampling (a), and independent sampling (b) methods from the BLSTM prediction tool.

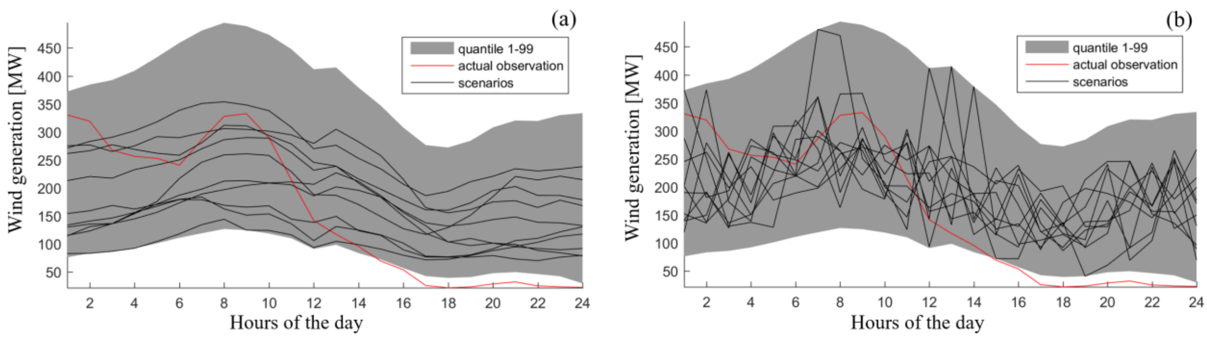


Figure 3.22 – Scenarios of wind generation for the third day of the considered week obtained using the copula-based sampling (a), and independent sampling (b) methods from the MLP prediction tool.

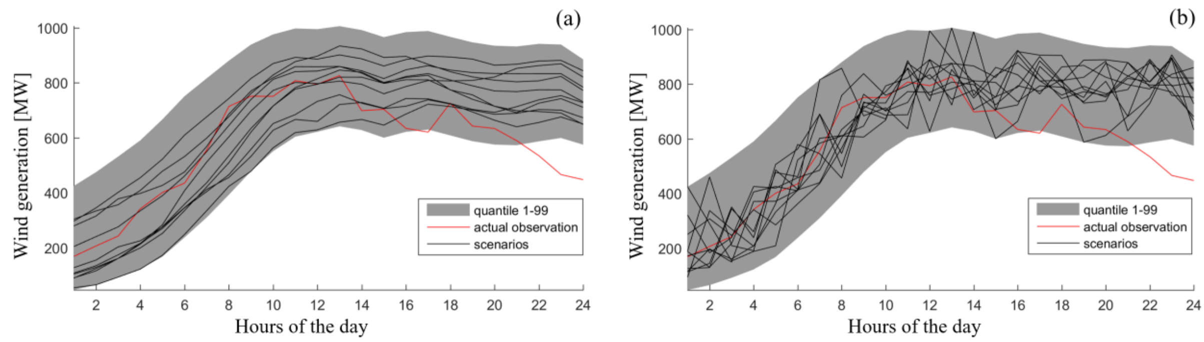


Figure 3.23 – Scenarios of wind generation for the seventh day of the considered week obtained using the copula-based sampling (a), and independent sampling (b) methods from the MLP prediction tool.

Firstly, it is observed that the quantiles provided by the BLSTM are tighter and more accurate than those issuing from the MLP. Then, it can be seen that the proposed copula-based sampling strategy allows to better capture the statistical information of the multivariate time-varying distribution of interest. Indeed, the independent sampling leads to scenarios with numerous sharp ramps that do not represent the smoother time profile of the aggregated wind power. In order to quantify this effect, the interdependence structure of forecast errors is studied. To that end, the autocorrelation function (ACF) of scenarios is compared with the one associated with the original variables. The ACF yields indeed the (linear) correlation between two values of the same variable at times *t* and *t+i*.

$$\mathrm{ACF}_t = \frac{\sum_{i=1}^{n-t}\left(d_i - \overline{d}\right)\left(d_{i+t} - \overline{d}\right)}{\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2} \tag{3.48}$$

The results are summarized in Table 3.3, where the mean ACF deviation (i.e. deviations between the ACF of the generated scenarios and the one of the actual data averaged on the first representative lags of the serial correlation) are presented.

Table 3.3
Temporal properties of generated scenarios (deviation of autocorrelation function on representative lags) with respect to historical observations.

	Copula-based sampling	Independent sampling
Wind	0.24	0.75
PV	0.17	0.92
Load	0.19	0.62
DA prices	0.09	0.63

The results show that the studied variables (wind, PV, load and electricity prices in the day-ahead market) do not come from a random processes (high values of autocorrelation between consecutive time steps), and that the copula-based sampling, contrary to the independent policy, appropriately captures this time-dependent information.

Visually (from Figures 3.20 to 3.23), the quality of scenarios seems more sensitive to the sampling strategy than to slight improvements in prediction accuracy. In this way, the scenarios generated with the (improved) copula-based sampling from the (basic) MLP probabilistic forecasts appear more realistic and accurate than scenarios generated with the (basic) independent sampling from the (improved) BLSTM probabilistic forecasts. Such an observation will be further investigated through a dedicated case study.

3.5.4 Value of probabilistic forecasts

Finally, we analyze the practical value of generating more accurate scenarios, by studying the economic benefits resulting from the use of these scenarios in the subsequent decision-making procedure. Here, the day-ahead optimization faced each day by an electricity retailer having its own renewable generation capacity is used as a case study. The portfolio is composed of one percent of the Belgian load (~ 140 MW of peak consumption) as well as twenty percent of the installed (onshore) wind (~ 350 MW) and PV (~ 600 MW) capacity. Then, a storage station (maximum output power of 50 MW, energy capacity of 250 MWh, and ramping capabilities of 10 MW/minute) is also considered so that dependencies between decisions at each time step of the scheduling horizon are important. Basically, the retailer aims at balancing its portfolio on a quarter-hourly basis (so as to avoid the financial penalties in case of imbalance) by exchanging (the surplus or deficit of energy) in the day-ahead electricity market (tariff arbitrage). In real-time, the retailer can then use the flexibility provided by the storage station to face prediction errors.

In this context of time-dependent decisions under uncertainty, we estimate the value of the different techniques to generate day-ahead scenarios, which is here realized through the procedure depicted in Figure 3.24.

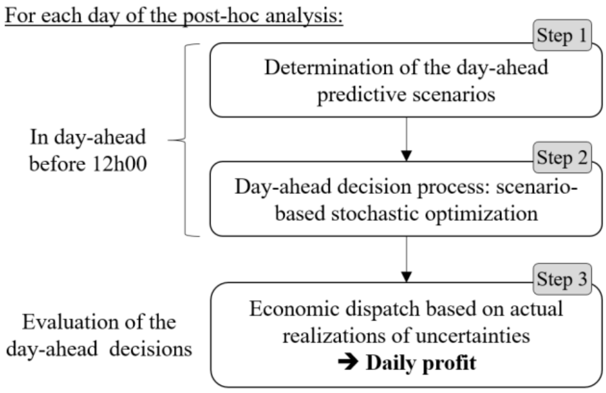


Figure 3.24 – Procedure used to compare the quality of day-ahead decisions based on the different techniques to characterize the forecast uncertainty.

The methodology is carried out for three different variants, which differ by the way scenarios are generated (Step 1):

1) MLP + copula-based sampling
2) Probabilistic BLSTM + independent sampling
3) Probabilistic BLSTM + copula-based sampling

The practical quality of scenarios is then analyzed through a post-hoc analysis, which consists in confronting the day-ahead decisions (obtained at the end of the day-ahead stochastic optimization of Step 2) with respect to the actual realizations (observations) of uncertainties during the whole month of January 2017. To that end, an economic dispatch (Step 3) has to be performed. The objective is to compute the profit actually generated based on the actual trajectories of uncertain variables as well as the day-ahead decisions, i.e. energy exchanged in the day-ahead market for each of the 24 hours. This procedure is performed for each day of the studied month, and the results (daily profit for the three investigated variants) are represented in Figure 3.25.

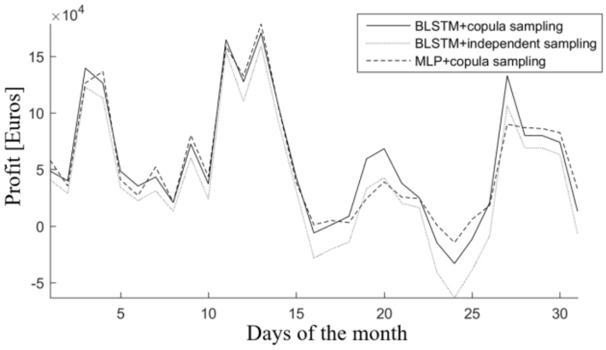


Figure 3.25 – Daily profit generated by the electricity aggregator with respect to the stochastic scenarios used to model uncertainties.

By comparing variants #2 and #3 (which use the same probabilistic forecasts but differ by the way the scenarios are generated), it can be concluded that using representative scenarios in the stochastic optimization process of step 2 (scenarios that account for the complex dependence structure among variables) is an highly important factor to take reliable decisions, which is here associated with an increase of profit of around $4*10^5$ Euros (i.e. relative increase of more than 10 %) over the simulated month. Moreover, the quality of predictions, in our case the fact of using the BLSTM neural networks (variant #3) instead of traditional feedforward networks (variant #1) with the same subsequent approach to generate the stochastic scenarios, plays also an important role to improve decisions in an uncertain environment. In this way, improved predictions (variant #3) enable decision makers to avoid taking overly conservative policies (so as to guarantee their robustness towards extreme scenarios). Here, the total profit throughout the considered month in variant #3 exceeds by $0.5*10^5$ Euros (~4%) the profit realized in variant #1.

3.6 Conclusions and perspectives

In this chapter, a new approach to generate short-term multivariate predictive scenarios is presented. The methodology attempts to address the main challenges associated with such a task, i.e. obtaining accurate forecasts that efficiently catch the contextual information contained in the explanatory variables, while capturing both temporal and cross-variable dependencies when generating scenarios. The results demonstrates that the proposed methodology yields accurate, calibrated forecast distributions learned from the historical dataset, and that the generated scenarios enables to increase the economic profit of energy aggregators participating in power markets.

As a next step (which is already under implementation), we investigate improvements in the prediction framework to tailor the architecture for shorter-term time horizons (e.g. wind forecasting for the next few seconds to minutes). Similarly, longer-term predictions (over a week) can be considered as well.

It may also be of interest to explore the benefit of further improving the sampling strategy to take into account that the contextual information can influence the interdependence structure of variables (e.g. stronger autocorrelation during windy days). Likewise, it would be interesting to perform a thorough (over a representative learning horizon) sensitivity analysis on the number of epochs required for efficiently adapting the network parameters to new data. The idea is to find the best trade-off between conserving reliable information from remote historical realizations and efficiently accommodating new patterns in the data.

3.7 Chapter publications

This chapter has led to the following publications:

- J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, “Improved Day-Ahead Predictions of Load and Renewable Generation by Optimally Exploiting Multi-Scale Dependencies,” in IEEE Innovative Smart Grid Technologies, Auckland, New-Zealand, 2017.

- J. Bottieau, F. Vallée, Z. De Grève and J.-F. Toubeau, "Leveraging Provision of Frequency Regulation Services from Wind Generation by Improving Day-Ahead Predictions using LSTM Neural Networks," in IEEE Energycon, Limassol, Chyprus, 2018.

- J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, "Deep Learning-based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets," in *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203-1215, March 2019.

CHAPTER 4

DAY-AHEAD STOCHASTIC OPTIMIZATION OF VIRTUAL POWER PLANTS

4.1 Introduction

Once the uncertainties associated with the day-ahead decision procedure of market players have been properly characterized (chapter 3), the optimal day-ahead scheduling of their generation/consumption/storage portfolio has to be determined so as to efficiently hedge against the resulting uncertainty space. Within the framework of this thesis, a particular interest is drawn to small-to-medium pumped storage hydro (PSH) stations. However, the operation of such units is significantly constrained by their limited energy capacity (typically 4 to 6 hours at full power to completely exhaust the storage capacity). Consequently, their economic potential is fully leveraged when included within an existing virtual power plants, i.e. energy aggregators optimized as a single entity (mono-agent centralized control), participating jointly in energy and balancing services markets (see chapter 2).

Market players are financially incentivized to improve their ability to address the different sources of uncertainties within their portfolio or even to help the system (through participation in ancillary services) when the real-time balance between the total generation and consumption is not maintained [Matevosyan[06], Zapata[14]]. Indeed, such behaviors enhance grid security and allow system operators to decrease the quantity of necessary power reserve for restoring the balance, which positively impacts the electricity bill of end-users by reducing the total transmission grid costs associated to these ancillary market services. It is therefore of general interest to improve portfolio management of power market participants, especially as it can also contribute to the emergence of new actors investing in renewable energies. They can indeed rely on robust tools for managing risk in power markets, which will overall accelerate the energy transition.

Currently, a large part of the literature devoted to the day-ahead decision procedure of market players relates to the challenges associated with a single wind producer [Baringo[13], Bathurst[02], Botterud[12], Liang[11], Morales[10,b], Pinson[07], Sánchez de la Nieta[14], Zugno[13]] or a single hydroelectricity producer [De Ladurantaye[07], Pousinho[13]]. Then, another significant part of the proposed works studies the combined and coordinated use of technologies with complementary characteristics such as wind generation and energy storage [Abreu[12], Castronuovo[04], Ding[12], Sánchez de la Nieta[13], Khodayar[13,a]] or wind and conventional generation [Hellmers[16]].

Generally, these studies tend to focus on one fixed type of portfolio, and as a result, the developed methodologies often fail to consider all market opportunities such as using part of the flexibility for helping the system in real-time (during the imbalance settlement) or by participating in the different ancillary services. In this context, the proposed formulation is here designed to be as general as possible in order to suit any portfolio configuration. In this way, similarly to [Amin Tajeddini[14], Mashhour[11], Pandzic[13,b]], we do not make any assumption on the portfolio constitution, and the methodology is built to incorporate any type of electricity generation, consumption or source of flexibility (storage technologies, demand response, etc.). The objective is moreover to be able to consider all market opportunities, i.e. both energy markets (day-ahead, intraday, imbalance settlement) and ancillary services (including spinning and non-spinning reserves[24]). Finally, in contrast with traditional works that only consider a limited number of stochastic variables (typically wind generation and day-ahead electricity prices), it is here intended to include all relevant sources of uncertainty, while accounting for the risk in the decision strategy.

However, the main contribution targeted in this part is the proper consideration of the uncertain amount of energy that will be called on for operating reserves in the day-ahead scheduling of Virtual Power Plants. In most studies, the amount of balancing power reserves requested in real time is either neglected [Ugedo[06], Vargas-Serrano[17]], or simplified to a single hourly value [Chazarra[17], Connolly[11], Kazempour[09]]. In [Chazarra[18]], the uncertainty in the price of the aFRR capacity (but not regarding the actual activated energy) is added into a stochastic decision procedure. Here, we aim at generating adequate scenarios of the energy requested for ancillary services, and also to subsequently integrate them in a new formulation that jointly includes technical and economic effects arising from the uncertain real-time activation of allocated reserves. Specifically, the work takes into account both revenues from the actual provision of reserves and the variable cost structure of all considered technologies. This allows to obtain a cost-optimal allocation of assets to the different ancillary services over the scheduling horizon. In this way, downward regulation can be economically supplied by expensive-to-run units (due to operational cost savings), while high-performance technologies are cost-optimal for delivering the frequently activated reserves.

These reserves were historically provided by conventional power plants due to their ability to efficiently modify their output power (high ramping rates). However, environmental considerations (i.e. reduce pollution and health detriment caused by coal fired power plants, and, to a lesser extent, gas-fired power plants) associated with a strong willingness to open up the market of ancillary services to all flexible resources within the grid, is driving the emergence of new actors. The most popular ones are currently storage units, demand response strategies, and modulation of the output power of renewable energies through power electronic devices. The participation of these technologies to balancing services is however more complex due to their limited energy capacity. Indeed, for storage units, the uncertain deviations in the output power (due to the real-time activation of reserves to restore grid frequency) may lead to unintended amount of energy stored at the end of the planning horizon, which thus affect the economic value of the storage power plant for the following days. Such deviations may also

[24] The non-spinning reserve is defined as extra capacity that remains offline unless it is requested by the system operator in the framework of mFRR. From a practical point of view, this type of reserve is more complex to integrate in the formulation of the day-ahead scheduling of virtual power plants.

unexpectedly empty or fill up the available capacity during the considered period and therefore prevent the unit to fulfill its operational schedule or to provide the contracted operating reserves. Then, the participation of demand response (deferrable load) to operating services is hampered by two main challenges, namely the necessity to comply with the load intertemporal constraints (some processes have fixed load profiles and/or cannot be stopped during operation), while efficiently accounting for the load recovery effect, i.e. the fact that the load curtailment infers postponement in the energy consumption [Pourmousavi[14]]. Finally, the contribution of renewable energy sources (RES) in flexibility services is limited by the reliability of power forecasts [Morren[06]].

Overall, the probabilistic nature of the energy requested in the ancillary service markets induces uncertainty on both the profit (revenues from the real-time activation and expenses due to operating costs) and the energy capacity of participating resources. Considering worst-cases in the optimization, i.e. that the energy can be fully activated in either upward or downward direction during the whole scheduling horizon, prevents small-to-medium sized resources to participate in the service. Henceforth, in order to fully leverage the flexibility abilities of energy-constrained assets (so as to increase their economic value), it is important to efficiently model the uncertain fluctuations resulting from the real-time activation of operating reserves, and to incorporate them in a tailored stochastic scheduling procedure. This allows indeed to avoid relying on sub-optimal overly conservative approaches that aims to ensure availability of committed energy in case of full deployment of the reserve during the whole scheduling horizon.

The stochastic optimization framework is discussed in Section 4.2. Then, Section 4.3 introduces the methodology to model the uncertainties, whereas Section 4.4 presents the formulation of the day-ahead stochastic scheduling of Virtual Power Plants that considers the uncertainty of real-time procurement of ancillary services. In Section 4.5, the mathematical methodology to solve the resulting problem is briefly discussed, and finally, Section 4.6 presents the case study that is designed to estimate the added value of the developed model with regard to traditional approaches that lead to overly conservative solutions. Relevant conclusions and perspectives are summarized in Section 4.7.

4.2 Stochastic optimization

The main objective of the optimization model is to obtain a rational scheduling for allocating the available resources to the different markets platforms, while ensuring reliability of the results, i.e. the operational constraints must be adequately modeled so as to reflect the actual technical requirements of the studied system and to avoid impractical outcomes. From this, a trade-off between the performance of the model and the associated simulation time has to be determined. However, decisions need to be made without perfect information about the decision-making problem, and we have to resort on stochastic optimization. Different techniques exists, and are here briefly described:

Scenario-based optimization

In this framework, each uncertain parameter is modeled as a random variable and represented by a finite set of discrete scenarios [Conejo[10]]. It is critical to generate a sufficient number of representative scenarios that cover the possible realization of the underlying stochastic processes. The objective function is not a single value but a random variable, and the problem is traditionally formulated as the optimization of the expected value of this function, e.g. maximizing the mean profit of the virtual power plant over the scheduling horizon.

Interval optimization

Instead of sampling scenarios, the interval optimization uses confidence intervals under the form of upper and lower bounds to represent the uncertainty space [Yu05], and therefore does not require the probability distributions of uncertain variables. The method then derives optimistic and pessimistic solutions for satisfying the problem constraints [Wu12].

Robust optimization

The stochastic variables are modeled by an uncertainty set that is specified a priori. The shape of this uncertainty set is chosen so as to fit the historical data, but is typically selected to obtain a linear problem (e.g. rectangular set when two random variables are considered). The problem is optimized with respect to the worst-case realization within the uncertainty set, but the outcome remains feasible for any point contained in the uncertainty set [Sun17]. In comparison with the scenario-based approach (which leads to the best solution in expectation), robust optimization is a more conservative method, but has the advantage to be associated with lower calculation times.

Chance-constrained optimization

This is a less conservative version of the robust approach, which recently became popular due to its pragmatic (in terms of computational burden) characteristics. In this technique, the principle is to consider a probability for satisfying each constraint (e.g. based on all potential realizations of the real-time activation of balancing reserves, the probability that the energy limits of storage units are not violated should be equal to or higher than 95%). Such a formulation is very difficult to solve in practice (due to the non-convexity of the resulting probabilistic constraints), unless it is considered that the uncertainty set is following a Gaussian or a rectangular probability distribution function [Wu14].

Distributionally-robust optimization

This modern technique is an improved version of the chance-constrained and robust approaches. Instead of a single known uncertainty set, we have to deal with an ambiguity set that includes an infinite number of uncertainty sets [Bian15, Wei16]. This framework is well-suited when the exact distribution of the uncertainty at hand is unknown. In such a case, the objective is to optimize the objective function against the worst-case distribution, while the obtained solution remains feasible for any other distribution within the ambiguity set.

The day-ahead scheduling of VPPs has to be carried out on a daily basis and it is thus important that the formulation does not lead to conservative solutions. Indeed, the savings realized in case of realization of the worst-case scenario is unlikely to compensate the accumulation of opportunity losses during the other more classical days. Then, **since providers of balancing capacity are disqualified if they fail to provide the service (Section 2.5), it is important to ensure that the provision of ancillary services is satisfied for all possible scenarios. Henceforth, the scenario-based approach is here privileged.**

The proposed day-ahead model has a multi-stage structure, and is constructed based on the actual decision process (subdivision into three consecutive trading floors) faced by a portfolio manager, namely:

1. The VPP operator must submit its offering curve in the day-ahead market and establish the schedule of its slow (inflexible) generators before knowing the actual realization of

stochastic parameters (market prices, load and generation within the portfolio) influencing the portfolio operation.
2. The VPP operator must decide on the operation of its fast (flexible) conventional power plant and storage utilities once the uncertainty is resolved, which corresponds to the real time economic dispatch of the portfolio. It should be noted that the imbalance price remains unknown at this stage.
3. The imbalance tariffs are revealed. However, no decision has to be made at this third stage. As a result, the stochastic imbalance prices can be replaced by their expectations in the preceding stage and the proposed model can be expressed as a two-stage stochastic programming problem.

The two-stage decision procedure is sketched in Figure 4.1, where ω is the scenario index, N_Ω is the number of scenarios considered, and Ω represents the set of all scenarios.

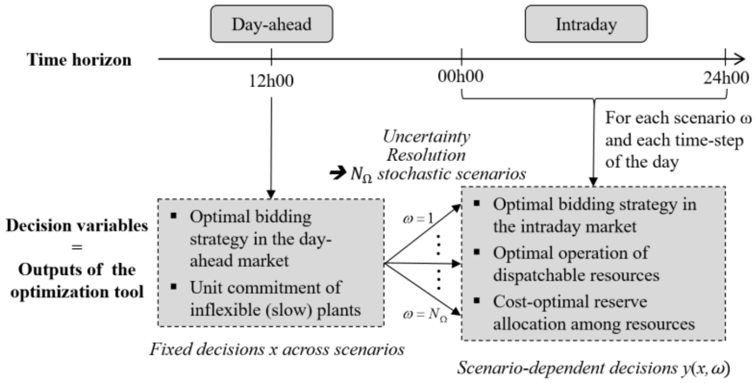


Figure 4.1 – Day-ahead decision procedure of Virtual Power Plants.

In the two-stage stochastic decision-making process, we differentiate two different types of decisions:
1. First-stage or here-and-now decisions x. These decisions are taken before the realization of the uncertainties Ω. It results that these here-and-now decisions variables are scenario-independent (do not depend on the realization of the stochastic processes).
2. Second-stage or wait-and-see decisions $y(x, \omega)$. These decisions are made after the actual realization of uncertainties is disclosed/revealed. Consequently, these decisions are scenario-dependent (i.e. these are differentiated for each single scenario), and depend upon both the first stage decisions and the realization of the stochastic variables.

A two-stage stochastic **linear programming** problem can be generally expressed as presented in (4.1). It should be noted that a nonlinear version can be straightforwardly deducted [Conejo[10]].

$$\begin{aligned} \min_{x} \quad & c^{\mathrm{T}} x + \mathrm{E}\left\{Q_\omega\right\} \\ \text{subject to} \quad & Ax \leq b \\ & x_{\min} \leq x \leq x_{\max} \\ & x \in \Re^{n_1} \end{aligned} \tag{4.1}$$

where Q_ω is the optimal value of the second stage problem (referred to as recourse problem), which is formulated as follows:

$$\begin{aligned} Q_\omega = \min_y \; & q_\omega^T y_\omega \\ \text{subject to} \quad & T_\omega x + W_\omega y_\omega \le h_\omega \\ & y_\omega \in \Re^{n_2} \end{aligned} \tag{4.2}$$

where x is the vector of the n_1 first-stage decision variables (limited by lower bounds $x_{\min}$ and upper bounds $x_{\max}$), whereas y_ω encompass the n_2 second-stage decision variables of the scenario ω. Since there are N_Ω second-stage sub-problems, one for each considered scenario, the problem is therefore characterized by $n_1 + N_\Omega n_2$ decisions variables. Then, c, q_ω, b, h_ω, A, T_ω, and W_ω are known vectors and matrices of appropriate size (regarding the problem formulation).

The two-stage stochastic problem can be equivalently expressed as the following deterministic program [Birge97]:

$$\begin{aligned} \min_{x, y_\omega} \quad & c^{\mathrm{T}} x + \sum_{\omega \in \Omega} \pi_\omega \left(q_\omega^T y_\omega \right) \\ \text{subject to} \quad & Ax \le b, \\ & T_\omega x + W_\omega y_\omega \le h_\omega, \; \forall \omega \in \Omega \\ & x \in \Re^{n_1}, \; y_\omega \in \Re^{n_2}, \; \forall \omega \in \Omega \end{aligned} \tag{4.3}$$

where π_ω is the probability of occurrence associated with scenario ω. Such a formulation intrinsically presents an exploitable structure that is well suited for decomposition (so as to accelerate the convergence of the resolution algorithm).

Illustrative example

The following example is drawn from [Conejo10], and is aimed at illustrating the two-stage stochastic formulation. It is centered on an industrial electricity consumer that participates in a pool-based electricity market to cover its load throughout a week (168 hourly periods). It is facing uncertainty regarding both its future demand and the pool-based market price, but it is assumed for simplicity that these stochastic parameters are constant over the considered week. The buyer has moreover the possibility to purchase up to 90 MW at 45 €/MWh (the same quantity is purchased for each period of the week), through bilateral contracting in week-ahead (before knowing its actual consumption and pool prices). The uncertainty associated with the decision process (predictive scenarios of demand and pool price) is provided in Table 4.1

Table 4.1
Scenario data for illustrative example.

Scenarios	Probability of occurrence	Demand [MW]	Price [Euros/MWh]
ω_1	$\pi_1 = 0.2$	110	50
ω_2	$\pi_2 = 0.6$	100	46
ω_3	$\pi_3 = 0.2$	80	44

To summarize the decision procedure (Figure 4.2), at the first stage, the consumer has to decide how much to buy from the bilateral contract (ignoring both its future consumption and the electricity prices in the pool-based market). Then, during the second stage, knowing its electrical load, it has to buy the remaining quantity (at the corresponding market price) to cover its load. It should be noted that the possibility to buy energy in week-ahead to sell it afterwards in the pool market is not considered.

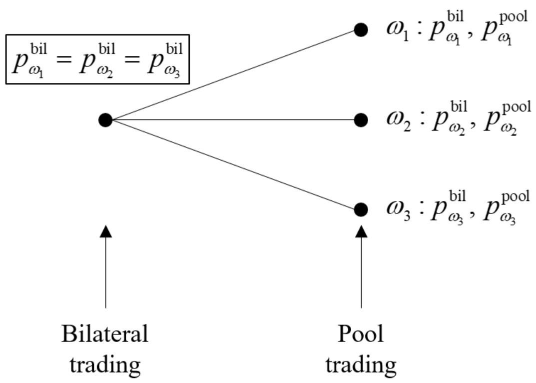


Figure 4.2 – Decision scenario tree associated with the illustrative example [Conejo[10]].

This two-stage stochastic programming problem can be formulated as the minimization of the expected cost faced by the consumer to supply its uncertain demand.

$$\min 168\times\left[0.2\times\left(45p_{\omega_1}^{\text{bil}}+50p_{\omega_1}^{\text{pool}}\right)+0.6\times\left(45p_{\omega_2}^{\text{bil}}+46p_{\omega_2}^{\text{pool}}\right)+0.2\times\left(45p_{\omega_3}^{\text{bil}}+44p_{\omega_3}^{\text{pool}}\right)\right]$$

subject to:

$$p_{\omega_1}^{\text{bil}} + p_{\omega_1}^{\text{pool}} \geq 110$$
$$p_{\omega_2}^{\text{bil}} + p_{\omega_2}^{\text{pool}} \geq 100$$
$$p_{\omega_3}^{\text{bil}} + p_{\omega_3}^{\text{pool}} \geq 80$$
$$0 \leq p_{\omega_1}^{\text{bil}}, p_{\omega_2}^{\text{bil}}, p_{\omega_3}^{\text{bil}} \leq 90$$
$$0 \leq p_{\omega_1}^{\text{pool}}, p_{\omega_2}^{\text{pool}}, p_{\omega_3}^{\text{pool}}$$
$$p_{\omega_1}^{\text{bil}} = p_{\omega_2}^{\text{bil}} = p_{\omega_3}^{\text{bil}}$$

The first-stage decision variables p^{bil} represent the power purchased through the bilateral contract for the three considered scenarios. The second-stage variables p^{pool} stand for the power bought in the pool-based market.

The first three constraints enforce energy supply for the three scenarios of demand, whereas the two following constraints express the bounding limits of the bilateral trading and pool-based market respectively. Finally, **the last constraint describes the non-anticipativity condition translating the fact that the bilateral contract is independent of the scenario realization** (and has to be fulfilled whatever the future realization of stochastic parameters).

The solution to this problem is characterized by a consumer buying p^{bil} = 80 MW using the bilateral contract.

Including risk in the stochastic decision procedure

Traditional two-stage stochastic programming is risk-neural. Indeed, by optimizing the expected value of the objective function, the formulation does not consider the remaining parameters characterizing the distribution. In this way, in the context of profit maximization of a Virtual Power Plant, even though the mean profit is positive, it is possible that some scenarios lead to negative profits or losses. The effect of risk can be taken into account by controlling the shape of the profit distribution, in particular the probability of experiencing low revenues. The purpose is to obtain the best trade-off between the conflicting objectives of maximizing both the expected profit and the worst-case profit scenarios (typically increase one contribution reduces the second), while satisfying the technical constraints of the problem. Practically, the most common method of managing the risk is to include in the formulation a term measuring the risk related to the profit distribution. Here, the conditional value-at-risk (*CVaR*) is used .

As represented in Figure 4.3, for a given $\alpha \in [0, 1]$, the conditional value-at-risk (*CVaR*) is defined as the expected value of the profit smaller than the $(1-\alpha)$ % quantile of the profit distribution [Rockafellar00].

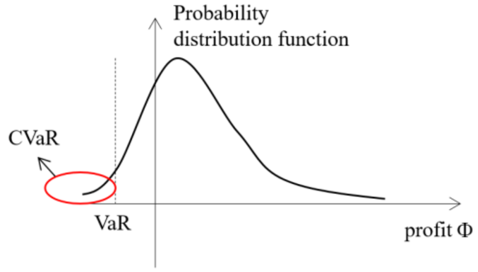


Figure 4.3 – Illustration of the conditional value-at-risk.

In the context of the maximization of the profit Φ, the $CVaR_\alpha$ is the expected value of the profit lower than the Value-at-risk (VaR_α).

$$CVaR_\alpha(\Phi) = \mathrm{E}\left(\Phi \mid \Phi \le VaR_\alpha(\Phi)\right) \tag{4.4}$$

where the *VaR* at confidence level α, noted VaR_α is the upper bound of the 100(1-α) % least profitable scenarios of the profit distribution, i.e. the largest value η ensuring that the probability of having a profit lower than η is less than 1-α.

$$VaR_\alpha = \max\left\{\eta \middle| P(\Phi \le \eta) \le 1-\alpha\right\} \tag{4.5}$$

The $CVaR_\alpha$ can be expressed by means of linear expressions [Rockafellar00]. To that end, it suffices to introduce an additional positive variable z_ω (for each scenario), which measures negative profit deviations from VaR_α.

$$z_\omega \ge 0,\ \forall \omega \in \Omega \tag{4.6}$$

$$z_\omega \ge VaR_\alpha(\Phi_\omega) - \Phi_\omega,\ \forall \omega \in \Omega \tag{4.7}$$

Hence, the $CVaR_\alpha$ can be expressed as:

$$CVaR_\alpha = VaR_\alpha - \frac{1}{1-\alpha}\sum_{\omega\in\Omega}\pi_\omega z_\omega \tag{4.8}$$

The *CVaR* can be included into the formulation (4.3) using $\beta = \{0, 1\}$ as the trade-off between the maximizations of the expected profit and of the average profits in the 1-α % of scenarios with the lowest revenues.

$$\begin{aligned}
\min_{x,y_\omega} \; & (1-\beta)\left(c^{\mathrm{T}}x+\sum_{\omega\in\Omega}\pi_\omega\left(q_\omega^T y_\omega\right)\right)+\beta\left(VaR_\alpha-\frac{1}{1-\alpha}\sum_{\omega\in\Omega}\pi_\omega z_\omega\right) \\
\text{subject to} \quad & Ax\le b, \\
& T_\omega x+W_\omega y_\omega\le h_\omega,\ \forall\omega\in\Omega \\
& VaR_\alpha-\left(c^{\mathrm{T}}x+q_\omega^T y_\omega\right)\le z_\omega,\ \forall\omega\in\Omega \\
& z_\omega\ge 0,\ \forall\omega\in\Omega \\
& x\in\Re^{n_1},\ y_\omega\in\Re^{n_2},\ \forall\omega\in\Omega
\end{aligned} \tag{4.9}$$

where z_ω is a continuous non-negative variable.

Challenges associated with scenario-based stochastic optimization

Currently, stochastic programming still presents two important challenges [Papavasiliou[11]].

The first one consists in choosing the most appropriate set of scenarios modeling the complex dependency structure of the stochastic variables (Section 4.3). Each scenario has to be statistically representative of the interdependencies among variables while conserving the information about the time structure (i.e. autocorrelation) of each variable individually.

The second challenge is to reduce the computational burden of the problem due to its numerous degrees of freedom. This challenge is particularly pronounced in two cases, namely when some decision variables are binary and when the objective function or some constraints are nonlinear. Here, the sources of nonlinearity are appropriately linearized and the number of integer decision variables is minimized thanks to an efficient modeling framework, capitalizing on the knowledge reported in the existing literature. The problem is indeed formulated as a compact mixed-integer linear program (Section 4.4).

4.3 Scenarios generation

In day-ahead, the uncertain variables need to be predicted as accurately as possible in order to take adequate market decisions while scheduling the optimal sequence of actions of portfolio units for the next day. The robustness of these decisions is ensured by the decision-making procedure that statistically accounts for the possible deviations using a set of representative scenarios $\omega \in \Omega$. It is therefore essential to precisely model and generate these scenarios. However, it may not be relevant to rely on an advanced forecasting tool for all uncertain variables, especially the real-time grid imbalance or the amount of reserve that will have to be activated by the system operator, since these values are highly volatile and governed by ultra-short-term events. The prediction accuracy of such variables over a multi-hours horizon is thus questionable. The distinction between the two set of variables is presented in Table 4.2.

Table 4.2
Stochastic variables included into the formulation.

Set 1: Variables predicted using the BLSTM-based approach	**Set 2**: Endogenous variables modeled using statistically representative scenarios
Wind generation PV generation Load Electricity prices in the day-ahead market	Electricity prices and liquidity in the intraday market Grid imbalance Amount of energy requested in real-time for FCR, aFRR and mFRR

The generation of the predictive scenarios (for stochastic variables pertaining to set 1) is presented in Chapter 3. As a reminder, the probabilistic forecasts (under the form of intervals) for the 24 hours of the following day for all considered variables are first obtained using a BLSTM-based approach. Then, the multivariate scenarios are generated by sampling the resulting high-dimensional distribution.

The procedure for identifying the representative scenarios (of the endogenous uncertain parameters pertaining to set 2) is structured around two consecutive steps. The first one consists in identifying the relevant set of historical data for constructing the model. Then, the type of model has to be appropriately selected and the building phase (i.e. model training) can be carried out. Since market and grid conditions are currently in a transition state mainly due to the increasing contribution of renewable-based generation in the energy mix, the model is trained using relevant data of at most 2 years old (so as to avoid modeling irrelevant past regimes).

Traditionally, the different uncertain variables of the problem are generated independently [Glasserman[04]]. This originates from the fact that the temporal dynamics may significantly vary among the different variables, which necessitates diverse modeling tools (e.g. Markov chains, autoregressive (integrated) moving-average models, neural networks, etc.) that cannot be easily coupled. This approach may accurately model the statistical properties of each variable individually but ignores the dependency structure between variables and can therefore lead to unrealistic scenarios. Moreover, as represented in Figure 4.4, this independent generation yields a multiplying effect of the number of scenarios when the number of stochastic parameters increases. This issue can be addressed by using scenario reduction techniques [Römisch[03]], but at the expense of less likely scenarios. These may potentially generate extreme outcomes (highly profitable or highly adverse) and should therefore not be neglected.

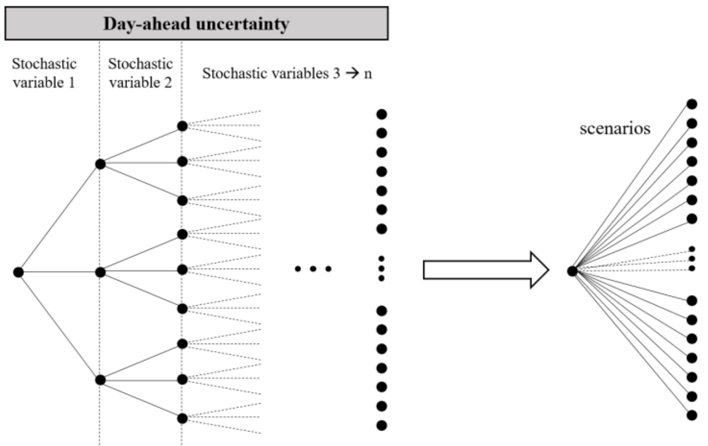


Figure 4.4 – Independent generation of scenarios.

Against this background, a comprehensive model including all uncertain variables may not only increase the quality of the model but also reduce the number of scenarios required for capturing the uncertainty sphere [Tastu[13]]. Here, a probabilistic approach is envisaged, and involves representing the multivariate distribution function with a large number of dimensions in order to encompass the temporal information of all variables of the problem.

Firstly, the number of dimensions is thus preliminarily decreased for each endogenous variable separately using principal component analysis (PCA). This technique attempts indeed to find a linear subspace of lower dimensionality than the original space, in which the new dimensions include the largest amount of information about the original dataset.

Secondly, a copula model is used to incorporate the complex dependence structure among variables [Sklar59]. The model encompasses both the temporal transitions characteristics of each single parameter as well as the dependence pattern among variables. It is worth noting that the model intrinsically catches fluctuations of the dependence structure over time (e.g. a stronger correlation between variables at some specific periods of the day). The method used to estimate the copula model and generate subsequent random vectors (or scenarios, or time trajectories) with the multivariate distribution of interest is based on [Strelen07, Strelen09], in which the copula is estimated thanks to a non-parametric (empirical) approach. Such a parameter-free method offers a greater generality allowing to represent any type of dependence.

Illustrative example

The methodology for generating scenarios of endogenous variables (pertaining to set 2 of Table 4.2) is illustrated in a simple bivariate case including the imbalance tariffs and the real-time activation of upward aFRR for a typical day of January.

Both these variables are measured with a quarter-hourly time step, leading to 96 daily data. The first stage of the modeling procedure presented above consists therefore in decreasing the dimensionality of the uncertainty space through a PCA analysis. The transformation is indeed such that the first principal component accounts for the largest possible variance contained in the original data, and each succeeding component comprises, in turn, the largest possible part of the remaining information (with the constraint that each component is orthogonal to the preceding one). The percentage of variability associated with each component for imbalance prices and activation of upward aFRR is represented in Figure 4.5.

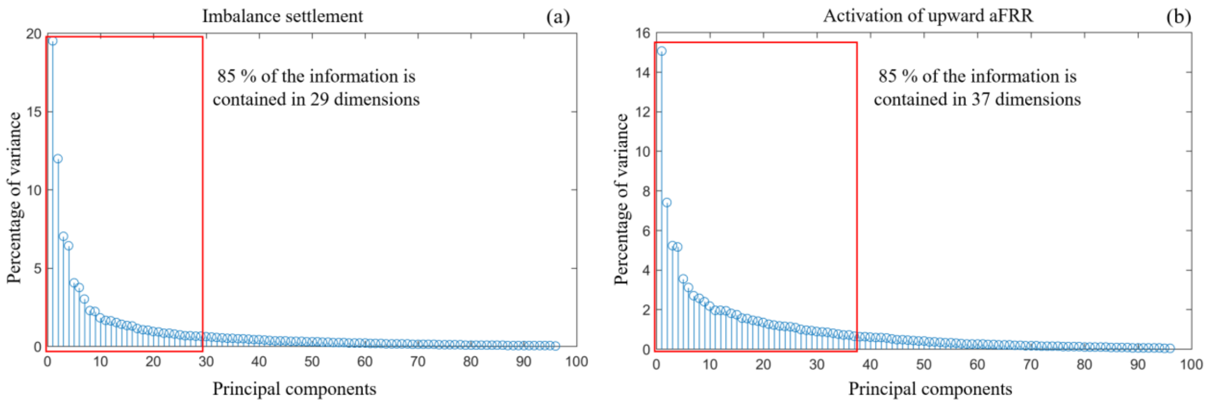


Figure 4.5 – Outcomes from the principal component analysis for the imbalance price variable (a), and the upward activation of aFRR (b).

It is interesting to observe that 85 % of the information is captured using around 35 dimensions, thereby demonstrating that a lot of redundant information is included within the studied time series. A copula model including both reduced-size variables is then constructed based on historical measurements (months of January 2015 and 2016). This model can then be exploited to generate representative scenarios that account for both time and cross-variable dependencies. Here, 10 scenarios aiming at representing a typical day of January 2017 are generated, and the comparison between the actual realizations of 2017 with the simulated time trajectories (using the copula model trained on 2015-2016 data) is illustrated in Figure 4.6 for the imbalance tariffs and in Figure 4.7 for the amount of aFRR activated by the system operator.

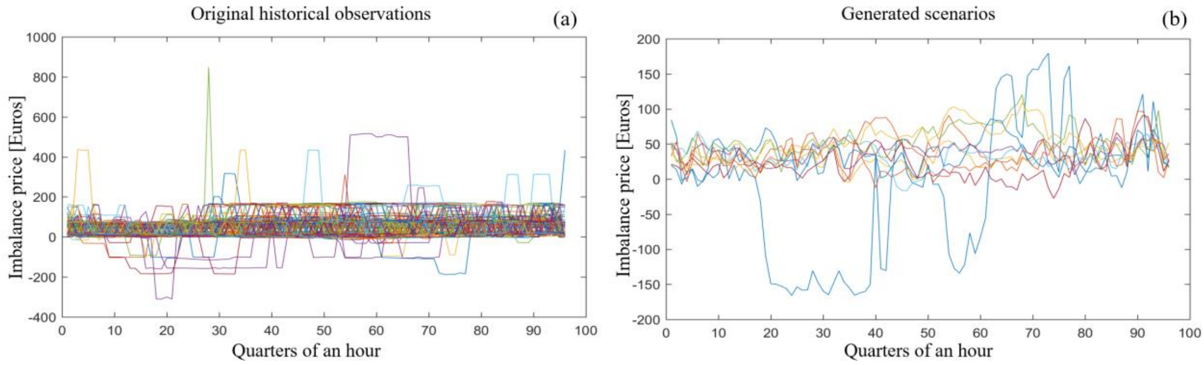


Figure 4.6 – Test set (a) and generated data (b) for the imbalance tariffs of January 2017.

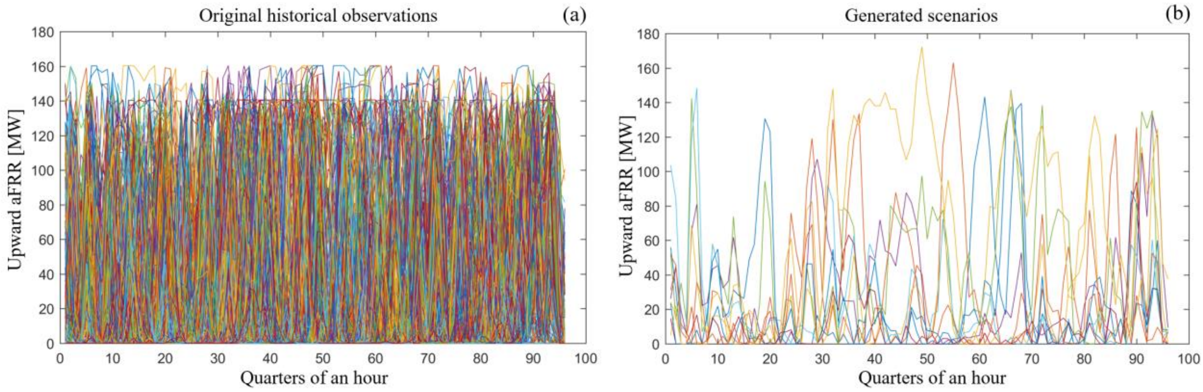


Figure 4.7 – Test set (a) and generated data (b) for the imbalance tariffs of January 2017.

It can be observed that the generated scenarios appear statistically close to the actual observations. In order to mathematically endorse this visual conclusion, the autocorrelation function (ACF) of both series are represented in Figure 4.8.

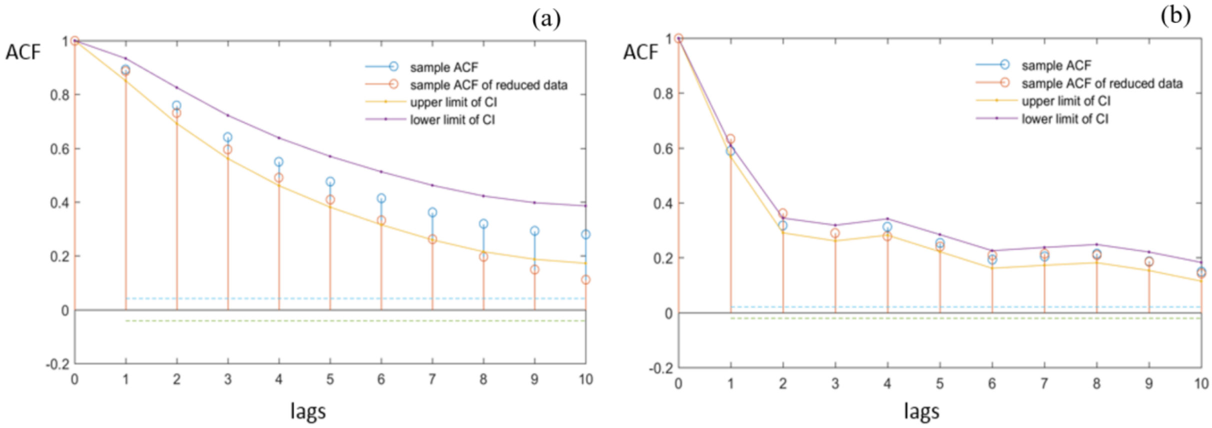


Figure 4.8 – Comparison of Autocorrelation functions (ACFs) between the training set and generated scenarios for the imbalance settlement (a) and activation of upward aFRR (b)

4.4 Mathematical formulation

The nomenclature relative to the stochastic formulation of the day-ahead scheduling of Virtual Power Plants (containing pump storage hydro stations) is exposed hereunder. When the formulation refers to other variables, the latter are properly introduced in the text.

Sets and indexes

$\tau \in \mathrm{T}$	Time step
$\omega \in \Omega$	Stochastic scenario
$g \in G$	Conventional power plants
$g_s \in G_s \subseteq G$	Slow conventional power plants
$g_f \in G_f \subseteq G$	Fast conventional power plants
$p \in PSH$	Pumped-storage hydro station
$\upsilon \in \Upsilon \equiv (G \cup PSH)$	Conventional and storage plants
$dr \in DR$	Demand response technology
$r \in RES$	Renewable generation technology
$s \in \Sigma$	Reserve product
$\Sigma^{\mathrm{up}} \subseteq \Sigma$	Upward reserve product
$\Sigma^{\mathrm{down}} \subseteq \Sigma$	Downward reserve product

Mid-term decision variables (= parameters in the day-ahead optimization)

$\mathrm{Res}_s^{\mathrm{tot}}$	Total balancing capacity reserved by the VPP
$\mathrm{E}_\tau^{\mathrm{LTM}}$	Energy exchanged on the long-term markets

Decision variables (in bold characters)

$\mathbf{Res}_{s,\upsilon,\omega,\tau}$	Allocated reserve capacity
$\mathbf{Res}_{s,\upsilon,\omega,\tau}^{\mathrm{on}}$	Spinning reserve capacity
$\mathbf{Res}_{s,g,\omega,\tau}^{\mathrm{su}}$	Non-spinning reserve capacity
$\mathbf{Res}_{s,g,\omega,\tau}^{\mathrm{sd}}$	Spinning reserve capacity through shut-down
$\mathbf{Res}_{s,\omega,\tau}^{\mathrm{dr}}$	Reserve capacity offered by demand response
$\mathbf{E}_\tau^{\mathrm{DAM}}$	Energy exchanged on the day-ahead market
$\mathbf{E}_{\omega,\tau}^{\mathrm{IM}}$	Energy exchanged on the Intraday market
$\mathbf{E}_{\omega,\tau}^{\mathrm{imb}}$	Imbalance of the portfolio
$\mathbf{P}_{r,\omega,\tau}^{\mathrm{cut\text{-}off}}$	Renewable output power curtailed
$\boldsymbol{\alpha}_{g,\omega,\tau}^{\mathrm{com}}$	Binary variable indicating the committed status of the unit (before activation of reserves)
$\boldsymbol{\alpha}_{g,\omega,\tau}^{\mathrm{real}}$	Binary variable indicating the actual status of the unit (after activation of reserves)
$\boldsymbol{\alpha}_{g,\omega,\tau}^{\mathrm{su}}$	Binary variable indicating if the unit is starting up
$\boldsymbol{\alpha}_{g,\omega,\tau}^{\mathrm{sd}}$	Binary variable indicating if unit is shutting down
$\boldsymbol{\alpha}_{g,\omega,\tau}^{\mathrm{off,su}}$	Binary variable indicating if the unit is allocated to start-up (for reserve provision)
$\boldsymbol{\alpha}_{g,\omega,\tau}^{\mathrm{on,sd}}$	Binary variable indicating if the unit is allocated to shut-down (for reserve provision)
$\boldsymbol{\delta}_{g,\omega,\tau}^{\mathrm{com}}$	Output power (above the minimum level)
$\boldsymbol{\alpha}_{p,\omega,\tau}^{\mathrm{turb}}$	Binary variable indicating the turbining status
$\boldsymbol{\alpha}_{p,\omega,\tau}^{\mathrm{pump}}$	Binary variable indicating the pumping status
$\mathbf{P}_{p,\omega,\tau}^{\mathrm{turb}}$	Output power in turbine mode
$\mathbf{P}_{p,\omega,\tau}^{\mathrm{pump}}$	Output power in pump mode
$\mathbf{c}_{p,\omega,\tau}^{\mathrm{SU,pump}}$	Pump start-up costs

$\mathbf{c}_{p,\omega,\tau}^{\mathrm{SU,turb}}$	Turbine start-up costs
$\mathbf{c}_{p,\omega,\tau}^{\mathrm{SD,pump}}$	Pump shut-down costs
$\mathbf{c}_{p,\omega,\tau}^{\mathrm{SD,turb}}$	Turbine shut-down costs
$\mathbf{P}_{\omega,\tau}^{\mathrm{DR}}$	Output power of demand response

Parameters

α_{CVaR}	Confidence level used for the *CVaR*
β	Risk-aversion parameter
π_{ω}	Probability of occurrence of scenario ω
Δ_{τ}	Time resolution, in hours
$\lambda_{s}^{\mathrm{res}}$	Price for availability of reserves, in €/MWh
$\lambda_{\omega,\tau}^{\mathrm{DAM}}$	Price of day-ahead market, in €/MWh
$\lambda_{\omega,\tau}^{\mathrm{IM}}$	Price of Intraday market, in €/MWh
$\lambda_{\tau}^{\mathrm{LTM}}$	Price of long-term markets, in €/MWh
$\lambda_{\omega,\tau}^{\mathrm{imb}}$	Penalty price, differing between positive and negative imbalances, in €/MWh
NRV_{ω}	Net regulation volume within the system, in MW
$\kappa_{1,\omega,\tau}$	Risk-aversion parameter in case of positive system imbalance
$\kappa_{2,\omega,\tau}$	Risk-aversion parameter in case of negative system imbalance
$\lambda_{s}^{\mathrm{act}}$	Price for activation of reserve, in €/MWh
$\lambda_{r}^{\mathrm{GC}}$	Price of green certificates, in €/MWh
$\lambda_{\upsilon}^{\mathrm{inj}}$, $\lambda_{\upsilon}^{\mathrm{cons}}$	Grid fees for generated (consumed) energy, in €/MWh
$R_{\omega,\tau}^{s}$	Proportion of reserve (∈ [0, 1]) requested by the system operator
C_{g}^{min}	Minimal operating costs of conventional generation units
C_{g}^{marg}	Marginal operating costs of conventional generation units
P_{g}^{min}, P_{g}^{max}	Minimum (maximum) stable output power, in MW
R_{g}^{s}	Maximum power variation for service *s*, in MW
SU_{g}, SD_{g}	Maximum start-up (shut-down) rate, in MW
C_{p}^{op}	Variables operating costs, in €/MWh
$\mathrm{SOC}_{p}^{\mathrm{min}}$	Minimum state-of-charge, in MWh
$\mathrm{SOC}_{p}^{\mathrm{max}}$	Maximum state-of-charge, in MWh
$\mathrm{SOC}_{p}^{\mathrm{target}}$	Target value of the state-of-charge (at the end of the scheduling horizon), in MWh
$C_{p}^{\mathrm{SU,pump}}$	Costs incurred in case the pump is stated up, in €
$C_{p}^{\mathrm{SU,turb}}$	Costs incurred in case the turbine is started up, in €
$C_{p}^{\mathrm{SD,pump}}$	Costs incurred in case the pump is shut down, in €
$C_{p}^{\mathrm{SD,turb}}$	Costs incurred in case the turbine is shut down, in €
$\eta_{p}^{\mathrm{pump}}, \eta_{p}^{\mathrm{turb}}$	Efficiencies in pump and turbine modes
$P_{\omega,\tau}^{\mathrm{load}}$	Total load in portfolio, in MW

Objective function

The objective function (4.10)-(4.11) is expressed as the maximization of the expected profit of the portfolio over the whole set of scenarios, while taking into consideration the risk-aversion of the VPP using the conditional value-at-risk. The formulation is expressed as a mixed-integer linear program (MILP), and includes:

(i) the fixed revenues from the energy exchanges in long-term markets;
(ii) the fixed revenues for the availability (capacity) of balancing reserves during the 24 hours of the scheduling horizon;
(iii) the revenues from the energy exchanges in the day-ahead market;
(iv) the revenues from the energy exchanges the Intraday markets;
(v) the financial penalties in case of portfolio imbalance;
(vi) the revenues from the actual provision of ancillary services (based on the signal $R^{s}_{\omega,\tau}$ requested by the system operator);
(vii) the operating costs (including the contribution to reserves) of all units;
(viii) the costs of curtailment of renewable generation (loss of green certificates).

$$\max\ \left(1-\beta\right)\Phi+\beta.CVaR \tag{4.10}$$

$$\Phi=\underbrace{\sum_{\tau\in\mathrm{T}}\lambda_{\tau}^{\mathrm{LTM}}\,\mathrm{E}_{\tau}^{\mathrm{LTM}}}_{\text{(i)}}+\underbrace{\sum_{s\in\Sigma}24\lambda_{s}^{\mathrm{res}}\mathrm{Res}_{s}^{\mathrm{tot}}}_{\text{(ii)}}+\sum_{\omega\in\Omega}\sum_{\tau\in\mathrm{T}}\pi_{\omega}\left\{\begin{array}{c}\underbrace{\lambda_{\omega,\tau}^{\mathrm{DAM}}\mathbf{E}_{\tau}^{\mathrm{DAM}}}_{\text{(iii)}}+\underbrace{\lambda_{\omega,\tau}^{\mathrm{IM}}\mathbf{E}_{\omega,\tau}^{\mathrm{IM}}}_{\text{(iv)}}+\underbrace{\lambda_{\omega,\tau}^{\mathrm{imb}}\mathbf{E}_{\omega,\tau}^{\mathrm{imb}}}_{\text{(v)}}\\ \underbrace{+\sum_{s\in\Sigma}\Delta_{\tau}R_{\omega,\tau}^{s}\lambda_{s}^{\mathrm{act}}\mathrm{Res}_{s}^{\mathrm{tot}}}_{\text{(vi)}}\\ \underbrace{-\sum_{\upsilon\in\Upsilon}\mathbf{c}_{\upsilon,\omega,\tau}^{\mathrm{op}}}_{\text{(vii)}}\underbrace{-\sum_{r\in\mathrm{RES}}\lambda_{r}^{\mathrm{GC}}\mathbf{P}_{r,\omega,\tau}^{\text{cut-off}}}_{\text{(viii)}}\end{array}\right\} \tag{4.11}$$

In order to adequately model the operation of the imbalance settlement, the optimization is carried out with a 15 minutes time resolution. This allows moreover that the power trajectories of all resources can be modeled as quarter-hourly piecewise linear functions instead of hourly functions, which enables to better represent ramping properties of units. This results in a clear distinction between power and energy, and eliminates power discontinuities (and thus energy imbalances) in the scheduling phase.

Day-ahead market

The profit made in the day-ahead market is positive if energy has been sold, and negative if energy has been purchased. In both cases, the profit value is equal to the volume exchanged multiplied by the market price. In the day-ahead market, the energy is exchanged for hourly periods, whereas the formulation is carried out with a quarter-hourly time resolution. It is therefore imposed that the hourly energy is equally distributed among its constitutive intra-hour (quarter-hourly) intervals.

To reflect the non-anticipativity constraint associated with the bidding strategy in the day-ahead market, the optimal values $\mathbf{E}_{\tau}^{\mathrm{DAM}}$ are the same for all scenarios ω ∊ Ω. These decisions variables are thus not associated with the subscript ω (scenario-independent variables) into the formulation.

Intraday market

Similarly to the day-ahead market, the energy in the Intraday market is traded for hourly periods, and we make the necessary adaptations to ensure consistency with the quarter-hourly time intervals used in the formulation.

Due to the continuous trading model used in the Intraday market, different market prices can be defined for each transaction. Here, for simplicity, the different prices corresponding to the same hours are weighted with respect to the volume of the transaction so as to obtain a single price.

$$\lambda_{\omega,\tau}^{\mathrm{IM}} = \sum_{tr \in TR} \frac{q_{tr}}{q_{tot}} \lambda_{tr} \tag{4.12}$$

where q_{tr} and q_{tot} stand respectively for the volume (in MWh) of the transaction $tr \in TR$ and the aggregated volume of all transactions associated with period τ, whereas λ_{tr} is the trading price (in €/MWh) of the transaction tr.

The intraday market liquidity is also modeled using historical information:

$$\left|\mathbf{E}_{\omega,\tau}^{\mathrm{IM}}\right| \leq \mathrm{IM}_{\mathrm{liquidity}} \tag{4.13}$$

It should be emphasized that the VPP behaves as a price taker in day-ahead and intraday markets (i.e. it is assumed that the market price is not impacted by the portfolio decisions). It results that the day-ahead and intraday prices can be modeled as exogenous variables.

Energy balance equation – Portfolio imbalance settlement

The imbalance penalty applied to the VPP for each quarter of an hour is equal to the imbalance price (in €/MWh) multiplied by the unbalanced energy of the portfolio (in MWh). This imbalance energy $\mathbf{E}_{\omega,\tau}^{\mathrm{imb}}$ is considered as positive in case of positive imbalance (i.e. the portfolio is in surplus of energy) and negative otherwise.

The VPP cannot be considered as a price-taker in the balancing market (since the sensitivity of this market is in the order of a few MWh). This issue of accounting for the market power in the imbalance settlement has been tackled in [Zugno[13]] for a wind producer aiming at maximizing its profit from both day-ahead and balancing markets (i.e. imbalance settlement) in an uncertain environment. The problem has a bi-level (nested) structure that cannot be solved directly. However, if a problem is convex and satisfies some regularity conditions (e.g. if the constraints are affine functions, then no other conditions is required), then it can be replaced by its Karush-Kuhn-Tucker (KKT) conditions [Luenberger[84]]. The lower-level problem is consequently equivalently represented by its KKT conditions, and the bi-level problem can then be reformulated as a single MILP. However, due to the complexity of the resulting formulation, a single time period needs to be considered in the work.

To bypass this issue, traditional formulations impose that the imbalance should be equal to 0. However, such a risk-averse strategy does not exploit the potential of the single pricing mechanism that rewards actors who are helping the system. Here, we attempt to find a compromise between both approaches (computationally demanding game-theoretical model that does not allow to consider multiple time steps and risk-averse policy that does not harness the economic potential of the imbalance settlement) with a solution that allows the portfolio to intentionally deviate in real-time to improve its energy management policy.

However, in two-stage stochastic programming, second stage decisions are taken once the uncertainty is fully disclosed. This implies that a decision to purposefully deviate from a balanced position is made without the risk that is normally faced in real-time by a portfolio manager that ignores the actions of other market participants. Hence, if these decisions are let uncontrolled in the formulation, this will lead to overly optimistic strategies since the portfolio will fully take advantage of the global imbalance of the grid, whereas it is much more risky in practice due to the important volatility of the grid imbalance. Furthermore, it should not be forgotten that the VPP is unlikely to have perfect information about its own current position, especially if load and renewable generation is included in the portfolio.

Hence, the imbalance position is decomposed into three components, a term $\mathbf{E}_{\omega,\tau}^{\text{imb,play}}$ which defines the imbalance that is deliberately played in the market, and two other terms $\mathbf{E}_{\omega,\tau}^{\text{imb,pos}}$ and $\mathbf{E}_{\omega,\tau}^{\text{imb,neg}}$ that are used to strongly penalized the portfolio in case of undesired imbalance.

$$\mathbf{E}_{\omega,\tau}^{\text{imb}}\lambda_{\omega,\tau}^{\text{imb}} = \mathbf{E}_{\omega,\tau}^{\text{imb,play}}\lambda_{\omega,\tau}^{\text{imb}} + \mathbf{E}_{\omega,\tau}^{\text{imb,pos}}\lambda_{\omega,\tau}^{\text{imb,pos}} - \mathbf{E}_{\omega,\tau}^{\text{imb,neg}}\lambda_{\omega,\tau}^{\text{imb,neg}} \tag{4.14}$$

where the tariffs in case of positive and negative portfolio imbalances are respectively set to $\lambda_{\omega,\tau}^{\text{imb,pos}} = 0$ € and $\lambda_{\omega,\tau}^{\text{imb,neg}} = 100$ €.

In this dissertation, the following risk strategy is applied:

$$\kappa_{1,\omega,\tau}\ \text{NRV}_{\omega} \leq \mathbf{E}_{\omega,\tau}^{\text{imb,play}} \leq \kappa_{2,\omega,\tau}\ \text{NRV}_{\omega} \tag{4.15}$$

$$\kappa_{1,\omega,\tau} = \begin{cases} 0.2 & \text{if } \text{NRV}_{\omega} \leq \text{-100 MW} \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

$$\kappa_{2,\omega,\tau} = \begin{cases} 0.2 & \text{if } \text{NRV}_{\omega} \geq 100 \text{ MW} \\ 0 & \text{otherwise} \end{cases} \tag{4.17}$$

where $\kappa_{1,\omega,\tau}$ and $\kappa_{2,\omega,\tau}$ are constant values, and allow to adopt different risk strategies. The net regulation volume (NRV) is fully explained in Section 2.6, and represents the amount of reserves activated by the system operator to restore the balance within the system. This value is used to determine the imbalance tariff, which is why the variable is included into the formulation.

Overall, the imbalance of the VPP is the difference between two terms, i.e. (1) the total energy generated within the portfolio and the energy purchased in energy markets, and (2) the total consumption and the energy sold in markets. It is worth reminding that the energy provided in the context of ancillary services is neutralized in the imbalance perimeter of a Belgian portfolio, and is therefore not taken into account[25].

$$\begin{aligned} \mathbf{E}_{\omega,\tau}^{\text{imb}} &= -\text{E}_{\tau}^{\text{LTM}} - \mathbf{E}_{\tau}^{\text{DAM}} - \mathbf{E}_{\omega,\tau}^{\text{IM}} \\ &+ \Delta_{\tau}\left(\sum_{g\in G}\mathbf{P}_{g,\omega,\tau}^{\text{real}} + \sum_{p\in PSH}\left(\mathbf{P}_{p,\omega,\tau}^{\text{turb}} - \mathbf{P}_{p,\omega,\tau}^{\text{pump}}\right) + \sum_{r\in RES}P_{r,\omega,\tau}^{\text{actual}} - P_{\omega,\tau}^{\text{load}} - \sum_{dr\in DR}\mathbf{P}_{\omega,\tau}^{\text{dr}}\right) \end{aligned} \tag{4.18}$$

Allocated reserve capacity

The procurement of balancing capacity $\text{Res}_{S}^{\text{tot}}$ occurs in mid-term (in week-ahead). The resulting capacity is then dynamically allocated (on a 15-min basis) to the different resources of the portfolio. It must be ensured that, at each time period, the portfolio is able to provide the total contracted amount of balancing reserves in both directions. Each unit can provide asymmetric contributions (e.g. participating only in upward regulation). The upward reserve can be delivered by online (conventional and storage) units, offline power plants that can start-

[25] Technically, the activation of FCR is not neutralized in the portfolio of Belgian market players.

up sufficiently quickly (non-spinning reserve) and load shedding (4.19). The downward regulation is provided by online (conventional and storage) stations, power plants able to shut-down sufficiently quickly, curtailment of renewable generation and demand response by load activation (4.20).

$$\forall s \in \Sigma^{\text{up}} : \sum_{\upsilon \in \Upsilon} \left(\mathbf{Res}^{\text{on}}_{s,\upsilon,\omega,\tau} + \mathbf{Res}^{\text{su}}_{s,\upsilon,\omega,\tau} \right) + \sum_{dr \in DR} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} = \text{Res}^{\text{tot}}_{s} \tag{4.19}$$

$$\forall s \in \Sigma^{\text{down}} : \sum_{\upsilon \in \Upsilon} \left(\mathbf{Res}^{\text{on}}_{s,\upsilon,\omega,\tau} + \mathbf{Res}^{\text{sd}}_{s,\upsilon,\omega,\tau} \right) + \sum_{r \in RES} \mathbf{Res}_{s,r,\omega,\tau} + \sum_{dr \in DR} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} = \text{Res}^{\text{tot}}_{s} \tag{4.20}$$

It is worth noting that the allocation of reserves within the portfolio is differentiated among the different scenarios of stochastic variables to ensure the optimal dispatch of available resources with regard to actual realizations of uncertainties.

The real-time activation of balancing energy for each product s is modeled as a normalized vector $R^{s}_{\omega,\tau} \in [0, 1]$ that reflects the severity of the global imbalance within the grid for each scenario $\omega \in \Omega$. The power actually delivered for service s is thus equal to this value $R^{s}_{\omega,\tau}$ multiplied by the total power allocated. Only downward reserve by shut-down and non-spinning reserves are binary stochastic variables $\in \{0, 1\}$.

Conventional power plants

The investment costs (capital expenditure) as well as the fixed operating and maintenance costs (fixed O&M)[26] are not considered in the short-term decision procedure (not relevant at this operational stage). In fact, any costs that would be incurred regardless of whether a decision is made or not (e.g. depreciation of assets) are irrelevant.

The main technical and cost-related specifications of traditional conventional power plants (CPP) technologies are presented in Table 4.3. The dynamic features are broadly divided into base, mid and peak technologies and the associated ramping rates are presented. Then, the characteristics in terms of capital and operational costs are illustrated.

Table 4.3
General characteristics of conventional technologies.

Technology	Generator type	$\%P^{max}$ /minute	CAPEX	OPEX
Nuclear	Base load	1-5	High	Low
Fossil fuel fired steam turbine	Base load	1-5	High	Low
Open Cycle Gas Turbine (OCGT)	Mid load	5-10	Mid	Mid
Combined Cycle Gas Turbine (CCGT)	Mid load	5-20	Mid	Mid
Diesel	Peak	40	Low	High
Hydro	Peak	20-50	High	Very low

With the increasing need of operational flexibility in power systems to alleviate system imbalances, it is important to accurately account for the cycling of power plants (i.e. any change in the output power due to ramping, start-up or shut-down), both in terms of technical and costs-related aspects. Indeed, one should not only ensure that the output power transitions are technically feasible (in terms of ramping rates) but also take into consideration that these power variations involve additional CO_2 emissions, while accelerating wear and tear on the unit

[26] The fixed operating and maintenance costs (fixed O&M) represent the overheads, i.e. costs that do not vary with the unit scheduling, such as wages and salaries, insurances and periodic maintenance. For units that have their own sources of energy, these costs also encompass fixed costs associated with the coal/gas extraction.

components (mainly due to excessive turbine shaft tensions and rotor temperatures), ultimately leading to forced outages [Wu13]. For instance, when a power plant is shut down, its components are subject to important temperature and pressure stresses that are accelerating the ageing of the unit, ultimately leading to forced outages. These stresses are even more exacerbated during the start-up phase. It has been shown in [Van den Bergh15,b] that these cycling costs can be in practice decreased by 40 % when they are fully included in the scheduling process. However, these costs are very difficult to quantify and depend on many factors, among which the most relevant are the type of the unit, its age and the usage pattern. The latter is furthermore measured with a long-term perspective (e.g. costs of lost opportunity due to unexpected outage, capital and maintenance costs due to accelerated ageing, decrease of global efficiency due to repeated cycling), and is thus very difficult to properly estimate in the short-term decision procedure.

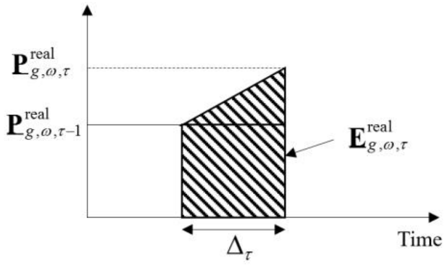


Figure 4.9 – Energy generated during cycling phase of conventional power plants.

The running costs $\underline{\mathbf{c}}^{\text{gen}}_{g,\omega,\tau}$ of conventional power plants consist in a fixed part $C^{\min}_g$ (fuel consumption and CO_2 emissions when the plant is operating at its minimum stable output power) and a variable contribution C^{marg}_g reflecting marginal costs at higher generation levels, whereas the grid fees $\underline{\mathbf{c}}^{\text{fees}}_{g,\omega,\tau}$ are proportional to the energy $\underline{\mathbf{E}}^{\text{real}}_{g,\omega,\tau}$ actually injected by the unit into the grid (4.21). As represented in Figure 4.9, this generated energy (4.22) takes into account the output power trajectory (ramping processes between consecutive time steps).

$$\underline{\mathbf{c}}^{\text{gen}}_{g,\omega,\tau} + \underline{\mathbf{c}}^{\text{fees}}_{g,\omega,\tau} = \Delta_\tau \left(C^{\min}_g \boldsymbol{\alpha}^{\text{real}}_{g,\omega,\tau} + C^{\text{marg}}_g \boldsymbol{\delta}^{\text{real}}_{g,\omega,\tau} \right) + \lambda^{\text{inj}}_g \underline{\mathbf{E}}^{\text{real}}_{g,\omega,\tau} \tag{4.21}$$

$$\underline{\mathbf{E}}^{\text{real}}_{g,\omega,\tau} = 0.5\Delta_\tau \left(\underline{\mathbf{P}}^{\text{real}}_{g,\omega,\tau} + \underline{\mathbf{P}}^{\text{real}}_{g,\omega,\tau-1} \right) \tag{4.22}$$

It should be emphasized that, due to the provision of non-spinning reserve as well as downward regulation by shut-down, the actual status of power plants may differ from their operation schedule (committed status). Consequently, in order to properly account for activation costs of such balancing services into the formulation, it is necessary to discriminate the scheduled output power $\underline{\mathbf{P}}^{\text{com}}_{g,\omega,\tau}$ and the real output power $\underline{\mathbf{P}}^{\text{real}}_{g,\omega,\tau}$.

$$\underline{\mathbf{P}}^{\text{real}}_{g,\omega,\tau} = \underline{\mathbf{P}}^{\text{com}}_{g,\omega,\tau} + \sum_{s \in \Sigma^{\text{up}}} R^s_{\omega,\tau} \mathbf{Res}_{s,g,\omega,\tau} - \sum_{s \in \Sigma^{\text{down}}} R^s_{\omega,\tau} \mathbf{Res}_{s,g,\omega,\tau} \tag{4.23}$$

$$\underline{\mathbf{P}}^{j}_{g,\omega,\tau} = \boldsymbol{\alpha}^{j}_{g,\omega,\tau} P^{\min}_g + \boldsymbol{\delta}^{j}_{g,\omega,\tau}, \; j \in [\text{com, real}] \tag{4.24}$$

Then, ramping costs are incurred at each output power variation. The contributions from scheduled power variations $\mathbf{c}^{\text{ramp}}_{g,\omega,\tau}$ and deployment of reserves $\underline{\mathbf{c}}^{\text{ramp,r}}_{g,\omega,\tau}$ are differentiated so as to accurately reflect costs of delivering reserves. Finally, costs C^{su}_g are incurred at each start-up ($\underline{\mathbf{c}}^{\text{su}}_{g,\omega,\tau}$), whether it is self-committed or by activation of non-spinning reserve (4.28).

$$\mathbf{c}^{\text{ramp}}_{g,\omega,\tau} \geq C^{\text{ramp}}_{g} \left(\boldsymbol{\delta}^{\text{com}}_{g,\omega,\tau} - \boldsymbol{\delta}^{\text{com}}_{g,\omega,\tau-1} \right) \tag{4.25}$$

$$\mathbf{c}^{\text{ramp}}_{g,\omega,\tau} \geq C^{\text{ramp}}_{g} \left(\boldsymbol{\delta}^{\text{com}}_{g,\omega,\tau-1} - \boldsymbol{\delta}^{\text{com}}_{g,\omega,\tau} \right) \tag{4.26}$$

$$\underline{\mathbf{c}}^{\text{ramp,r}}_{g,\omega,\tau} = C^{\text{ramp}}_{g} \sum_{s \in \Sigma} R^{s}_{\omega,\tau} \mathbf{Res}^{\text{on}}_{s,g,\omega,\tau} \tag{4.27}$$

$$\underline{\mathbf{c}}^{\text{su}}_{g,\omega,\tau} = C^{\text{su}}_{g} \left(\boldsymbol{\alpha}^{\text{su}}_{g,\omega,\tau} + R^{\text{off,su}}_{\omega,\tau} \boldsymbol{\alpha}^{\text{off,su}}_{g,\omega,\tau} \right) \tag{4.28}$$

In order to model power-trajectories and costs during start-up and shut-down with respect to previous operation states (e.g. higher start-up costs when the unit has been offline during a long period), one can refer to the formulation proposed by [Morales-Espana[13]].

The generation limits of each unit are defined by (4.29)-(4.31). These constraints make sure that the scheduled power respects the capacity margins allocated for the reserves so that part of the generation capacity is not doubly booked (for both reserves provision and strategic level of generation). In this way, it is ensured that the allocated reserves can be provided regardless of the actual realization of uncertainties.

$$\mathbf{P}^{\text{com}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} P^{\max}_{g} - \sum_{s \in \Sigma^{\text{up}}} \mathbf{Res}^{\text{on}}_{s,g,\omega,\tau} \tag{4.29}$$

$$\left(\boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} - \boldsymbol{\alpha}^{\text{on,sd}}_{g,\omega,\tau} \right) P^{\min}_{g} + \mathbf{Res}^{\text{on,sd}}_{s,g,\omega,\tau} + \sum_{s \in \Sigma^{\text{down}}} \mathbf{Res}^{\text{on}}_{s,g,\omega,\tau} \leq \mathbf{P}^{\text{com}}_{g,\omega,\tau} \tag{4.30}$$

$$0 \leq \boldsymbol{\delta}^{\text{real}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{real}}_{g,\omega,\tau} \left(P^{\max}_{g} - P^{\min}_{g} \right) \tag{4.31}$$

The output power transitions between consecutive time steps are constrained by the ramping abilities of the considered units, which differ between normal conditions and start-up or shut-down phases. When balancing capacity is allocated, it must be ensured that ramping capacity is not doubly booked (accounting for dynamic characteristics of the reserve products). In this way, FCR must be able to be fully activated in 0.5 minutes, aFRR in 7.5 minutes and mFRR in 15 minutes. Practically, the following constraints limit the provision of upward FCR (4.32), aFRR (4.33) and mFRR (4.34), as well as downward FCR (4.35), aFRR (4.36) and mFRR (4.37).

$$\mathbf{Res}^{\text{FCR}^{\text{up}}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} R^{\text{FCR}^{\text{up}}}_{g} \tag{4.32}$$

$$\mathbf{Res}^{\text{FCR}^{\text{up}}}_{g,\omega,\tau} + \mathbf{Res}^{\text{aFRR}^{\text{up}}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} R^{\text{aFRR}^{\text{up}}}_{g} \tag{4.33}$$

$$\mathbf{Res}^{\text{FCR}^{\text{up}}}_{g,\omega,\tau} + \mathbf{Res}^{\text{aFRR}^{\text{up}}}_{g,\omega,\tau} + \mathbf{Res}^{\text{mFRR}^{\text{up}}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} R^{\text{mFRR}^{\text{up}}}_{g} \tag{4.34}$$

$$\mathbf{Res}^{\text{FCR}^{\text{down}}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} R^{\text{FCR}^{\text{down}}}_{g} \tag{4.35}$$

$$\mathbf{Res}^{\text{FCR}^{\text{down}}}_{g,\omega,\tau} + \mathbf{Res}^{\text{aFRR}^{\text{down}}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} R^{\text{aFRR}^{\text{down}}}_{g} \tag{4.36}$$

$$\mathbf{Res}^{\text{FCR}^{\text{down}}}_{g,\omega,\tau} + \mathbf{Res}^{\text{aFRR}^{\text{down}}}_{g,\omega,\tau} + \mathbf{Res}^{\text{mFRR}^{\text{down}}}_{g,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{com}}_{g,\omega,\tau} R^{\text{mFRR}^{\text{down}}}_{g} \tag{4.37}$$

Then, the remaining ramping capacity can be used for ensuring a better strategic position in terms of energy (balancing position of the portfolio). In other words, power transitions scheduled by the operator are constrained by the total reserve capacity that can be requested during the time period considered (4.38)-(4.39).

$$\left(\mathbf{P}_{g,\omega,\tau}^{\text{com}} + \sum_{s\in\Sigma^{\text{up}}} \mathbf{Res}_{s,g,\omega,\tau}^{\text{on}}\right) - \mathbf{P}_{g,\omega,\tau-1}^{\text{com}} \leq 60\Delta_\tau R_g^{\text{up}} \boldsymbol{\alpha}_{g,\omega,\tau-1}^{\text{com}} + \left(SU_g - P_g^{\min}\right)\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{su}} \tag{4.38}$$

$$\mathbf{P}_{g,\omega,\tau-1}^{\text{com}} - \left(\mathbf{P}_{g,\omega,\tau}^{\text{com}} - \sum_{s\in\Sigma^{\text{down}}} \mathbf{Res}_{s,g,\omega,\tau}^{\text{on}}\right) \leq 60\Delta_\tau R_g^{\text{down}} \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{com}} + \left(SD_g - P_g^{\min}\right)\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{sd}} \tag{4.39}$$

The minimum up and down times are modeled following the methodology presented in [Rajan[05]] since this formulation has demonstrated higher performances than other methods [Ostrowski[12]]. The minimum up time expresses that once a unit has been started up, it may not be shut down immediately. Likewise, the minimum down time reflects the technical constraint that a unit that has been shut down must stay offline during a minimum period to ensure the safety of the equipment.

Fast-starting units (with no minimum up and down times) can deliver non-spinning mFRR (4.40) as well as downward reserve by shut-down (4.41). The formulation ensures that units are effectively brought offline when spinning reserve via shut-down is requested by the system operator (4.42), and brought online in case of emergency start-up (4.43). Finally, logical conditions (4.44)-(4.45) respectively impose that units can be allocated to start-up (shut-down) only if the unit has been committed to be offline (online).

$$\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{off,su}} P_g^{\min} \leq \mathbf{Res}_{g,\omega,\tau}^{\text{off,su}} \leq \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{off,su}} SU_g \tag{4.40}$$

$$\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{on,sd}} P_g^{\min} \leq \mathbf{Res}_{g,\omega,\tau}^{\text{on,sd}} \leq \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{on,sd}} SD_g \tag{4.41}$$

$$\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{com}} - \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{real}} = R_{\omega,\tau}^{\text{on,sd}} \boldsymbol{\alpha}_{s,g,\tau}^{\text{on,sd}} \tag{4.42}$$

$$\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{real}} - \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{com}} = R_{\omega,\tau}^{\text{off,su}} \boldsymbol{\alpha}_{s,g,\tau}^{\text{off,su}} \tag{4.43}$$

$$\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{off,su}} \leq 1 - \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{com}} \tag{4.44}$$

$$\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{on,sd}} \leq \boldsymbol{\alpha}_{g,\omega,\tau}^{\text{com}} \tag{4.45}$$

It is worth noting that these modeling equations necessitate to define only $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{off,su}}$, $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{on,sd}}$ and $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{com}}$ as binary variables, since $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{su}}$, $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{sd}}$ and $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{real}}$ are then automatically enforced to take binary values even if they are considered as continuous ones.

Then, in order to ensure the compactness (limited size) of the formulation, decision variables and constraints are differentiated between slow and fast power plants. In this way, equations (4.40)-(4.45) and related decision variables $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{off,su}}$, $\boldsymbol{\alpha}_{g,\omega,\tau}^{\text{on,sd}}$ are only modeled for fast power plants, whereas minimum up and down times are only limited to slow units.

Pumped storage hydro stations

Energy storage facilities have high dynamic characteristics and can thus quickly and cost-effectively modulate their output power to provide ancillary services or to accommodate unexpected deviations regarding the energy balance of the VPP. Representing 99% of the worldwide installed storage capacity, pumped-storage hydro (PSH) is the most widespread technology, and allows to store energy cost-effectively.

However, the operation of pumps and turbines is constrained by their operating domain, outside which the safe operation is not guaranteed (due to cavitation effects). Hence, as detailed in Chapter 5, in order to be able to provide balancing services, PSH units must continuously operate even when it is uneconomic to run the station since these are not flexible around 0 (idle mode). To prevent this problem, another more expensive solution, referred to as hydraulic short-circuit operation [Argonne[13]], was developed so as to simultaneously pump (at a fixed rate) and turbine (over the operating range), thus assimilating the station to a controllable load.

In this way, different configurations of PSH stations can be envisaged, each one differing in the number of hydraulic and electric machines. Each topology has its advantages and drawbacks, and defines the operation modes and ancillary services that can be achieved. In this way, the PSH modeling depends on the power plant configuration. For instance, with a reversible Francis pump-turbine, both main operation modes are mutually exclusive (either pump or turbine) whereas the hydraulic short-circuit operation allows to pump and turbine at the same time. It should be noted that a single holistic formulation able to model any configuration is not optimal as it results in overly complex structure (e. g. with too many equations with unnecessary binary variables).

Here, a variable-speed operation of a reversible Francis pump-turbine is considered. Indeed, recent progress in power electronics have enabled these PSH units to operate with a reliable variable-speed feature in both pump and turbine modes, consequently fostering their dispatchability [Mercier[17,a]]. In this way, the upward reserve ($s \in \Sigma^{\text{up}}$) can be provided either by increasing the generated power in turbine mode $\mathbf{Res}_{s,p,\omega,\tau}^{\text{turb}}$ or by reducing the pumping power $\mathbf{Res}_{s,p,\omega,\tau}^{\text{pump}}$. Similarly, downward reserves ($s \in \Sigma^{\text{down}}$) are supplied by lowering the power in turbine or by rising up the output power when pumping. Moreover, the reversible Francis pump-turbine has lower installation costs and allows providing balancing reserves in both pump and turbine modes but not when the station is idle (shut-down). The resulting formulation involves using two binary variables to discriminate operation modes, whereas the transitions between modes are considered sufficiently quick to be neglected (4.46).

$$\boldsymbol{\alpha}_{p,\omega,\tau}^{\text{pump}} + \boldsymbol{\alpha}_{p,\omega,\tau}^{\text{turb}} \leq 1 \quad \forall p,\omega,\tau \tag{4.46}$$

Usually, operating and maintenance costs (O&M costs) are taken as a fraction of the total capital expenditures, i.e. 1.5 to 2% in [Zhang[12]]. These costs are divided into two parts: the fixed costs that are have to be paid regardless of the pump and turbine cycles (these costs are thus not relevant in the day-ahead decision stage), and the variable ones that are subject to the utilization of the plant. In [Connolly[11]], the fixed O&M costs $C_{\text{O\&M,f}}$ are estimated as a fraction of the total installed power, whereas the variable costs $C_{\text{O\&M,v}}$ depend on the actual electricity exchanges.

$$\begin{aligned} C_{\text{O\&M,f}} &= 2.8\,€\,/\,\text{kW}/\,\text{year} \\ C_{\text{O\&M,v}} &= 3.8\,€\,/\,\text{MWh} \end{aligned} \tag{4.47}$$

Both operating costs and grid fees depend on the actual energy exchanges with the grid (and must thus account for the contribution from reserves), and are expressed as follows:

$$\begin{aligned}\mathbf{c}^{\text{op}}_{p,\omega,\tau} &= \left(C^{\text{op}}_{p}+\lambda^{\text{inj}}_{p}\right)\Delta_{\tau}\left(\mathbf{P}^{\text{turb}}_{p,\omega,\tau}-\mathbf{P}^{\text{turb}}_{p,\omega,\tau-1}+\sum_{s\in\Sigma}R^{s}_{\omega,\tau}\mathbf{Res}^{\text{turb}}_{s,p,\omega,\tau}\right)\\ &+\left(C^{\text{op}}_{p}+\lambda^{\text{cons}}_{p}\right)\Delta_{\tau}\left(\mathbf{P}^{\text{pump}}_{p,\omega,\tau}-\mathbf{P}^{\text{pump}}_{p,\omega,\tau-1}+\sum_{s\in\Sigma}R^{s}_{\omega,\tau}\mathbf{Res}^{\text{pump}}_{s,p,\omega,\tau}\right)\end{aligned} \tag{4.48}$$

Then, it is important to model the start-up and shut-down costs so as to properly consider the wear and tear of the hydraulic and electrical machines. Indeed, the operating costs employed in the literature often correspond to the historical usage of PSH power plants (i.e. turbine during peak hours and pump at night) and are not appropriate when a more fluctuating and intensive use is considered. The ramping costs, however, are less straining for the equipment (than they are for thermal power plants) and can be neglected.

$$\mathbf{c}^{\text{SU,pump}}_{p,\omega,\tau} \geq C^{\text{SU,pump}}_{p}\left(\boldsymbol{\alpha}^{\text{pump}}_{p,\omega,\tau}-\boldsymbol{\alpha}^{\text{pump}}_{p,\omega,\tau-1}\right) \tag{4.49}$$

$$\mathbf{c}^{\text{SU,turb}}_{p,\omega,\tau} \geq C^{\text{SU,turb}}_{p}\left(\boldsymbol{\alpha}^{\text{turb}}_{p,\omega,\tau}-\boldsymbol{\alpha}^{\text{turb}}_{p,\omega,\tau-1}\right) \tag{4.50}$$

$$\mathbf{c}^{\text{SD,pump}}_{p,\omega,\tau} \geq C^{\text{SD,pump}}_{p}\left(\boldsymbol{\alpha}^{\text{pump}}_{p,\omega,\tau-1}-\boldsymbol{\alpha}^{\text{pump}}_{p,\omega,\tau}\right) \tag{4.51}$$

$$\mathbf{c}^{\text{SD,turb}}_{p,\omega,\tau} \geq C^{\text{SD,turb}}_{p}\left(\boldsymbol{\alpha}^{\text{turb}}_{p,\omega,\tau-1}-\boldsymbol{\alpha}^{\text{turb}}_{p,\omega,\tau}\right) \tag{4.52}$$

$$\mathbf{c}^{\text{SU,pump}}_{p,\omega,\tau},\mathbf{c}^{\text{SU,turb}}_{p,\omega,\tau},\mathbf{c}^{\text{SD,pump}}_{p,\omega,\tau},\mathbf{c}^{\text{SD,turb}}_{p,\omega,\tau} \geq 0 \tag{4.53}$$

The schedule of hydraulic machines is constrained by output power limitations, defining the safe operating ranges of both pump and turbine modes [Ardizzon[14]]. Hence, operating range constraints are needed to ensure that the PSH unit can provide reserve in both pump and turbine modes, while ensuring that the energy capacity of the station is not doubly booked.

$$\mathbf{P}^{\text{turb}}_{p,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{turb}}_{p,\omega,\tau}P^{\text{turb,max}}_{p}-\sum_{s\in\Sigma^{\text{up}}}\mathbf{Res}^{\text{turb}}_{s,p,\omega,\tau} \tag{4.54}$$

$$\mathbf{P}^{\text{turb}}_{p,\omega,\tau}-\sum_{s\in\Sigma^{\text{down}}}\mathbf{Res}^{\text{turb}}_{s,p,\omega,\tau} \geq \boldsymbol{\alpha}^{\text{turb}}_{p,\omega,\tau}P^{\text{turb,min}}_{p} \tag{4.55}$$

$$\mathbf{P}^{\text{pump}}_{p,\omega,\tau} \leq \boldsymbol{\alpha}^{\text{pump}}_{p,\omega,\tau}P^{\text{pump,max}}_{p}-\sum_{s\in\Sigma^{\text{down}}}\mathbf{Res}^{\text{pump}}_{s,p,\omega,\tau} \tag{4.56}$$

$$\mathbf{P}^{\text{pump}}_{p,\omega,\tau}-\sum_{s\in\Sigma^{\text{up}}}\mathbf{Res}^{\text{pump}}_{s,p,\omega,\tau} \geq \boldsymbol{\alpha}^{\text{pump}}_{p,\omega,\tau}P^{\text{pump,min}}_{p} \tag{4.57}$$

PSH have very high ramping rates (able to switch from full pumping capacity to the maximum generation power in less than 7.5 minutes). Hence, given the 15-min time resolution considered, it is guaranteed that the ramping capacity is not doubly allocated. Nonetheless, due to the high dynamic requirements related to balancing reserves (e.g. FCR fully delivered in 30 seconds), it must be ensured that the allocated capacity respects these ramping constraints. To that end, analogous equations to (4.32)-(4.37) have to be considered.

When allocating reserves to PSH stations, one should ensure that the limits on the energy content have to be respected at each time step, i.e. the unit must be able to provide the requested energy in the worst-case scenario (full deployment of all scheduled reserves in one direction). Disregarding these constraints may mislead the VPP operator into believing that resources are cost-effectively scheduled, while it may actually lead to the real-time unavailability of the reserved balancing capacity, resulting in costly financial penalties.

$$\underline{\mathbf{SOC}}_{p,\omega,\tau-1} + \Delta_\tau \left(\eta_p^{\text{pump}} \mathbf{P}_{p,\omega,\tau}^{\text{pump}} - \frac{\mathbf{P}_{p,\omega,\tau}^{\text{turb}}}{\eta_p^{\text{turb}}} \right) + \Delta_\tau \sum_{s\in\Sigma^{\text{down}}} \left(\mathbf{Res}_{s,p,\omega,\tau}^{\text{pump}} + \mathbf{Res}_{s,p,\omega,\tau}^{\text{turb}} \right) \le \text{SOC}_p^{\max} \quad (4.58)$$

$$\underline{\mathbf{SOC}}_{p,\omega,\tau-1} + \Delta_\tau \left(\eta_p^{\text{pump}} \mathbf{P}_{p,\omega,\tau}^{\text{pump}} - \frac{\mathbf{P}_{p,\omega,\tau}^{\text{turb}}}{\eta_p^{\text{turb}}} \right) - \Delta_\tau \sum_{s\in\Sigma^{\text{up}}} \left(\mathbf{Res}_{s,p,\omega,\tau}^{\text{pump}} + \mathbf{Res}_{s,p,\omega,\tau}^{\text{turb}} \right) \ge \text{SOC}_p^{\min} \quad (4.59)$$

The state-of-charge (SOC) accounts for energy losses originating from pump and turbine inefficiencies, and integrates the actual contribution of reserves.

$$\begin{aligned}\underline{\mathbf{SOC}}_{p,\omega,\tau} = \underline{\mathbf{SOC}}_{p,\omega,\tau-1} &+ \Delta_\tau \left(\eta_p^{\text{pump}} \mathbf{P}_{p,\omega,\tau}^{\text{pump}} - \frac{\mathbf{P}_{p,\omega,\tau}^{\text{turb}}}{\eta_p^{\text{turb}}} \right) \\ &+ \Delta_\tau \eta_p^{\text{pump}} \left(\sum_{s\in\Sigma^{\text{up}}} R_{s,\omega,\tau}^{\text{SO}} \mathbf{Res}_{s,p,\tau}^{\text{pump}} - \sum_{s\in\Sigma^{\text{down}}} R_{s,\omega,\tau}^{\text{SO}} \mathbf{Res}_{s,p,\tau}^{\text{pump}} \right) \\ &- \frac{\Delta_\tau}{\eta_p^{\text{turb}}} \left(\sum_{s\in\Sigma^{\text{up}}} R_{s,\omega,\tau}^{\text{SO}} \mathbf{Res}_{s,p,\tau}^{\text{turb}} - \sum_{s\in\Sigma^{\text{down}}} R_{s,\omega,\tau}^{\text{SO}} \mathbf{Res}_{s,p,\tau}^{\text{turb}} \right)\end{aligned} \quad (4.60)$$

In order to account for the economic value of the energy stored at the end of the day, the final state-of-charge $\underline{\mathbf{SOC}}_{p,\omega}^{\text{final}}$ is imposed to a value optimally determined through a medium-term (e.g. week-ahead) analysis [Deane[13]].

$$\underline{\mathbf{SOC}}_{p,\omega}^{\text{final}} \ge \text{SOC}_p^{\text{target}} \quad (4.61)$$

It should be noted that the formulation associated with other storage technologies (e.g. battery, compressed-air energy storage, etc.) can be simplified since no binary variables are then necessary to model the discontinuous operating range, and only one continuous variable can be used to define the output power.

Demand response

Some end-users can shift part of their consumption (e.g. electric heating systems, refrigeration loads and electric vehicles) to periods that would be more economic. However, they cannot participate on their own to advanced load shifting programs. An aggregator therefore acts on behalf of its customers by collecting the load data and by submitting aggregated demand response bids to power markets. In such a case, the uncertainty about the load prediction boils down to the base (non-responsive) part of the total consumption.

Similarly to cycling capabilities of conventional power plants, this DS resource can be exploited at different time scales, either for arbitrage opportunities in energy markets (load shifting during peak hours) or for proving real-time flexibility [Karangelos[12]]. In the latter case, the flexibility can be either used to hedge against forecast errors or unexpected events within the portfolio or can be valued towards the system operator by contributing to the grid stability (frequency control). However, due to the nature of DR flexibility such as static load profile that cannot be altered (but that can be shifted in time), few loads meet the technical requirements for participating to ancillary services.

In order to take the intertemporal load characteristics into consideration without modeling each individual system, a technology-clustered formulation of the demand response strategy is considered. Here, thermostatically controlled loads, namely industrial fridges and residential heat pumps, are the two DR clusters included in the studied portfolio. In both cases, the availability of these controllable loads is limited by comfort constraints imposed by the end-user (temperatures within fridges and buildings must stay within defined ranges).

The thermal behavior of utilities (evolution of temperature over time) is here simplified by a set of constraints (4.62)-(4.63) modeling a sliding time window T^{window} within which bounds on the consumption level $E^{\text{dr,lim,min}}$ and $E^{\text{dr,lim,max}}$ are imposed (so as to avoid excessive temperature deviations). These constraints are moreover modeled to guarantee that DR-based reserve is allocated only if its deployment does not cause violation of comfort levels.
$\forall \tau = 1 \ldots N_T - T^{\text{window}}$

$$\sum_{q=\tau}^{\tau+T^{\text{window}}} \Delta_\tau \left(\mathbf{P}^{\text{dr}}_{\omega,q} + \sum_{s\in\Sigma^{\text{down}}} \mathbf{Res}^{\text{dr}}_{s,\omega,q} \right) \le E^{\text{dr,lim,max}} \tag{4.62}$$

$$\sum_{q=\tau}^{\tau+T^{\text{window}}} \Delta_\tau \left(\mathbf{P}^{\text{dr}}_{\omega,q} - \sum_{s\in\Sigma^{\text{up}}} \mathbf{Res}^{\text{dr}}_{s,\omega,q} \right) \ge E^{\text{dr,lim,min}} \tag{4.63}$$

It is then ensured that all load $E^{\text{dr,max}}$ is consumed at the end of the scheduling horizon (load recovery effect).

$$\sum_{\tau\in\mathrm{T}} \Delta_\tau \left(\mathbf{P}^{\text{dr}}_{\tau,\omega} + \sum_{s\in\Sigma^{\text{up}}} R^{s}_{\omega,\tau} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} - \sum_{s\in\Sigma^{\text{down}}} R^{s}_{\omega,\tau} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} \right) \ge E^{\text{dr,max}} \tag{4.64}$$

Then, the ancillary services that can be actually provided by DR resources are constrained by the dispatch of the flexible load. In this way, upward reserves are offered through load curtailment (and are thus limited to the load level $P^{\text{dr,max}}_{\tau}$) whereas the downward reserve consists in increasing the consumption.

$$\sum_{s\in\Sigma^{\text{up}}} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} \le \mathbf{P}^{\text{dr}}_{\omega,\tau} \tag{4.65}$$

$$\sum_{s\in\Sigma^{\text{down}}} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} \le P^{\text{dr,max}}_{\tau} - \mathbf{P}^{\text{dr}}_{\omega,\tau} \tag{4.66}$$

It is assumed that the aggregator has the ability to precisely adjust the state of all controllable loads (perfect control over DR resources).

Renewable generation

When renewable generation is authorized to provide downward reserves (typically for installations of sufficient size), the offered reserve should be limited to the output power $P^{\text{reliable}}_{r,\tau}$ available with sufficient reliability (4.67). This level is here fixed to the lowest value of renewable generation among all predictive scenarios (worst-case scenario) considered in the stochastic optimization procedure.

$$\mathbf{Res}_{s,r,\tau} \le P^{\text{reliable}}_{r,\tau} \tag{4.67}$$

Then, the real-time curtailment of renewable generation decided by the portfolio operator for strategic reasons is limited to the actual generation level $P^{\text{actual}}_{r,\omega,\tau}$ decreased by the energy allocated for balancing reserves.

$$0 \leq \mathbf{P}^{\text{cut-off}}_{r,\omega,\tau} \leq P^{\text{actual}}_{r,\omega,\tau} - \mathbf{Res}_{s,r,\tau} \tag{4.68}$$

This RES curtailment induces a loss of green certificates (or any other financial incentive) due to the non-generated energy, which has to be taken into account in the objective function.

Constraints used to model the risk (CVAR)

Including the conditional value-at-risk into the formulation necessitates to integrate the following additional constraints:

$$VaR_{\alpha} - \left\{ \begin{array}{c} \sum_{\tau \in \mathrm{T}} \lambda^{\text{LTM}}_{\tau} \, \mathrm{E}^{\text{LTM}}_{\tau} \\ + \sum_{s \in \Sigma} 24 \lambda^{\text{res}}_{s} \text{Res}^{\text{tot}}_{s} \end{array} \right\} - \sum_{\tau \in \mathrm{T}} \left\{ \begin{array}{c} \lambda^{\text{DAM}}_{\omega,\tau} \mathbf{E}^{\text{DAM}}_{\tau} + \lambda^{\text{IM}}_{\omega,\tau} \mathbf{E}^{\text{IM}}_{\omega,\tau} + \lambda^{\text{imb}}_{\omega,\tau} \mathbf{E}^{\text{imb}}_{\omega,\tau} \\ + \sum_{s \in \Sigma} \Delta_{\tau} R^{s}_{\omega,\tau} \lambda^{\text{act}}_{s} \text{Res}^{\text{tot}}_{s} \\ - \sum_{\upsilon \in \Upsilon} \mathbf{c}^{\text{op}}_{\upsilon,\omega,\tau} - \sum_{r \in \text{RES}} \lambda^{\text{GC}}_{r} \mathrm{P}^{\text{cut-off}}_{r,\omega,\tau} \end{array} \right\} \leq z_{\omega}, \; \forall \omega \in \Omega \tag{4.69}$$

$$z_{\omega} \geq 0, \; \forall \omega \in \Omega \tag{4.70}$$

where VaR_{α} and z_{ω} are auxiliary decision variables.

4.5 Performance of MILP formulation

Solving mixed-integer linear programming (MILP) problems by rounding the solution from a traditional linear programming solver (e.g. simplex method) is usually not optimal, not to mention that there is no guarantee that the solution will be feasible, especially for large scale formulations. The integer constraints have therefore to be explicitly taken into account in the problem resolution.

Due to the combined improvements of algorithmic techniques to solve MILP problems as well as of the computer capabilities, solving such MILP has become considerably faster during the last two decades [Koch[11]].

However, these problems remain intrinsically very difficult to solve, especially for large-scale formulations (such as stochastic programming), and it is up to the problem designer to implement an efficient formulation. Indeed, it is interesting to keep in mind that an infinite number of formulations can be obtained based on the way the physical constraints of the problem are translated into mathematical equations.

The objective is therefore to design a computationally efficient (quickly solved) formulation that accurately model the original problem so that the resulting solution is practically feasible and close to the actual optimum.

In this way, two different formulations of the same MILP problem are illustrated in Figure 4.10. The ideal formulation (LP1) encompasses all feasible points of the original problem is such a way that each vertex is an integer solution. Hence, it allows solving the MILP (non-convex) problem as a LP (convex-problem), and the solution of LP1 will coincide with the actual optimal solution. This ideal formulation is commonly referred to as *convex hull*, and is defined as the smallest convex region encompassing all feasible integer solutions [Wolsey[98]]. The formulation LP2 is called a relaxed LP version of the problem.

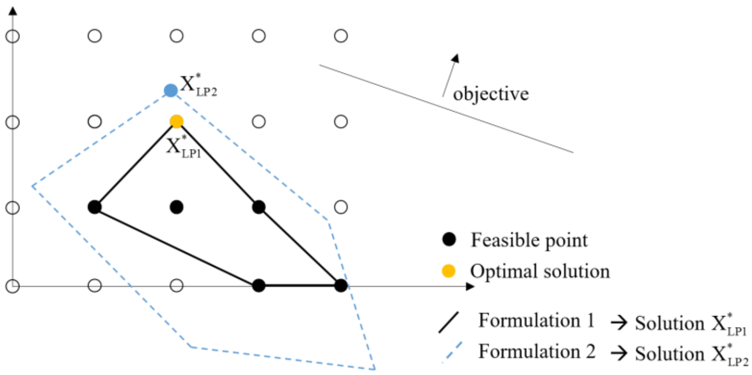


Figure 4.10 – MILP formulations [Morales-Espana[14]].

In practice, obtaining the convex hull is a complex task that cannot be realistically achieved for large-scale problems. But, it is possible to tighten the region of the relaxed LP problem in order to reach computational improvements [Nemhauser[99]]. Overall, given two different formulations of the same problem, the one that lies the nearest from the convex hull (i.e. the tighter formulation) should be favored since it provides stronger lower bounds (in case of a minimization problem) and its solution is nearer to the optimal (desired) integer solution.

Apart from the **tightness**, the performance of an MILP formulation depends also on its **compactness** (i.e. quantity of information to be considered when solving the problem). The compactness of a MILP formulation relates to its size, not only in terms of number of constraints but also regarding the number of integer decision variables and non-zero elements in the formulation [Bixby[00]].

Mixed-integer linear programming problems are usually solved using **branch-and-cut algorithms** (i.e. branch-and-bound algorithm combined with cutting planes[27] to further tighten LP relaxations). The principle of this methodology for a minimization problem is described.

First, the relaxed LP problem (original MILP formulation in which the integer constraints are neglected) is solved. If the solution X_{LP} contains a non-integer value for a variable that is supposed to be integer (i.e. fractional solution), a cutting plane algorithm attempts to tighten the LP relaxation (find additional linear constraints which are violated by the current solution X_{LP} but not by the feasible integer points).

Then, the branch-and-bound algorithm is started. The basic idea of this algorithm is to rely on two subroutines that compute respectively a lower and an upper bound of the optimal solution. The upper bound is an integer solution pertaining to the feasible set. The lower bound is a fractional solution (originating from relaxation). The objective is to minimize the difference between upper and lower bounds (optimality tolerance) by partitioning the search space into convex sets, and find lower/upper bounds for each set. The procedure necessitates thus to solve a sequence of LP relaxations, and is stopped when the difference is small enough.

[27] Cutting planes methods aims at refining the feasible set of a problem by adding linear inequalities, referred to as cuts, in order to tighten the formulation.

Overall, the tightness of the MILP formulation is an image of the search space that needs to be explored before reaching the optimal integer solution (the length of the search can be reduced by relying on relaxations close to the optimal integer solution). The compactness is an image of the searching speed to reach the optimal integer solution (the computation time can be reduced the appropriately decreasing size of the formulation).

The recent progress of MILP solver comes from different (tightening and compacting) strategies, but the inclusion of cutting planes has been acknowledged as the most effective strategy [Bixby[07], Rothberg[03]], by tightening the formulation around the integer feasible solution. Current research for improving the solving procedure of MILP problem is thus focused on how to obtain more tight formulations rather than more compact ones [Morales-Espana[14]].

On the one hand, a MILP formulation is tightened by adding supplementary constraints, which increases the problem size. In this way, although this tightened formulation reduces the search space, solvers may take more time since they are required to repeatedly solve larger LPs in the branch-and-bound procedure. On the other hand, compact formulations usually provide weak lower bounds (distant from the optimal integer solution).

In conclusion, obtaining a good trade-off between tightness and compactness of a MILP formulation is a complex task since the obvious mathematical equations to model the original problem often lead to very weak (not tight) or very large formulations.

In parallel, problems such as scenarios-based stochastic programs have a particular geometrical structure that can be exploited by **decomposition algorithms**, which consist in iterative approaches that replace the original problem into smaller sub-problems so as to accelerate the convergence speed [Kazempour[12]]. Devising and applying these techniques to solve our MILP problem is outside the scope of this work, and the resulting MILP problem of the day-ahead scheduling of a portfolio manager is solved using the traditional branch-and-cut algorithm of Matlab.

4.6 Case study

The formulation is applied to an electricity retailer having 2 medium-sized conventional generation units, a 130 MW slow power plant and a fast unit with a 80 MW capacity. The portfolio is moreover constituted of one PSH station with an output power in both pump and turbine modes of 24 MW (with an energy/power ratio of 5). The VPP has also wind turbines for a total installed power of 20 MW and supply both industrial and residential clients (peak power of 30 MW). Some are equipped with rooftop photovoltaic (PV) units for a total installed power of 5 MW. The electrical energy generated from PV installation distributed among low-voltage networks is treated as a negative load and cannot be curtailed. Finally, the VPP is responsible for optimally operating DR resources from heat pumps (10 MW) and industrial fridges (15 MW).

The computational size of the problem is firstly analyzed. The number of decision variables and constraints are respectively presented in Table 4.5 and Table 4.6.

Table 4.5
Number of decisions variables associated with the different contributions of the formulation.

Contribution	# continuous decision variables	# binary decision variables
Day-ahead market	1*N_T*N_Ω	0
Intraday market	1*N_T*N_Ω	0
Imbalance settlement	3*N_T*N_Ω	0
Slow conventional generation	11*(number of units)*N_T*N_Ω	2*(number of units)*N_T*N_Ω
Fast conventional generation	13*(number of units)*N_T*N_Ω	4*(number of units)*N_T*N_Ω
Pumped storage hydro units	18*(number of units)*N_T*N_Ω	2*(number of units)*N_T*N_Ω
Demand response	3*(number of categories)*N_T*N_Ω	0
Renewable generation	2*(number of technologies)*N_T*N_Ω	0
Conditional value-at-risk	1 + N_Ω	0

Table 4.6
Number of constraints associated with the different contributions of the formulation.

Contribution	# equality constraints for each time step in each scenario	# inequality constraints for each time step in each scenario
Day-ahead market	5*N_T*N_Ω	0
Intraday market	4*N_T*N_Ω	0
Imbalance settlement	1*N_T*N_Ω	0
Reserve allocation	6*N_T*N_Ω	0
Slow conventional generation	0	16*(number of units) *N_T*N_Ω
Fast conventional generation	0	22*(number of units) *N_T*N_Ω
Pumped storage hydro units	0	18*(number of units) *N_T*N_Ω
Demand response	0	(number of categories)*((N_T-T^{window}) *N_Ω + N_Ω + 2*N_Ω*N_T)
Renewable generation	0	1*(number of categories) *N_T*N_Ω
Conditional value-at-risk	0	N_Ω

4.6.1 Stochastic model analysis

The scenarios modeling the uncertainty of the second set of variables (grid imbalance and real-time activation of reserves) are generated as follows. The dimensionality of each individual variable is first decreased with PCA. The reduced-sized variables (i.e. after applying PCA) are determined such as preserving at least 85 % of the information of the original dataset.

The copula model is then used for constructing the scenarios. It requires less than 30 seconds to create the model and generate the related scenarios. Their validity and performance characteristics can then be estimated by comparing their statistical properties with a test set composed of the actual realizations of July 2016.

First, the statistical properties of variables are evaluated individually. The statistics deal with the mean, standard deviation and autocorrelation function (ACF) of variables. Then, the linear correlation between pairs of variables is computed using the Pearson correlation coefficient. The Kendall coefficient is also studied since it is able to measure nonlinear dependencies. The range of both coefficients is the [-1, 1] interval. The model outcomes are compared with the training and test set and all analytical results are summarized in Table 4.7.

Table 4.7
Statistical differences of generated scenarios with training set and actual realizations.

	Mean deviation with training set	Mean deviation with test set
Mean	7 %	19 %
Standard deviation	4 %	22 %
ACF	0.11	0.11
Pearson coefficient	0.08	0.11
Kendall coefficient	0.07	0.09

The relative difference (in %) between the mean and standard deviation of the generated vectors and real observations is taken as a precision measurement. The model properties are close to the training data with low statistical deviations but present decreased performances when compared to the test set. Such differences can be largely explained by the transition state that characterizes the current power market and grid environment.

The accuracy of the time correlation and cross-variable dependence structure is estimated by the absolute difference between coefficients computed with the model and the actual realizations. It is interesting to notice that the performance of the model is quite similar when respectively compared to the training and test sets. This observation tends to show that the evolution of market conditions is not affecting the whole dependency structure but is only impacting the marginal distribution of variables.

4.6.2 Optimization specifications

The MILP scheduling model is implemented and solved in Matlab, and the simulations have been performed on an Intel® Core™ i7-3770 CPU @ 3.4 GHz with 16 Go RAM. The stochastic optimization has been run with N_{Ω} = 6 scenarios during a typical day of July, which corresponds to the maximum number that can be included, considering the space memory of the computer. In theory, this number of scenarios should result from a trade-off between accuracy of results (obtain a solution resilient to uncertainties originating from forecasts errors) and the computational burden. This time has indeed to be kept sufficiently low (typically less than 1 hour) to take advantage of the most recent forecasts within the decision procedure.

The objective is to optimize the day-ahead scheduling, which is characterized by decisions that are made on a 15 minutes temporal resolution (96 daily decisions).

Discussion on optimal valorization

One of the main purpose of this chapter is to evaluate the added value of properly accounting for the uncertainty associated with the real-time activation of balancing services. However, the procurement of balancing capacity is currently carried out in mid-term (week-ahead for FCR and aFFR and month-ahead for mFRR), and the amount of reserve capacity is therefore considered as a constraint in the day-ahead scheduling. Without relying on a mid-term decision tool (as the one presented in chapter 6), it is therefore impossible to properly consider the benefit of our approach (cost-optimal allocation to the different market opportunities). Consequently, the formulation in the case study is slightly biased to consider that the procurement of the total balancing capacity $\mathbf{Res}_s^{\mathrm{tot}}$ for each service $s \in \Sigma$ occurs in day-ahead (with fixed offering prices to keep the problem linear). It results that the equations (4.19)-(4.20) are therefore expressed as follows:

$$\forall s \in \Sigma^{\text{up}} : \sum_{\upsilon \in \Upsilon} \left(\mathbf{Res}^{\text{on}}_{s,\upsilon,\omega,\tau} + \mathbf{Res}^{\text{su}}_{s,\upsilon,\omega,\tau} \right) + \sum_{dr \in DR} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} = \mathbf{Res}^{\text{tot}}_{s} \tag{4.71}$$

$$\forall s \in \Sigma^{\text{down}} : \sum_{\upsilon \in \Upsilon} \left(\mathbf{Res}^{\text{on}}_{s,\upsilon,\omega,\tau} + \mathbf{Res}^{\text{sd}}_{s,\upsilon,\omega,\tau} \right) + \sum_{r \in RES} \mathbf{Res}_{s,r,\omega,\tau} + \sum_{dr \in DR} \mathbf{Res}^{\text{dr}}_{s,\omega,\tau} = \mathbf{Res}^{\text{tot}}_{s} \tag{4.72}$$

First, the global risk-strategy is established by fixing the value of the β parameter in the objective function (4.10) that characterizes the trade-off between the expected value of the profit distribution and the conditional value-at-risk (CVaR) quantifying the generated revenue of the less-profitable scenarios. Figure 4.11 shows the dependence between the mean and CVaR of the profit distribution for different risk strategies.

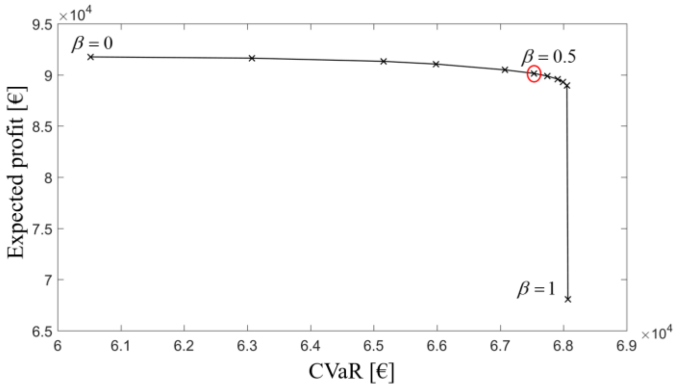


Figure 4.11 – Dependence between expected profit and CVaR with regard to risk-aversion.

Consistently with regard to previous works on risk-constrained decision-making, the expected profit increases when risk aversion decreases (i.e. lower β). Traditionally, a good trade-off is obtained on the elbow of the curve. A slightly risk-averse strategy is here considered with a β-value of 0.5.

4.6.3 Comparison with a conservative formulation

The outcomes of the proposed approach are compared with a conservative strategy that does not consider the uncertainty of real-time activation of balancing reserves.

This approach is fully conservative in the sense that decisions are taken considering that the reserved capacity can be continuously requested in either upward or downward direction during the whole day. It should be noted that the calculation time of both formulations is very similar since the additional complexity of the proposed approach (to efficiently allocate balancing reserves to available resources while adequately integrating related revenues and operating costs) is quite small in comparison with constraints associated with the technical operation of constitutive assets.

The reliability and performance of both methods is then analyzed via an out-of-sample validation of their respective day-ahead scheduling. This validation procedure consists in confronting the day-ahead decisions with respect to a new set of scenarios representing different possible realizations (daily trajectories) of uncertainties. For each realization, an economic dispatch of the VPP has to be performed. This model stems from the formulation described in Section 4.4, in which day-ahead decisions variables, i.e. energy exchanged in the day-ahead market, amount of power allocated to the different reserve products and the commitment of slow power plants, are fixed to the values determined at the end of the scheduling problem. Practically, the day-ahead schedules are thus tested within a Monte Carlo environment, in which new simulations are carried out until convergence is achieved, i.e. when the coefficient of variation C_v is lower than 1 %.

$$C_v = \frac{\mathrm{var}(\Phi_i)}{\mathrm{mean}(\Phi_i)} \tag{4.73}$$

where Φ_i is the vector of profits computed during the last 10 Monte Carlo simulations.

The results of the out-of-sample analysis are represented in Figure 4.12, and illustrates the two main benefits of the proposed formulation compared to traditional approaches. Firstly, the impact of accounting for realistic scenarios of real-time activation of reserves is highlighted. Secondly, the economic advantages of considering a time-varying cost-optimal mix of flexibility providers is analyzed. To that end, a day-ahead (fixed) allocation of reserves is compared with the intraday dispatch of resources. The profit distribution over all scenarios is here represented by its expected value and standard deviation (whisker plot).

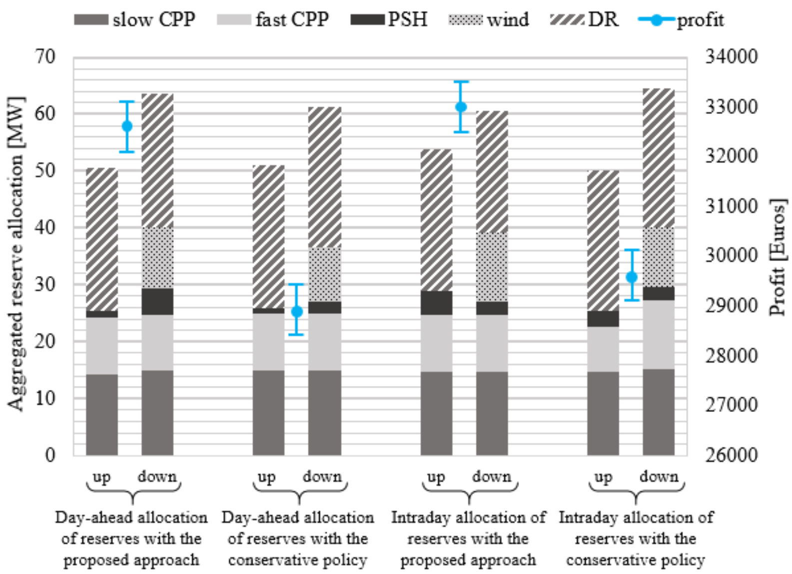


Figure 4.12 – Added value of the proposed formulation.

The proposed formulation allows to procure balancing capacity more cost-efficiently than the conservative approach, which is highlighted by the higher economic value of the portfolio. Then, allocating reserves in Intraday on a quarter-hourly basis, based on previous decisions and actual values of uncertain parameters, enables to increase the participating of the PSH station and RES to reserves by respectively 15 % and 20 %. Overall, it results in an increased expected profit of around 11 %, which illustrates the need to avoid static (daily) allocation of balancing resources.

4.6.4 Analysis of the proposed formulation

First, it is interesting to notice that the inclusion of a PSH station within the portfolio dramatically increases the simulation time. In this way, the global optimization is solved in around 20 minutes, whereas it takes only 30 seconds when no PSH unit is considered. Such a difference mainly originates from the discontinuity in the safe operating ranges of these stations.

However, neglecting these constraints leads to unfeasible real-time operating conditions, and, more importantly, to an over-optimistic estimation of available flexibility margins to provide regulation services. The real-time provision of reserves is therefore not guaranteed, which may infer portfolio imbalances, ultimately lowering the profit generated by PSH units.

Then, the economic value associated with the provision of flexibility services by energy-constrained units is analyzed within a risk-neutral approach. To that end, simulations are divided into five dispatch strategies, which differ by the way resources are allocated. In variant #1, flexible resources participate only in energy arbitrage, and are thus exclusively used to accommodate temporary surpluses or deficits in energy. The individual benefits of regulating services delivered by PSH units, output power modulation of RES and DR resources are respectively estimated in variants #2, #3 and #4. Finally, in variant #5, all resources (including conventional generation) can provide both energy arbitrage and regulation services.

The outcomes are summarized in Figure 4.13. Practically, the variants are compared regarding the total balancing capacity allocated by the portfolio as well as the generated profit.

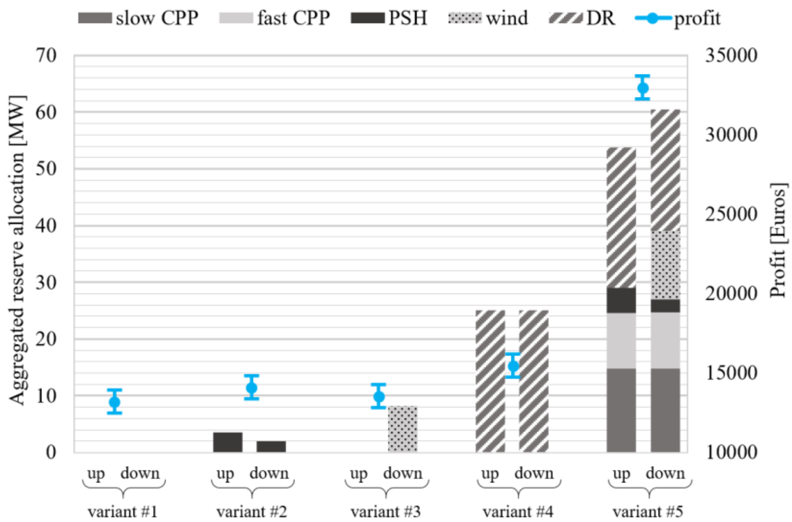


Figure 4.13 – Cross-comparison of different methods for allocating resources.

Generally, all available flexibility is not allocated to ancillary services. This can be explained by the necessity to avoid portfolio imbalances, which requires significant ramping abilities to compensate quick fluctuations of the residual demand (non-responsive load minus the aggregated distributed renewable generation) associated with prediction errors, reaching up to 10 MW/15 minutes. Indeed, since exchanges on energy markets consist of hourly electricity blocks whereas portfolio imbalances are penalized on a 15-minutes basis, these energy variations have to be handled with available assets.

All the three energy-constrained resources prove to be cost-efficient flexibility providers. However, their joint contribution is greater than the sum of their individual abilities. Here, the portfolio effect allows to increase by 20 % the contribution of storage in balancing services (from 3.58 MW and 2 MW in up- and downward regulation to respectively 4.33 MW and 2.34 MW). Interestingly, the contribution of DR-based regulation services is almost unaffected when other flexibility providers are available, which highlights their strong economic value, regardless of the portfolio in which they are included.

Moreover, this aggregation of technologies results in a more efficient use of assets. In this way, new resources able to provide downward regulation (e.g. PSH stations in pumping mode) enable to avoid shutting-down slow conventional units when electricity prices are lower than their marginal costs. It is also observed that fast-starting power plants can absorb occasional large prediction errors more cost-effectively than other power plants due to their ability to provide downward regulation by shut-down (since such mFRR reserves are seldom needed in practice).

4.7 Conclusions and perspectives

In this chapter, a new formulation for the joint day-ahead bidding strategy in energy and operating reserve markets is presented. The objective is to properly consider a dynamic allocation of flexible resources so as to determine the actual cost-optimal mix of flexibility providers. The results demonstrate that, in presence of multiple energy-constrained resources, the dynamic allocation of reserves foster the participation to ancillary services, which results in higher economic value of the global portfolio. Moreover, it is observed that neglecting or misrepresenting the real-time activation of operating reserves can lead to overly conservative solutions that do not fully exploit the potential of available resources.

4.8 Chapter publications

This chapter aims to lead to the following publication:

- J.-F. Toubeau, Z. De Grève, and F. Vallée, "Fostering Provision of Operating Reserves in a Short-Term Multimarket Optimization Framework," working paper.

CHAPTER 5

MODELING NONLINEAR EFFECTS IN THE DAY-AHEAD SCHEDULING OF PUMPED STORAGE HYDRO STATIONS

5.1 Introduction

As already mentioned, the increased contribution of uncertain and fluctuating renewable generation, originating mainly from wind and photovoltaic (PV) sources, is substantially impacting the planning and operation of power systems. In order to efficiently hedge against these uncertainties, there is a growing need of flexibility that can be provided by pumped storage hydropower (PSH) plants due to their ability to quickly and cost-effectively respond to mismatches between generation and consumption.

A pumped-storage plant is an energy storage device with water being recycled between upper and lower reservoirs. In a vertically integrated system, such units are used to reduce the fuel costs of the system by letting the pump-storage plants serve the peak load and then pump the water back into the upper reservoir at light-load periods. In the current competitive framework of the electricity sector, these units are exploited with an objective of return on investment (profit maximization).

These stations can indeed store large amounts of energy with low operating costs. Then, recent progress in power electronics have enabled PSH units to operate with a reliable variable-speed feature in both pump and turbine modes, consequently fostering their dispatchability (ability to quickly adjust their output power). This flexibility is highly valuable, not only to improve the economic efficiency of existing assets such as wind farms or thermal power plants [Abreu[12], Plazas[05], Sánchez de la Nieta[13]], but also to provide ancillary services (i.e. power reserves used to ensure the grid stability) such as frequency control or congestion management. This favorable environment leads to the development of new technologies such as underground PSH units, in which the lower reservoir is located into the ground, for instance when end-of-life mines or quarries are exploited as natural basins for saving civil engineering expenses. These stations have indeed very limited impacts on landscape, vegetation and wildlife, and are not limited by topography so that more sites can be exploited [Alvarado[15]]. However, their operation is governed by two main nonlinear effects that cannot be easily modeled with traditional analytical models [Pérez-Diaz[15]].

Firstly, groundwater exchanges between the reservoirs and their hydrogeological porous surroundings may occur. This situation typically arises for underground PSH when the waterproofing work is not feasible or uneconomical [Pujades[17]]. It is worth noting that these groundwater exchanges vary endogenously with water volumes within reservoirs and differ thus from exogenous water inflows (originating from rainfall, snowmelt, natural evaporation, etc.) that can be independently forecasted. Modeling these nonlinear groundwater dependencies is not yet tackled in the current literature, and represents a challenging task, especially since a small simulation time step (1 minute maximum) is required to properly model the high dynamics of these nonlinear effects.

Secondly, these small to medium-sized PSH units are generally subject to important variations of the net hydraulic head (i.e. height difference between water levels of the reservoirs). These variations are referred to as the head effects [Ponrajah[98]], and are typically quantified through laboratory measurements on a scale model of the studied hydraulic machines [Pannatier[10]]. This characterization of head effects is important since the head value defines both the operating range of the station as well as the efficiency of both pump and turbine processes.

Thirdly, in underground PSH stations, the geometry of the reservoirs depend on the topological conditions and can potentially take any complex form, thereby leading to a nonlinear relationship between water volumes within reservoirs and the hydraulic head.

In this way, the safe operating limits in pump and turbine modes continuously vary over time with regard to head variations. The performance curves of PSH stations are difficult to model since they present a non-convex and non-concave behavior. In most optimization models developed in the literature, the head effects are thus neglected [Chang[01], Habibollahzadeh[86], Nilsson[98]]. However, in order to optimally operate these underground stations, it is essential to use accurate models.

In [Catalao[09]], a nonlinear programming model with some simplified assumptions is proposed. However, such a nonlinear formulation is intrinsically very complex to solve and the applicability of the method is limited to small-sized problems [Luenberger[16]].

On the contrary, with the high performance reached by mixed-integer linear programming (MILP) software, linearization techniques may constitute attractive alternatives. Another advantage of these MILPs is that the optimality level of the solution is known (the solver yields the gap between the final solution and an upper bound of the maximization problem). In this way, several relaxation based algorithms are developed to solve the PSH scheduling with head-dependent reservoirs [Guan[99], Xi[99]].

Hence, some attempts at linearizing the head effects have been presented [Alvarez[18], Borghetti[08], Chen[17], Conejo[02,a], Diniz[08]]. However, such linear approximations require to include additional (integer and continuous) variables as well as new constraints for ensuring the formulation consistency. Since head variations are time-dependent, these additional modeling equations have to be implemented for each time step of the scheduling horizon. In the context of aggregation of several assets, the problem size can thus quickly become overwhelming. In this way, it is worth noting that all the above-mentioned models were implemented within a deterministic optimization framework and have nonetheless face substantial issues regarding the simulation time.

In this chapter, the challenge of modeling the complex characteristics of PSH power plants within the day-ahead stochastic scheduling of an aggregator participating jointly in energy and ancillary services markets is tackled by using a hybrid approach. The methodology consists in decomposing the problem into two complementary modules, embedded within an iterative learning procedure. In this way, the decisions are first optimized at the portfolio level (centralized scheduling of all assets), using the formulation presented in chapter 4 that can be efficiently solved. Then, a holistic simulation model, encompassing hydraulic, electro-mechanical and geological aspects with a high time resolution, is used to evaluate the resulting scheduling of PSH units. This simulation model has been developed by Multitel, the research center leading the Smartwater project (the source of funding of this thesis [Smartwater[18]]). A feedback adjustment of relevant portfolio parameters is thereafter carried out and the procedure is reiterated until reaching an optimal feasible solution. The main benefits of the methodology are three-fold.

Firstly, the proposed formulation takes into full consideration all nonlinearities inherent in the operation of PSH stations. In this way, thanks to the simulation module, both head effects and groundwater exchanges are accurately modeled with a high time resolution. This enables to properly capture their fast dynamics (i.e. around 10 seconds to observe changes in operating conditions), while guaranteeing the practical feasibility and optimality of the scheduling obtained at the end of the optimization.

Secondly, the iterative nature of the hybrid methodology allows to better manage the simulation time so that the method can be applied within a stochastic framework so as to efficiently deal with the different sources of uncertainty (market prices, renewable generation, activation of balancing services, etc.). The approach moreover handles large portfolios co-optimizing different technologies in a multi-market environment

Thirdly, the principle of the hybrid approach allows to easily integrate other sources of nonlinearity (e.g. state-space model of the thermal behavior of buildings with heat pumps supplying operational flexibility to the grid) without significantly affecting the simulation time. Indeed, different simulators can be run in parallel, each one providing feedback information allowing to adequately consider nonlinear effects in the day-ahead scheduling.

Overall, the proposed architecture is independent of both the underlying mathematical tool used for the portfolio scheduling (scenario-based stochastic optimization, robust optimization, chance-constrained programming, etc.) and the simulation models. The procedure is therefore very robust to changes in the portfolio configuration (e.g. if the aggregator expands such as to be considered as a price-maker instead of price-taker) as well as easily adaptable in case of evolutions regarding the market regulation policy.

This chapter is organized as follows. Firstly, the intricate nonlinear dependencies of underground PSH units are presented in Section 5.2. Then, possible topologies for improving the flexibility of PSH units are discussed. The global structure of the sequential hybrid approach is presented in Section 5.4. Then, the design of each of the constitutive blocks is described in Section 5.5. Section 5.6 analyses the practical value and feasibility of the optimization tool with a single PSH unit, whereas Section 5.7 evaluates the scalability of the methodology using a pool-commitment of technologies (in line with those presented in chapter 4). Finally, important conclusions and perspectives are exposed.

5.2 Nonlinear effects governing the operation of PSH units

In order to reliably consider pump-storage hydro plants within the scheduling of a larger portfolio, it is essential to understand and adequately model their operational constraints with an appropriate timescale so as to reflect the actual techno-economic requirements of the studied system and to avoid impractical outcomes.

5.2.1 Head effects

Due to the potentially complex geometry of the basins of underground PSH units, the dependence between the water volumes exchanged between reservoirs and the net hydraulic head may be intricate. However, it is important to accurately model the head variations. Indeed, the value of the hydraulic head has two significant impacts on the operation of variable-speed PSH stations. Firstly, since low and high flow rates can lead to severe erosion of the turbine blades due to cavitation, the head value limits the safe operating domain of hydraulic machines in both pump and turbine modes, leading to discontinuous operating ranges.

$$\Delta P_t = f_1\left(H_t\right) \tag{5.1}$$

The ranges associated with a variable-speed Francis pump-turbine are indicated in Figure 5.1 [Mercier[17,a]]. It can be observed that, even for variable-speed configurations, the operation of sites with large head fluctuations (e.g. former quarries for which the surface of reservoirs is limited by topological conditions) is strongly impacted [Ardizzon[14]]. In this way, a reduction of 10% of the hydraulic head with regard to its nominal value restricts the safe output power domain to around [0.675 pu, 0.875 pu] in pump and to [0.375 pu, 0.8 pu] in turbine. This involves that the allowed domain ΔP_t of the unit constantly varies over time.

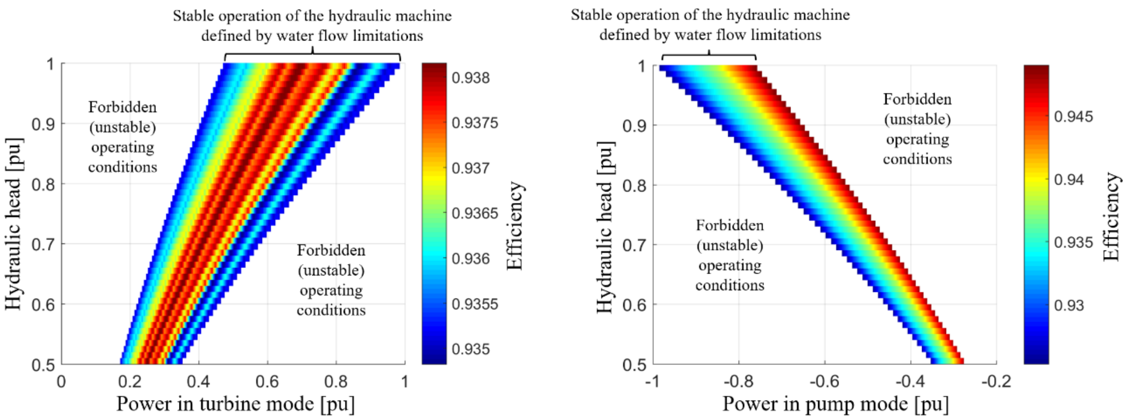


Figure 5.1 – Performance curve of a Francis pump-turbine in both turbine and pump modes [Mercier[17,a]].

Secondly, the head value also influences the efficiency of the PSH stations. Indeed, PSH units are characterized by nonlinear time dependencies between net hydraulic head H_t, flow-rate q_t and the output power P_t, which define the performance curve (efficiency range) of the hydraulic machine. For instance, if the unit is in pumping mode, the water is transferred from the lower to the upper reservoir, which increases the hydraulic head. Hence, the flow-rate must be progressively reduced if a constant output power has to be maintained. This three-dimensional relationship is unit-specific since it depends on the sizing of the storage plant (type of hydraulic machines, geometry of the reservoirs, etc.) and can generally be summarized as follows:

$$P_t = f_2\left(q_t, H_t\right) \tag{5.2}$$

As illustrated in Figure 5.2, the performance curves of PSH stations are intrinsically difficult to model as they present a non-convex and non-concave behavior. As mentioned in the introduction, some linearization techniques have been developed in the past few years. In [Conejo[02,a]], the performance curve was estimated for a hydro-producer (i.e. owning run-of-the-river systems with therefore no possibility of pumping) by discretizing the possible values of the net hydraulic head. Then, for each of these pre-fixed values, a piecewise linear approximation is used to model the resulting function $P_t = f_{2,\text{approx.}}(q_t, H_t)$. The non-concavity of this performance curve has to be modeled using binary decision variables for each period of the planning horizon as well as additional constraints for ensuring the formulation consistency. This head dependence model in turbine mode is improved in [Borghetti[08]] using an advanced approximation methodology designed to better evaluate the power generation between the pre-fixed head values. In [Diniz[08]], the hydro power generation function is approximated by using convexification procedures and regression tools. A method ensuring that the approximated linear formulation remains feasible is presented in [Tong[13]]. Finally, in [Alvarez[18], Chen[17]], the formulation was extended to integrate head effects on the pumping mode of a PSH station.

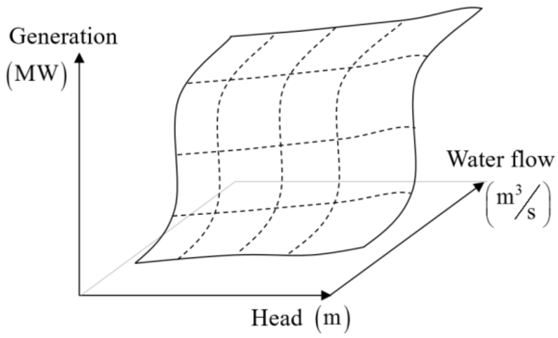


Figure 5.2 – Performance curve of PSH unit in turbine mode (similar shape during pumping).

All these formulations are associated with high computational requirements, and were thus either focused on a limited number of PSH stations, or incorporated into a deterministic optimization framework.

5.2.2 Groundwater exchanges

The rehabilitation of abandoned quarries or mines into small to medium-size PSH (typically between 1 MW to a few dozens of MW) has recently gained increased attention. However, underground PSH units may interact with the surrounding aquifers due to natural permeability of the reservoirs. In this way, when the water within the basin lies beneath the groundwater, then some water infiltrates by leaking through the reservoir walls (in accordance with hydrogeological dynamics), which consequently reduces the capacity of the unit. It results that the sum of water volumes within reservoirs can fluctuate over time (open system).

Generally, such water exchanges with the neighboring environment depend on the geometry and geological structure of the site. Preliminary simulations (using the simulation model presented in Section 5.5.3) carried out by the research center Multitel, aiming at studying the impact of groundwater exchanges on the generated profit, have been performed for a single Belgian PSH station (Maizeret site, which is further described in Section 5.6) with an old open pit mine as lower basin. The objective of this study was to evaluate the relevance of accounting for these water exchanges in subsequent decision-making procedure.

The outcomes (averaged on the year 2015) are presented in Table 5.1 for three configurations: a waterproof reservoir, a reservoir whose height of edges has not been modified to account for groundwater exchanges and a reservoir with an optimized geometry (i.e. edges height has been increased so that water infiltrations do not reduce the unit capacity).

Table 5.1
Influence of groundwater exchanges on generated profit of a single PSH unit.

	Waterproof reservoir	Under-optimized geometry	Optimized geometry
Profit [Euros]	178.046	136.248	171.762
Difference with reference	reference	-23,48 %	-3,53 %

The results demonstrate that groundwater exchanges not only impact the generated profit but also modify the optimal schedule of the PSH station. In this way, it is essential to accurately consider the groundwater exchanges for both the long-term analysis (to make the necessary sizing adjustments) and the shorter-term optimization (to avoid unexpected empty or full upper reservoir that will infer adverse recourse decisions, while ensuring the desired final state at the end of the planning horizon).

5.2.3 Transient effects

Both head effect and water infiltrations are characterized by fast dynamics (around 10 seconds to observe changes in operating conditions). However, the importance of these phenomena strongly depends on the type of hydraulic machine (e.g. in turbine operation, the efficiency of pump-as-turbine (PAT) is more sensitive to power variations than variable-speed Francis turbine).

Moreover, when the station is participating to balancing services, its output power is constantly varying to alleviate grid imbalances. This is illustrated in Figure 5.3 with the real-time activation of the automatic frequency restoration reserves (aFRR), i.e. spinning reserves that are automatically activated (within 7.5 minutes) to restore the balance in the control zone of the system operator.

It can be observed that averaged signals (e.g. at a 15-min basis) may not be representative since they do not fully consider the variability of the real-time activation of the reserves. The resulting smoothing effect may indeed hide some extreme situations (e.g. short periods with full activation) that have to be accounted for.

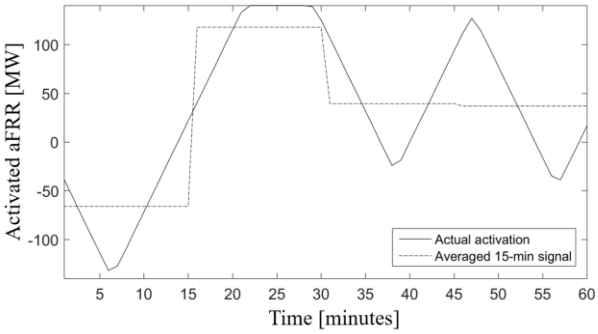


Figure 5.3 – Actual signal for aFRR and same signal averaged on a15-min basis.

5.3 Improving flexibility of PSH units

There exist currently two main options to increase the flexibility of PSH units, namely the variable-speed technology and the hydraulic short-circuit mode.

5.3.1 Variable speed technology

Currently, there are two main configurations of reversible pump-turbine operating at variable-speed. The first configuration, represented in Figure 5.4(a), is characterized by a conversion chain where the pump-turbine is connected to a synchronous machine whose stator is connected to the grid through a power converter. The second possibility, illustrated in Figure 5.4(b) and the current most popular one, is to equip the pump-turbine with a doubly-fed induction machine (DFIM) and a power converter between the rotor and the grid (architecture based on wind turbines). It should be mentioned that pump-as-turbines (PATs) with an asynchronous machine connected to the grid (with a converter) can also be considered.

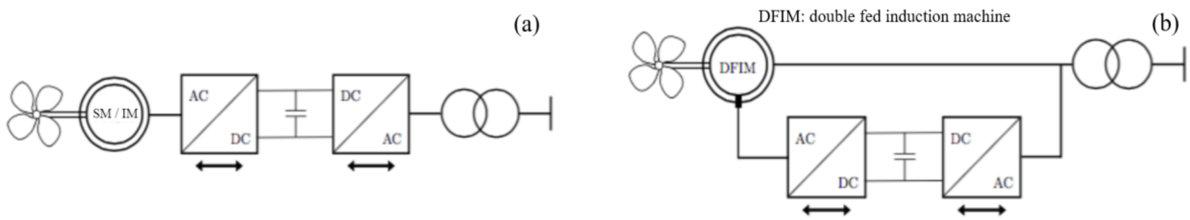


Figure 5.4 – Configurations of variable-speed pump-turbine [Mercier[17,b]].

As presented in Figure 5.5 for a fixed value of the hydraulic head, the variable speed operation allows to regulate the output power even while in pumping mode (thereby acting as a fully controllable load), and therefore provides wider operation ranges in both generation $\left[\underline{p}_g^{VS}, \overline{p}_g^{VS}\right]$ and consumption modes $\left[\underline{p}_p^{VS}, \overline{p}_p^{VS}\right]$.

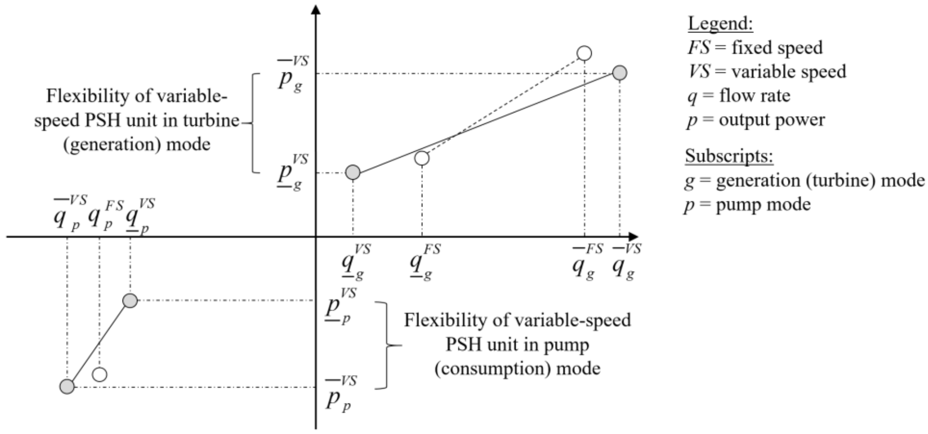


Figure 5.5 – Operating range of a pump-turbine unit with variable speed (solid lines and grey dots) and with fixed speed (dotted lines and white dots) for a given (fixed) hydraulic head [Chazarra[18]].

By integrating head variation, the ranges associated with a Francis pump-turbine can be summarized as in Figure 5.6 [Mercier[17,b]]. The grey area corresponds to the safe domain of the variable-speed technology, whereas red segments indicate the ranges if the speed has to be kept at its nominal value. It can be observed that variable-speed machines are needed so that the pump mode can support head variations and thus operate within its stability margins.

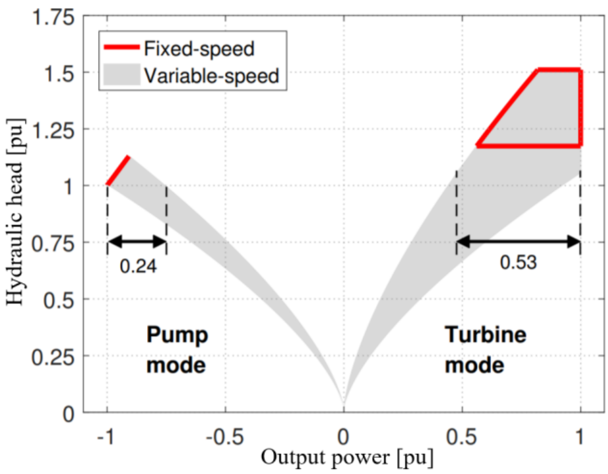


Figure 5.6 – Operation ranges of a Francis pump-turbine.

5.3.2 Hydraulic short-circuit operation

The configuration of a pumped storage hydro station operating in hydraulic short-circuit mode is presented in Figure 5.7. The system is composed of a turbine, a fixed-speed pump, both vertically connected to a synchronous electric machine (operating as a generator when the system is in turbine mode and as a motor when the water is pumped form the lower to the upper reservoir). Thanks to a clutch able to engage and disengage the power transmission between the turbine and the pump, both can be operated either individually or simultaneously. If the plant is in generation mode, the clutch is disengaged, and the generation output is controlled by the position of the turbine guide vanes. If the plant is in pumping mode, the clutch is engaged, and the pump guide vanes are wide open so that there is no power regulation capability. If the plant is in the pump mode and that regulation is needed, the clutch is engaged and both the pump and the turbine operate by employing the hydraulic short circuit.

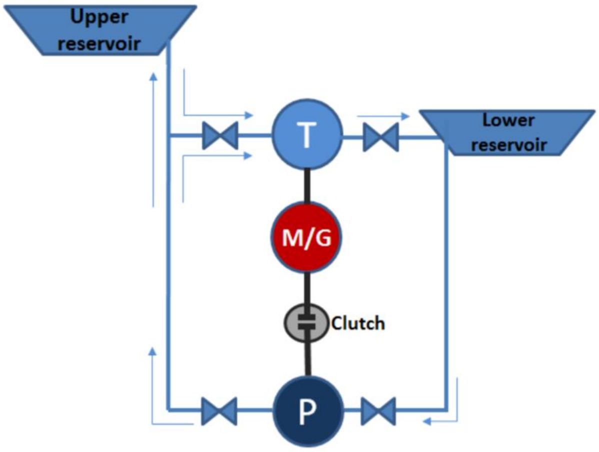


Figure 5.7 – Pumped storage hydropower plant operating in hydraulic short-circuit mode [Argonne[13]].

As represented in Figure 5.8, the hydraulic short-circuit gives the possibility to the installation to regulate power while it is consuming energy (with a power regulation range equal to that of the turbines in operation).

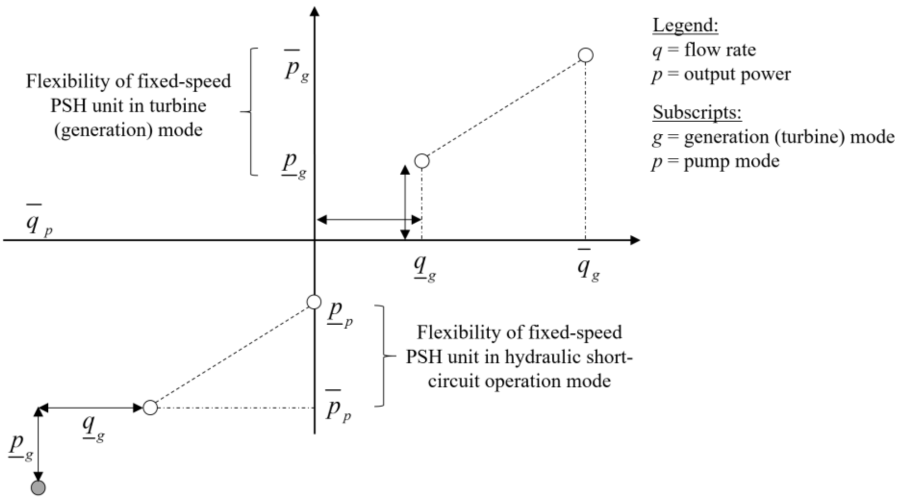


Figure 5.8 – Performance curve of a PSH unit in hydraulic short-circuit for a specific hydraulic head of the PSH installation (grey dot for pure pumping mode).

The main question faced by a potential investor in (underground) PSH applications in the current competitive context is the profitability of the unit. In this way, it is of interest for him to know if the flexibility offered by a more complex configuration (such as the variable-speed control) will bring sufficient additional revenues to recover the higher investment costs. In order to appropriately answer this question, it is important that the scheduling procedure exploits the full potential of the unit, while accurately estimating its economic value. This work is therefore focused on the optimal valorization (short-term scheduling) of these units within a larger portfolio (so as to mutualize the characteristics of the different technologies).

5.4 Model description

As highlighted by works on head-dependent models, accurate linear approximations of PSH operation are associated with high computational requirements. This issue is even more exacerbated when considering larger portfolios with several PSH units and other technologies with their own complexity, and can even prove to be overwhelming when a stochastic approach is implemented. However, in the context of day-ahead scheduling of a Virtual Power Plants participating to electricity markets, it is essential to ensure the robustness of the day-ahead decisions by hedging against the numerous uncertainties (e.g. load, renewable generation, electricity prices, real-time activation of balancing reserves, etc.) influencing the portfolio operation [Khodayar[13]].

In this way, the day-ahead scheduling problem faced by electricity aggregators having PSH units consists in finding a tradeoff between two conflicting objectives, i.e. devising a formulation that is sufficiently sophisticated to yield accurate and reliable solutions (avoid impractical outcomes) while ensuring that such a model is computationally efficient. In the current literature, the modeling equations of PSH stations are often simplified (e.g. considering a very limited number of head values to model the performance curve) so as to ensure tractability of the resulting problem, although these simplifications may result into inaccurate or even infeasible solutions. Here, this issue is addressed using the hybrid approach presented in Figure 5.9.

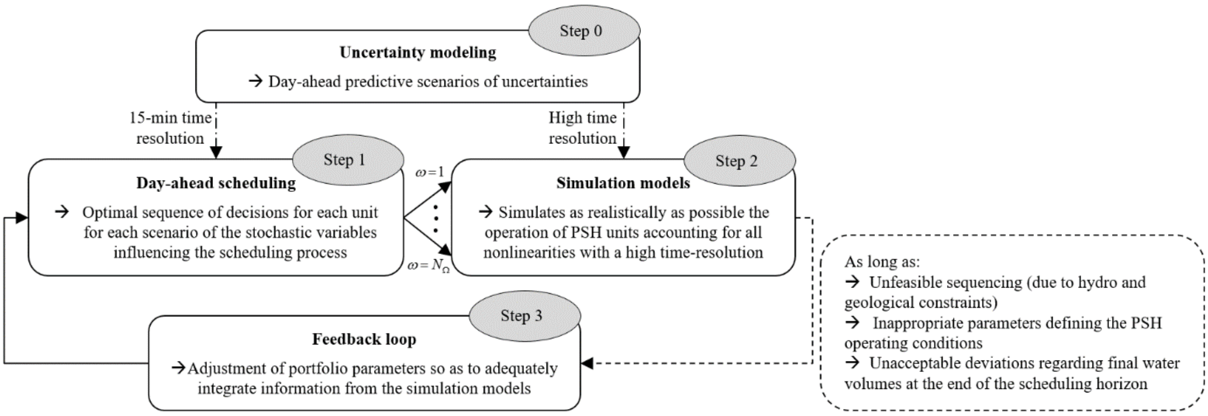


Figure 5.9 – Iterative hybrid approach for the day-ahead scheduling of VPPs.

Once the uncertainties have been modeled (Step 0 described in Section 5.5.1), the scheduling strategy is optimized with a 15-min timescale. The optimization is carried out at the portfolio level within a risk-constrained stochastic environment (Step 1 presented in chapter 4 and reminded in Section 5.5.2), in which uncertainties are characterized through N_{Ω} time-dependent scenarios. At this stage, nonlinear dependencies (i.e. head effects and groundwater exchanges of PSH stations) are therefore neglected to avoid tractability issues.

Since the economic value of PSH stations mainly originates from their flexibility (ability to quickly and efficiently change their output power), their schedule is typically not fixed in day-ahead but is rather used as energy reserves to hedge in real-time against unexpected portfolio imbalances (self-balancing capabilities) while providing ancillary services cost-effectively. In the day-ahead formulation, this intraday flexibility is therefore modeled by adjusting the PSH schedule with respect to the stochastic scenario ω. In this way, at the end of the VPP scheduling (Step 1), various PSH load profiles are defined, each one corresponding to a different scenario considered in the stochastic optimization procedure. Each of the resulting schedules is thereafter integrated into a simulation model that emulates in details the PSH unit operation with a high time-resolution (10 seconds) so as to efficiently mimic the system dynamics (Step 2 described in Section 5.5.3).

The outcomes of the simulation (averaged efficiencies, final water volumes, violation of operating constraints, etc.) are used to adjust the parameters of the portfolio optimization through a control loop feedback mechanism (Step 3 described in Section 5.5.4) and the procedure (Steps 1-2-3) is iterated until convergence of results is achieved.

5.5 Mathematical models

The modeling of the different blocks of the sequential decision procedure, as represented in Figure 5.8, are thoroughly described in subsections 5.5.1 to 5.5.4.

5.5.1 Uncertainty modeling

In order to efficiently account for unexpected events and prediction errors in the decision procedure, reliable day-ahead predictive scenarios have to be considered.

In the portfolio optimization (Step 1), the uncertainties have to be modeled with a 15-minutes time resolution (which corresponds to the time intervals of the scheduling). However, in the simulation model (Step 2), the time step is much more refined (intervals of 10 seconds). This allows considering the quick fluctuations of the real-time activation of balancing reserves (it should be noted that only these variables are available with such a fine resolution). But, these cannot be modeled with traditional tools due to the considerably high dimensionality resulting from the use of a 10 seconds time resolution (since it corresponds to 8640 daily time steps).

In this way, three different models are combined to characterize the different sources of uncertainty. Firstly, the aggregated load and renewable generation (wind and photovoltaic) within the portfolio, the electricity prices in the day-ahead markets are predicted under the form of scenarios with a 15-min time resolution (i.e. 96 sequential values over the next day) in accordance with the copula-BLSTM strategy presented in Chapter 3. Secondly, the prices and liquidity of the intraday market as well as the financial penalties in case of portfolio imbalance are generated (also with a quarter-hourly time resolution) using the copula model constructed in Section 4.3. Thirdly, the scenarios corresponding to the 10 seconds variations of the balancing reserves (FCR and aFRR) are obtained using a simple clustering of historical data (*k*-means with the Euclidean distance[28], where k is equal to the number N_Ω of scenarios considered in the stochastic decision procedure). It should be noted that these 10 seconds scenarios are only used in the simulation model (Step 2), and are then averaged with a 15-minutes time scale for the portfolio optimization (Step 1).

5.5.2 Day-ahead portfolio optimization

As a reminder, two-stage stochastic programming is used as a modeling framework for the day-ahead scheduling of the portfolio (Figure 4.1). Such a method, which enables to model

[28] This technique of time series classification is very basic, and necessitates methodological improvements to obtain more representative scenarios. Practically, the choice of an appropriate metric to aggregate the similar historical realizations with reliable criteria (so as to properly differentiate the different regimes) is a complex but critical task. Several recent research indicates Dynamic Time Warping (DTW) as the best distance for time series classification [Seto15]. Other approaches such as deep learning-based classification, relying on a previous feature selection (or dimensionality reduction) presents also a huge potential [Karim18].

uncertainties and time-dependent decisions, aims at maximizing the expected profit of the portfolio over the following day. In this way, in the first stage (at 12h in day-ahead), facing future uncertainties, the price-taker VPP has to decide for the 24 hours of the following day on the optimal bidding strategy to adopt in the day-ahead market as well as the schedules of the inflexible (slow) power plants. It represents the day-ahead decisions that cannot be modified in the future when the uncertainty is resolved. In this framework, uncertainties are modeled through N_Ω different possible scenarios (daily trajectories) weighted in accordance with their probability of occurrence. Hence, the second stage of the model corresponds to intraday decision (e.g. participation to the Intraday market and operation of PSH units) that aims at avoiding portfolio imbalances while providing the power requested for balancing services. These second-stage decisions are therefore scenario-dependent and can be adjusted according to the realization of the uncertainties.

Overall, this two-stage formulation therefore allows to include, within the day-ahead optimization, the recourse decisions that can be leveraged in intraday, while hedging against the uncertainties associated with the global decision procedure. The global problem can be formulated as a mixed-integer linear program (MILP), and is fully presented in Section 4.4.

For pumped hydro units, we recall that the objective function integrates the operating costs as well as the start-up and shut-down costs in both pump and turbines modes. Then, their operation is constrained in terms of output power (operating range defined by the characteristics of the hydraulic machine), energy capacity (state-of-charge of the unit) and ramping limitations, by ensuring that the ranges associated with these three parameters are not doubly allocated (for both energy and ancillary services contributions). The constraints are formulated such that the uncertainty associated with the real-time activation of balancing reserves is properly taken into consideration.

5.5.3 Simulation model of pumped hydro storage units

At the end of Step 1, each of the N_Ω sequences of decisions of PSH units (each one corresponding to a different stochastic scenario) are evaluated using a simulation model. This model has been fully developed by Multitel, the research center specialized, among other things, in system modeling, (and which was leader of the Smartwater project on which is based this work [Smartwater[18]]).

The model is implemented in RAO (Resource-Action-Operation) language [Artiba[98]], an object-oriented language dedicated to the modeling and simulation of complex systems in which sophisticated applications can be built. This system level simulation tool is here exploited to obtain a global model of PSH units, which allows to take into consideration the impact of any (geo-mechanical, hydrogeological and electrical) parameters in the system. Specifically, this method takes as inputs realistic models coming from partners specialized in electro-mechanics (for the accurate operation of hydraulic and electrical machines) and hydro-geology (for water exchanges between reservoirs and surrounding aquifers).

The input of the simulation model (represented in Figure 5.10) is the quarter-hourly scheduling of the PSH station, and the external balancing signals. The simulator then executes this sequence of decisions with a 10 seconds time resolution, which allows to obtain the complete state of the system (water levels in reservoir, head values, etc.) at each time step of the simulation horizon.

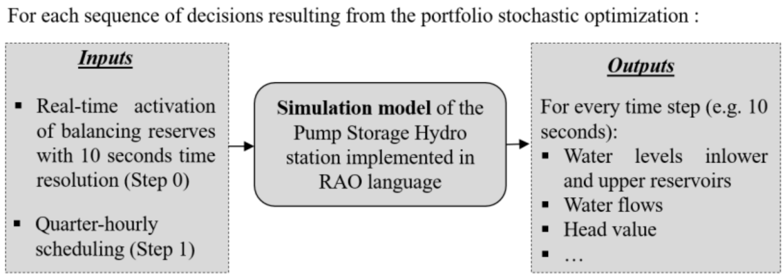


Figure 5.10 – Structure of the simulation model.

The procedure applied to solve the simulation problem, with a particular interest on the way nonlinear effects are modeled, namely:

- The nonlinear relationship between water levels and the head value (due to the potentially complex geometry of the basins coming from the natural form of the quarry or mine that is exploited).
- The forbidden zones (discontinuous operating range in both pump and turbine modes), which depends on the head value;
- The nonlinear relationship (performance curve) between net head, output power and water flows in both pump and turbine modes;
- The nonlinear groundwater exchanges.

The nomenclature relative to the simulation model is exposed hereunder (and illustrated in Figure 5.11).

Constant parameters

H^{base}	Vertical drop between reservoirs	[m]
$H^{up,grd}$	Height of the phreatic table surrounding the upper basin	[m]
$H^{low,grd}$	Height of the phreatic table surrounding the lower basin	[m]
Δ_τ^{sim}	Time step of the simulation model	[seconds]

State variables

H_τ^{up}	Water level in the upper basin	[m]
H_τ^{low}	Water level in the lower basin	[m]
H_τ	Net hydraulic head	[m]
S_τ^{up}	Surface of the upper basin at level H_τ^{up}	[m²]
S_τ^{low}	Surface of the lower basin at level H_τ^{low}	[m²]
Q_τ^{pump}	Water flow in pump mode	[m^3/s]
Q_τ^{turb}	Water flow in turbine mode	[m^3/s]
V_τ^{pump}	Total pumped water volume	[m^3]
V_τ^{turb}	Total turbined water volume	[m^3]
$Q_\tau^{up,grd}$	Groundwater flows in the upper basin	[m^3/s]
$Q_\tau^{low,grd}$	Groundwater flows in the lower basin	[m^3/s]

Firstly, the levels of the groundwater (phreatic) tables with respect to both upper $H^{up,grd}$ and lower $H^{low,grd}$ reservoirs are not impacted by daily (or even weekly) cycles of the pumped-storage units and are considered as constant in the day-ahead scheduling. Similarly, the sedimentation in the reservoirs as well as their impact on the available water volumes can be neglected at the operational stage. It should be noted that these constraints have nonetheless to be considered in the investment phase (sizing of the station topology) by discarding the configurations that violate geomechanical, hydrogeological and hydraulic requirements.

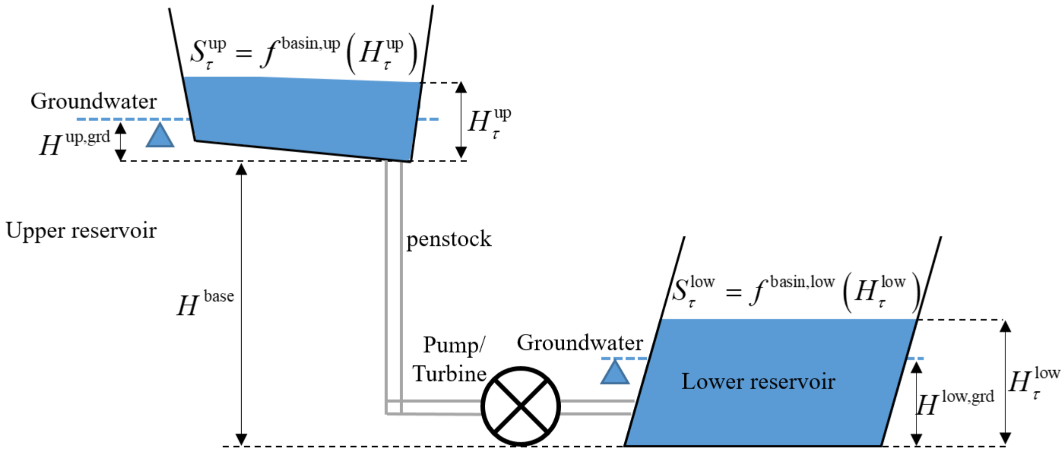


Figure 5.11 – Typical structure of an underground PSH station.

For a given head value, the allowed output power range is imposed (in accordance with Figure 5.1). Then, for the selected output power set point, the variable-speed technology allows to control the rotational speed that maximizes the efficiency. By determining the best efficiency points for the possible combinations of head values and allowed output power, we obtain the nonlinear performance curve (as also represented in Figure 5.1). Here, the performance curves in both pump and turbine modes are discretized so as to be modeled using specific arrays representation. In this way, a three-dimensional matrix reflects yields the efficiency associated with possible head and power values, and the water flows within the hydraulic machine can then be determined as follows:

$$Q_\tau^{pump} = \frac{\eta_\tau^{pump}\mathbf{P}_\tau^{pump}}{\rho g H_{\tau-1}} \tag{5.3}$$

$$Q_\tau^{turb} = \frac{\eta_\tau^{turb}\mathbf{P}_\tau^{turb}}{\rho g H_{\tau-1}} \tag{5.4}$$

with ρ = 1000 kg/m³ and g=9.81 m/s². The water volumes (pumped or turbined) are then straightforwardly calculated:

$$V_\tau^{pump} = Q_\tau^{pump}\Delta_\tau^{sim} \tag{5.5}$$

$$V_\tau^{turb} = Q_\tau^{turb}\Delta_\tau^{sim} \tag{5.6}$$

The groundwater exchanges are modeled as additional flows that vary with respect to the height difference between the water level within the reservoir and the groundwater. These water flows are determined during the sizing of the PSH unit. It is necessary to define two different models for each reservoir, one for a positive height difference between the groundwater and the water level, and the other for a negative difference. These groundwater flows are here approximated using third-order polynomials. However, it should be noted that any other representation could be used, regardless of its complexity.

$$Q_{\tau}^{\text{up,grd}} = \begin{cases} f_1^{\text{up,grd}}\left(H_{\tau-1}^{\text{up}}\right) \text{ if } H_{\tau-1}^{\text{up}} < H^{\text{up,grd}} \\ f_2^{\text{up,grd}}\left(H_{\tau-1}^{\text{up}}\right) \text{ if } H_{\tau-1}^{\text{up}} > H^{\text{up,grd}} \end{cases} \tag{5.7}$$

$$Q_{\tau}^{\text{low,grd}} = \begin{cases} f_1^{\text{low,grd}}\left(H_{\tau-1}^{\text{low}}\right) \text{ if } H_{\tau-1}^{\text{low}} < H^{\text{low,grd}} \\ f_2^{\text{low,grd}}\left(H_{\tau-1}^{\text{low}}\right) \text{ if } H_{\tau-1}^{\text{low}} > H^{\text{low,grd}} \end{cases} \tag{5.8}$$

The water levels within the basins are then adjusted with respect to volumes exchanged with both the other basins and the surrounding aquifers.

$$H_{\tau}^{\text{up}} = H_{\tau-1}^{\text{up}} + \left(V_{\tau}^{\text{pump}} - V_{\tau}^{\text{turb}} + Q_{\tau}^{\text{up,grd}}\Delta_{\tau}^{\text{sim}}\right) \Big/ S_{\tau-1}^{\text{up}} \tag{5.9}$$

$$H_{\tau}^{\text{low}} = H_{\tau-1}^{\text{low}} + \left(V_{\tau}^{\text{turb}} - V_{\tau}^{\text{pump}} + Q_{\tau}^{\text{low,grd}}\Delta_{\tau}^{\text{sim}}\right) \Big/ S_{\tau-1}^{\text{low}} \tag{5.10}$$

where the surface of the upper basin is determined based on the water level in accordance to the function $f^{\text{basin,up}}$ (5.11), whereas $f^{\text{basin,low}}$ gives the surface of the lower reservoir as a function of its water level (5.12).

$$S_{\tau-1}^{\text{up}} = f^{\text{basin,up}}\left(H_{\tau-1}^{\text{up}}\right) \tag{5.11}$$

$$S_{\tau-1}^{\text{low}} = f^{\text{basin,low}}\left(H_{\tau-1}^{\text{low}}\right) \tag{5.12}$$

The net hydraulic head is then given by:

$$H_{\tau} = H^{\text{base}} + H_{\tau}^{\text{up}} - H_{\tau}^{\text{low}} \tag{5.13}$$

The procedure is sequentially conducted for each 10 seconds time interval of the simulated day.

5.5.4 Control loop

The purpose of the feedback loop is to **exploit the information from the simulation model** (e.g. violation of operating constraints, deviation from the target value of water stored at the end of the day) **to adjust the parameters of the simplified optimization problem of Step 1** so as to ensure practical feasibility of the final optimal solution. Indeed, impractical outcomes would result into the inability to fulfill the scheduled power and to provide ancillary services, which may significantly deteriorate the profit of the whole portfolio and have to be absolutely avoided. Likewise, water volumes at the end of the day may differ from the desired (targeted) value, which impacts the economic value of the stored energy at the end of the day and leads to suboptimal management over a longer term perspective.

Specifically, the parameters of Step 1 that have to be adjusted (to take the nonlinear effects of PSH units into account) are the output power ranges and the efficiencies in pump and turbine modes (to adequately represent head variations) as well as water volumes within reservoirs (to avoid negative or excessive volumes within the reservoirs due to inaccurate estimation of water flows, while ensuring to reach the targeted amount of stored energy at the end of the day).

These water volumes are here reflected by the state-of-charge of the unit. In practice, these parameters are discretized with the 15-min time step within the VPP scheduling (Step 1), and are adjusted at each iteration using the outcomes of the simulator (Step 2). Practically, each model parameter is adjusted through an iterative learning process of the form:

$$u_{iter+1} = u_{iter} + \lambda_r e_{iter} \tag{5.14}$$

where u_{iter} is the value of the model parameter of Step 1 during iteration *iter*, λ_r is the learning rate (design parameter) and e_{iter} is the error between the simplified MILP optimization and the simulation model.

Henceforth, the performance of the hybrid tool strongly depends on the choice of an appropriate learning rate (i.e. magnitude of the refinement of model parameters at each iteration). An important step consists therefore in tuning this learning rate in such a way that the hybrid tool converges quickly and reliably towards optimal solution. This procedure, which is portfolio-dependent, is performed offline (before the online utilization of the global hybrid tool).

In this way, **the control loop has to be tuned only once for a particular portfolio, and consequently does not hamper the daily operation of the tool.**

5.6 Case study for a single PSH station

The presented methodology is studied for **a single actual PSH station** whose topology is depicted in Figure 5.12. It consists in a real Belgian site (Maizeret) for which the lower reservoir is a former underground open pit mine (Figure 5.13). The surface of both reservoirs is relatively limited, which incurs significant head effects. The groundwater exchanges are, for their part, determined and validated using ad hoc hydro-geological models [Poulain[18]] that are then included into the simulator (Step 2). The nominal output ranges of the unit (for the nominal value of the hydraulic head) are respectively [6, 8] MW and [4, 8] MW in pump and turbine modes and the energy capacity is of 80 MWh.

At this stage, the potential of flexibility of a single PSH station can already be discussed. By observing the range of variation (at the nominal head), it can be concluded that the participation in balancing services is highly restricted (1 MW for symmetrical products). In the case study, it is therefore considered that the unit is procuring 0.5 MW of symmetrical FCR and 0.5 MW of symmetrical aFRR.

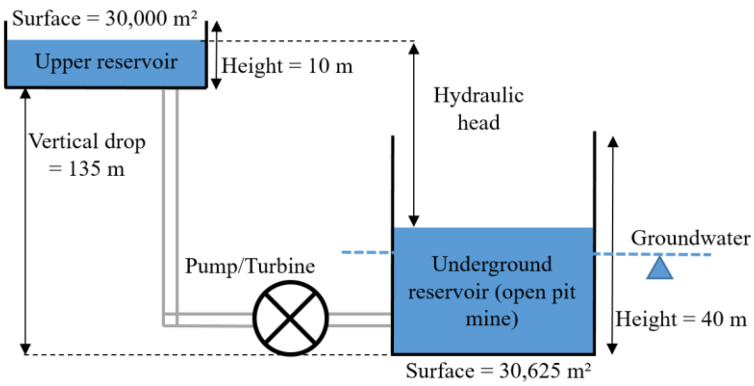


Figure 5.12 – Topology of the modeled PSH unit (Maizeret site).

Figure 5.13 – Actual topological situation of the Maizeret site (source: Google map).

The VPP optimization model (Step 1) is implemented and solved using Matlab. The simulation model (Step 2) is constructed in the RAO environment and converted into an executable (exe) that can be interfaced with Matlab. All computations have been performed on the same Intel® Core™ i7-3770 CPU @ 3.4 GHz (16 Go RAM) as the one used in chapter 4.

The stochastic MILP optimization (Step 1 for a single PSH station) has been run with $N_\Omega = 6$ scenarios during a typical day of the month of July. The calculation load of the RAO simulation model (Step 2) is not analyzed since it takes less than 1 second.

5.6.1 Discussion on the final state (boundary conditions)

In practice, this final value is important since the energy stored at the end of the day directly influences the scheduling of the following day. The simplest option is to implement a **free cycle**, i.e. the algorithm will optimize the scheduling without considering the future. It results that such strategies are likely to empty the upper reservoir to maximize the profit over the scheduling horizon, thereby discarding future opportunities.

Another (more optimal) solution is to consider a **fixed cycle**. The most common strategies are: 1) to systematically refill entirely the upper reservoir, 2) to store half the energy capacity, or 3) store at the end of the day the same amount than the initial value. However, these possibilities do not take future uncertainties into account, and do not harness the whole potential offered by the flexibility of storage utilities. Indeed, it may be useful to empty the storage unit at the end of the week to take advantage of the low prices during the weekend to fill in the upper reservoir.

Throughout this study, the boundary condition is considered with a fixed cycle imposing that the targeted amount of water stored in the upper reservoir at the end of the day has to be identical to its initial value (15 MWh were chosen arbitrarily).

However, it should be noted that, in order to determine the optimal storage value at the end of the day, it may be valuable to consider a **longer-term perspective** (so as to deal with the increased uncertainty over this longer period). This way achieved in [Deane[13]] using a stochastic framework with a weekly look ahead in which the storage trajectory is controlled between the different scenarios using non-anticipativity constraints (such that the energy stored at the end of each day is equal for all scenarios). The time resolution of the simulation has been decreased (hourly time steps) so as to guarantee problem tractability. It should be noted that the seasonal contribution of the storage (optimize week or month-ahead final state) is not relevant for small to medium-sized stations.

5.6.2 Design (sizing) of the control loop

After the validation phase (numerous analyses performed to evaluate the reliability and robustness of both optimization and simulation tools in various conditions), the design of the control loop is carried out. The objective is to determine the optimal values of learning rates λ_r associated with each parameter of the MILP optimization. In other words, we identify the step size of the adjustment of the optimization parameters (of Step 1) based on the outcomes of the simulation model (of Step 2), such as to ensure stability of the procedure and convergence of the solution.

The convergence criteria, for each scenario $\omega \in \Omega$ of the global hybrid tool, are the following:

1) Obtain a feasible scheduling, which amounts to ensure that the output power is always included within its safe operating range and that water volumes within reservoirs are consistent with topological limitations;

2) Restrict the difference between the target final stored energy and the actual value to 1% in order not to impede the economic potential of the station for the next days;

3) Restrict the difference between actual efficiencies (at the end of Step 2) and the values used in Step 1 beneath 1% so that accuracy of the solution is guaranteed.

The influence of learning rates related with output power ranges and water volumes limitations on the convergence speed (number of iterations to obtain the final solution) and final expected profit is represented in Figure 5.14.

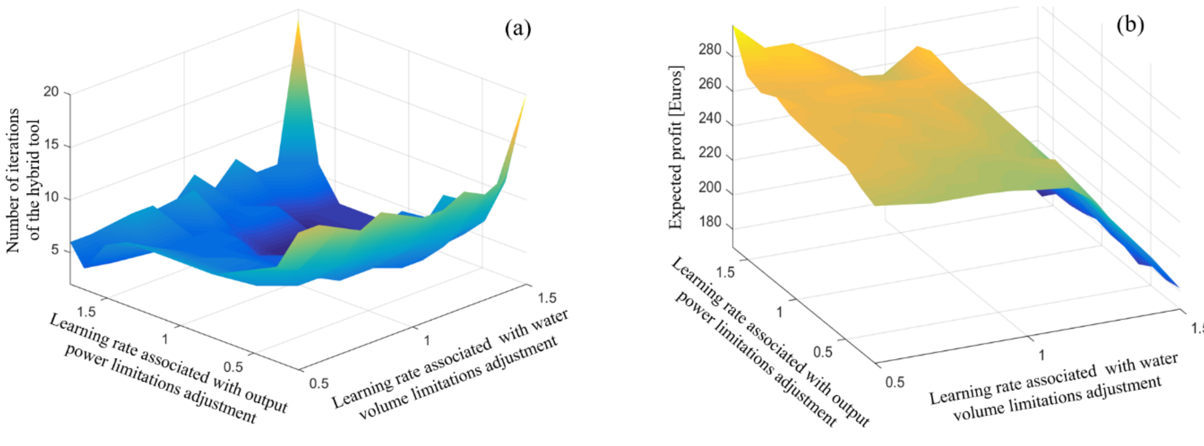


Figure 5.14 – Impact of the design parameters of the control loop on the number of iterations of the hybrid tool (a), and on the expected profit of the final solution (b).

Overall, if the learning rate is too small, the convergence of the optimization process will necessitate many iterations (Figure 5.14(a)), which is very time-consuming. On the other hand, with an oversized learning rate (significant changes in parameters between two consecutive optimizations), instabilities in the learning phase can occur. In this way, instead of smoothly converging towards the solution, the outcomes of the hybrid tool at each iteration oscillate around the solution. Furthermore, from Figure 5.14(b), it is also observed that, even if a feasible solution is obtained, high learning rates often lead to conservative solutions, resulting in loss of revenues. Indeed, the big step sizes between iterations lead to excessively stringent operating ranges, potentially preventing the algorithm from reaching the optimal outcome. **The choice of the learning rates influences not only the convergence speed of the algorithm, but also the optimality of the final solution.**

Based on these observations, a two-step procedure is implemented. At the first stage, the multidimensional (non-convex) relationship between the learning rates and the convergence speed is analyzed. Then, for the set of learning rates leading to a low computational burden (average number of iterations of the hybrid tool smaller or equal to 5), the configuration resulting in the highest value of the expected profit is selected.

Discussion on convergence

As sketched in Figure 5.15, the first run of the portfolio optimization (in which the operation of PSH units is simplified by neglecting non-linear characteristics) yields the starting point of the optimization, around which the search will be performed. In this way, if the initial solution is close to a feasible area, a local optimum can be easily found.

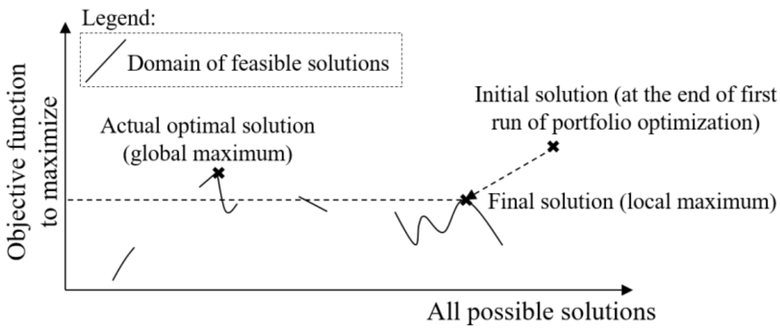


Figure 5.14 – Convergence of the proposed hybrid optimization tool.

However, there is no guarantee that the global optimum will be reached since the procedure can be trapped in a local optimum (maximum). This issue could be overcome by setting up a **multi-start local search algorithm** consisting in launching the first VPP optimization under different initial configurations of PSH parameters. These trial points can be composed by final configurations from previous similar days as well as new random points selected within the allowed bounds. The proposed hybrid sequential tool can then be used (in parallel to keep the computational time in the same range) for the different starting points. Each initial point may potentially lead to a different local optimum, resulting overall in a higher probability to reach the global optimal solution.

Due to the heuristic nature of the procedure, it is not easy to measure the optimality of the final solution. However, the initial outcome (obtained at the end of the first portfolio optimization (Step 1) with the parameters set at their nominal value) yields an upper bound that may be instructive.

5.6.3 Added value of considering the nonlinearities of PSH stations

Once the design of the tool is achieved, the impacts of considering the nonlinear behavior of the PSH unit in terms of operational profit and computational tractability are studied. To that end, simulations are carried out with four different formulations, which differ by their level of complexity.

- In variant #1, all nonlinear effects are neglected, including the definition of the safe (discontinuous) domain of operation characterized by constraints (4.54)-(4.57). The simulator is thus not necessary, and the methodology is condensed into a single run of the day-ahead scheduling (Step 1).
- Then, the discontinuous operating ranges of PSH station are integrated in variant #2, but the hydraulic head is assumed not to vary over time (reservoirs are assumed to have an infinite surface area) , which involves that these ranges are fixed with constant values. This variant is therefore also limited to a single optimization of Step 1.
- In variant #3, the effects of head-dependent reservoirs are adequately considered in the simulation model of Step 2 (but groundwater exchanges are still neglected). The whole sequential procedure is thus performed until convergence is achieved.
- Finally, in variant #4, all nonlinearities (head effects and groundwater exchanges) are included into the simulator, and this formulation therefore yields the solution of reference.

To compare solutions on a fair basis, the day-ahead scheduling of each variant is evaluated through a post-hoc analysis using the RAO simulator of variant #4, referred to as simulator of reference, which accounts for all nonlinear effects. This simulator allows to run each solution, henceforth obtaining the actual state of the PSH unit throughout the day. This enables to determine the violations of operating constraints (of variants #1, #2 and #3), and to subsequently compute the financial penalties that would have been faced by the PSH operator (due to energy imbalances).

The results are summarized in Table 5.2, and includes the total simulation time as well as the expected profit $E(\Phi^{init})$ obtained at the end of the optimization procedure, the adjusted profit $E(\Phi^{final})$ actually realized by the portfolio after accounting for imbalances due to violations of operating constraints (that arise from having neglected the nonlinear effects of the PSH operation), and the associated standard deviation $\sigma(\Phi^{final})$.

It should be mentioned that the grids fees and taxes associated with both consumption and generation modes of PSH stations have been neglected into the simulations. This assumption was necessary given the current Belgian context in which these contributions are so important that the optimal solution for the storage is to stay offline during the whole day (no opportunity for making profit).

Table 5.2
Comparison of formulations of varying complexities in terms of simulation time and profit distribution.

	Time	$E\left(\Phi^{init}\right)$	$E\left(\Phi^{final}\right)$	$\sigma\left(\Phi^{final}\right)$
variant #1	< 1 sec	1395.9 €	273.1 €	59.0 €
variant #2	74 sec	1079.6 €	813.4 €	54.6 €
variant #3	299 sec	932.0 €	860.0 €	79.5 €
variant #4	294 sec	873.1 €	873.1 €	58.9 €

Overall, **it is observed that disregarding the nonlinear behaviors of the PSH station not only leads to a systematic overestimation of the expected profit but also to a sub-optimal solution** (due to subsequent violations of real-time operating constraints). It results that adequately considering these effects constitutes an important step to optimally exploit the economic value of PSH units.

Then, it is interesting to notice that the inclusion of constraints defining the safe operating ranges of PSH stations within the simplified optimization tool (Step 1) considerably increases the simulation time. In this way, the problem is solved in around 74 seconds when the discontinuous operating domain is considered, whereas it takes less than 1 second with a continuous range. However, it can be seen that neglecting these constraints results in strong violations of the technical requirements of PSH stations, which considerably lowers their economic value. An interesting perspective to this work is therefore to improve the mathematical formulation associated with the (naïve) constraints (4.54)-(4.57) to speed up the MILP solving procedure.

Thereafter, the solution of variant #2 (in which the value of the hydraulic head is considered as constant, while groundwater exchanges are neglected) is evaluated using the simulator of reference, and the PSH power profile corresponding to the first stochastic scenario is represented in Figure 5.16.

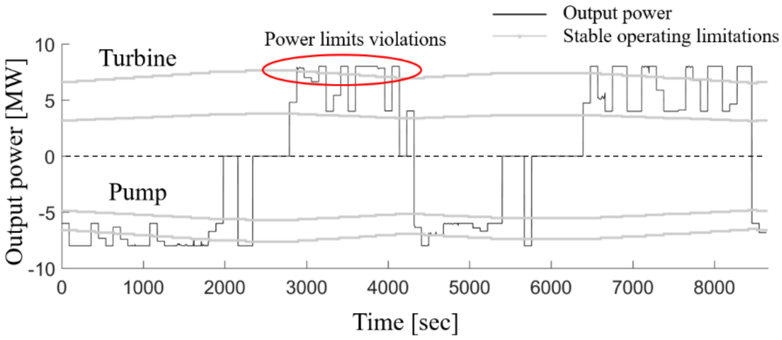


Figure 5.16 – Schedule of PSH station in variant #2 after first iteration of global hybrid tool.

It can be observed that the operational schedule is not feasible since the output power profile is often outside its safe operating range due to head effects. It should be mentioned that water volumes constraints were also violated, which explained the relative difference of -6.84 % in comparison with the reference solution.

Similarly, effects of groundwater exchanges are studied by emulating the solution of variant #3 in the simulator of reference. This solution was obtained after 3 iterations of the hybrid tool. Although the procedure adequately takes head effects into account, the schedule lead to violations of the operating ranges since the head values are incorrectly calculated due the contribution of groundwater exchanges. Figure 5.17 illustrates the evolution of the water level in the upper reservoir for the first scenario, and it is observed that the schedule is actually infeasible (water volume violations at the end of the day). This ultimately results in a relative loss of 1.5% with respect to the reference solution (without taking into account the loss of opportunity for the following day).

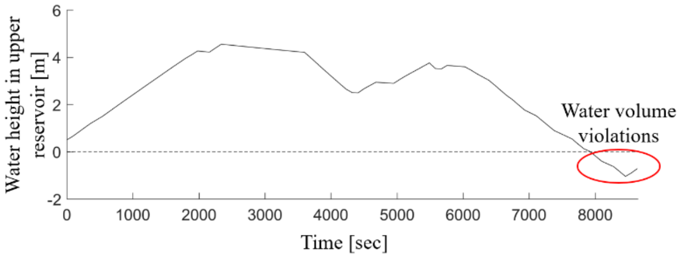


Figure 5.17 – Evolution of water height in upper reservoir if groundwater exchanges are neglected (variant #3).

It can be concluded that **accurately modeling groundwater exchanges for the day-ahead optimization of underground stations is important to avoid unexpected water levels within reservoirs, while ensuring the desired final state at the end of the scheduling horizon**.

To illustrate the convergence of the hybrid tool in variant #4, the violations of operational constraints among the 4 iterations are presented in Table 5.3. The possible violations encompass the final energy level at the end of the day, the safe operating ranges, the water volumes within reservoirs and efficiencies in both pump and turbine modes. A violated constraint is represented by "1" and is thus equal to "0" otherwise.

Table 5.3
Constraints violations across iterations for variant #4.

	Final energy	Power ranges	Water volumes	Efficiency turbine	Efficiency pump
Iter. 1	1	1	1	1	0
Iter. 2	1	1	1	1	0
Iter. 3	0	0	0	1	0
Iter. 4	0	0	0	0	0

5.7 Case study for a portfolio

Finally, the scalability of the methodology is studied through a pool-commitment of technologies. To that end, the same portfolio as the one presented in chapter 4 is used, namely 2 similar PSH stations are included within the portfolio of a Belgian electricity retailer providing ancillary services to the main grid. The VPP is also composed of 2 conventional power plants (CPPs) with a maximum output power of respectively 130 and 80 MW, as well as of 5 wind turbines for a total installed power of 20 MW. The retailer supplies energy to residential clients (peak power of 30 MW), among which 20 % are equipped with rooftop photovoltaic installation between 3 and 5 kVA.

The hybrid tool (corresponding to variant #4) requires 4 iterations to obtain the optimal solution, which shows that **the convergence speed is not affected by the portfolio effect (aggregation of assets) in the considered case study**. The total simulation time, however, now reaches 25 minutes, due to the increased complexity of the MILP optimization problem resulting from the additional technologies.

5.8 Conclusions and perspectives

The objective of the chapter was to implement an approach that integrates all relevant nonlinear characteristics of pump-storage hydro stations with a high time resolution within a computationally efficient formulation of the day-ahead scheduling of virtual power plants. The proposed approach combines two worlds (i.e. optimization and system modeling), and demonstrates that these can be highly complementary.

First, the advantage of the variable-speed technology for enhancing the flexibility of PSH units has been presented. However, it has been observed that, even for such flexible technologies, strong barriers are affecting the profitability of PSH units in the current Belgian regulatory framework. Indeed, such units have no legal recognition, and are therefore alternatively considered as load and generation (without even being supported by subsidies as are renewable energies). PSH units are thus subject to grid fees and taxes for both injection and energy offtakes, which prevents them to be economically operated (especially since lower price spreads are encountered on energy markets). The simulations have thus been performed by neglecting these fees to avoid that the optimal solution is to stay offline during the whole scheduling horizon.

The outcomes of the case study demonstrate that accurately considering these nonlinear effects is a key component to extract the full economic potential of underground stations, and suggest that the proposed hybrid tool (sequential operation of an optimization tool and a simulation model, both included into a control loop ensuring the convergence towards a feasible and optimal solution) offers an effective solution to achieve this goal.

The promising results also open the door to interesting perspectives.

In this way, a multi-start local search algorithm could be developed to help the algorithm finding the global optimum within the high-dimensional search space. However, any other approach allowing to avoid getting trapped in a local optimum can be envisaged.

Then, it seems important to improve the mathematical formulation employed for modeling the safe operating range of PSH units in both pump and turbine modes. It has indeed been highlighted that considering the discontinuous operating range with the (naïve) proposed equations leads to a very weak (not tight) formulation that substantially increases the computational burden of the search procedure.

The principle of the method can be extended to integrate other technologies with complex characteristics (e.g. complex model of responsive load that are associated with comfort constraints due, for instance, to the inertia between the control of the heat pumps and the resulting effect on the temperature within the building).

Finally, the methodology could also be very useful in the context of intraday rescheduling of virtual power plants. Indeed, thanks to ability of the RAO modeling language to simulate any agent-based decision strategy, the simulator can be used to model real strategic behaviors that are complex to implement using traditional optimization tools.

5.9 Chapter publications

This chapter has led to the following publication:

- J.-F. Toubeau, S. Iassinovski, E. Jean, J.-Y. Parfait, J. Bottieau, Z. De Grève, and F. Vallée, “A Nonlinear Hybrid Approach for the Scheduling of Merchant Underground Pumped Hydro Energy Storage,” in *IET Generation, Transmission & Distribution*, in press.

CHAPTER 6

MEDIUM-TERM MULTI-MARKET OPTIMIZATION OF VIRTUAL POWER PLANTS

6.1 Introduction

Currently, one of the major challenges in the management of Virtual Power Plants (VPPs) lies in the efficient planning of decisions relating to different time horizons [Conejo[10]]. However, the common trend in recent studies is to focus on the short-term operation of the portfolio. This can be explained by the emergence of renewable-based generation, whose stochastic nature has to be addressed close to real-time for avoiding deviation penalties arising from market settlements. Although such short-term decision-making is decisive for exploiting the potential of the portfolio, the mid-term perspective is also essential for scheduling general guidelines of the VPP for a longer period. More particularly, participating in forward/futures markets enables securing long-time prices and quantities, limiting interactions with the much more volatile spot market and consequently hedging risks of uncertainties and possible unit contingencies [Ausubel[10]]. Moreover, VPP with sufficient flexibility capabilities may find highly profitable to offer reserves to the balancing capacity market.

However, such mid-term decisions infer constraints on the short-term management through the obligation to uphold these longer-term commitments, and disregarding this dependence may lead to suboptimal or even unfeasible solutions. In this way, the units providing fast power reserves (i.e. FCR and aFRR) have to be committed (online) during the contracted periods (even during unprofitable times). Then, although it may seem tempting to contract as much power as possible in balancing markets (to maximize the revenues for availability of the reserves), one should not neglect the substantial financial penalties inflicted in case of failure to provide the requested power, which is of particular importance for energy-constrained units such as hydropower systems. Specifically, in Belgium, in case of several failures to provide the requested reserves (contract infringement), Elia will apply a temporary exclusion of the market player, which may eventually lead to a complete contract suspension. Such a scenario must be strictly avoided for portfolios whose incomes significantly depend on flexibility offers.

In this context, some studies tackle the medium-term problem but lack to rigorously focus on both the time dependence affecting the global decision process and the accurate modeling of the short-term decision process. In [Cabero[05]], the problem faced by a hydrothermal generation company concerning the planning over a one-year period of its resources in the presence of uncertainty is addressed. The formulation integrates risk-management perspective but strong assumptions are made for modeling the technical aspects of the units operation. In [Baslis[11]], a yearly self-scheduling solved thanks to a three-stage stochastic programming for a pumped-storage plant operator is presented. The producer is modeled as a price-maker but the source of short-term profit is limited to the optimal participation in the day-ahead spot market and the uncertainty is only modeled for the total demand and availability of units. A mixed-integer linear programming (MILP) model that maximizes the weekly VPP profit in the day-ahead market while fulfilling its long-term bilateral contracts is presented in [Pandzic[13,a]]. The work reported in [Hatami[01]] discusses the problem of a retailer having to determine sale prices of electricity while managing the different contracts in order to continuously satisfy the demand. In [Helseth[16]], the optimal scheduling of hydropower plants is addressed by taking into account water inflows and electricity prices as stochastic variables. Several works also investigate the optimal generation of power plants in the context of grid stability and efficiency over a planning horizon of a few weeks up to several months [Baslis[09], Dashti[16], Khodayar[13,b], Martins[14]], or for determining the maintenance plan of different facilities [Barot[08]].

The objective of this chapter is to implement a mid-term optimization tool for a VPP participating as a price-taker in energy markets (market results are independent of the VPP actions) but as a price-maker in reserve capacity markets (pay-as-bid system where the portfolio receives its bidding price, provided that it is competitive enough to be in-the-market). In line with Chapter 4, the methodology does not make any assumption on the portfolio constitution and is designed to incorporate any type of electricity generation, consumption and source of flexibility. Moreover, in Europe, although there exists a growing interest in standardizing (harmonizing) the market rules (e.g. reduction of the contractual period for the procurement of balancing capacity, change of balancing products, etc.), as highlighted in [ENTSOE[15]], many areas have their own subtleties and there exists almost as many regulation policies as there are countries. This work aims at accounting for this issue by introducing an adaptive, robust and flexible structure that can be easily tailored to follow possible evolutions of the market rules. From a practical point of view, the results are illustrated for the case of Belgium, whose regulation policy is close to countries such as Germany, Switzerland or Denmark in which the procurement of balancing capacity is carried out at a mid-term horizon, typically in month or week-ahead.

Up until now, the mid-term decisions of VPPs were usually taken by making simplifying assumptions concerning the short-term operation (neglect inter-temporal constraints of units, integrating a very limited number of stochastic parameters in the formulation, etc.) in order to rely on a single mathematical tool. Here, a different vision is tested for VPPs disposing of more stochastic (renewable sources) and flexible (storage) units. Indeed, for these increasingly numerous actors, the intraday decisions (i.e. ensuring the continuous energy balancing of the portfolio under large amount of uncertainty while being able to provide balancing energy for efficiently contributing to the grid stability) have a significant contribution in the total profit. In this context, it is essential that the short-term operation is exhaustively and adequately modeled in the mid-term decision procedure. Otherwise, some mid-term decisions that may look, at first sight, optimal are in fact seriously over-optimistic.

Here, tactical and operational decision levels are considered jointly in order to cope with the conflicting objectives between the different time horizons (e.g. a higher profit in mid-term reduces the short-term possibilities and a trade-off between these contribution has to be determined). This joint optimization allows taking adequate mid-term decisions based on accurate feedback coming from the short-term simulation. However, with a complex short-term formulation, it is not computationally realistic to add a new mid-term decision stage and the problem has to be adequately decoupled. Hence, in order to hedge against intractability of the resulting problem regarding both time and computer memory requirements, this work proposes to firstly learn (as a pre-processing task) the intricate relationship between mid-term decisions and the resulting profit that can be generated in short-term. Practically, this relationship is established by training a surrogate model of adequate complexity. Then, the medium-term decision process can be solved using the pre-determined model without having to simulate the optimal short-term VPP scheduling problem, whose resolution is very demanding [Carrion06].

The chapter is structured as follows. First, Section 6.2 aims at thoroughly motivating the formulation adopted for solving the mid-term decision problem, and Section 6.3 gives the theoretical background necessary to adequately exploit the mathematical tools involved in the problem solving procedure. Then, Section 6.4 deals with the methodology for handling uncertainty whereas Section 6.5 provides the mathematical formulation of the decision-process. A case study aiming at demonstrating the advantages of the proposed optimization tool compared to other approaches for a diversified portfolio participating in the Belgian electricity markets is illustrated in Section 6.6. Finally, the relevant conclusions and perspectives are summarized in Section 6.7.

6.2 Problem description

This section provides a high-level description of the implemented mid-term decision tool and is divided into two parts. The first one is devoted to the context and motivations of the study with a particular focus on the technical (computational) reasons that have led to the proposed formulation. Finally, the relationship between medium and short term perspectives is explained in more details. It should be noted that the underlying mathematical tools are further described in Sections 6.3 and 6.4.

6.2.1 Motivation of the proposed formulation

The mid-term decision procedure can be regarded as a multi-market optimization whose main issue is to determine the optimal use of available resources within a given portfolio. More particularly, the VPP has the choice of using its flexibility either for participating in energy markets (from long-term to intraday), or as reserve capacity for balancing services, or even as hedging capacity for avoiding imbalance penalties in case of unexpected event.

Generally, as illustrated in Figure 6.1, the mid-term decision procedure of a VPP consists in a three-stage stochastic decision process, which involves making optimal decisions at different time horizons (i.e. medium-term, day-ahead and intraday) before uncertain events are revealed.

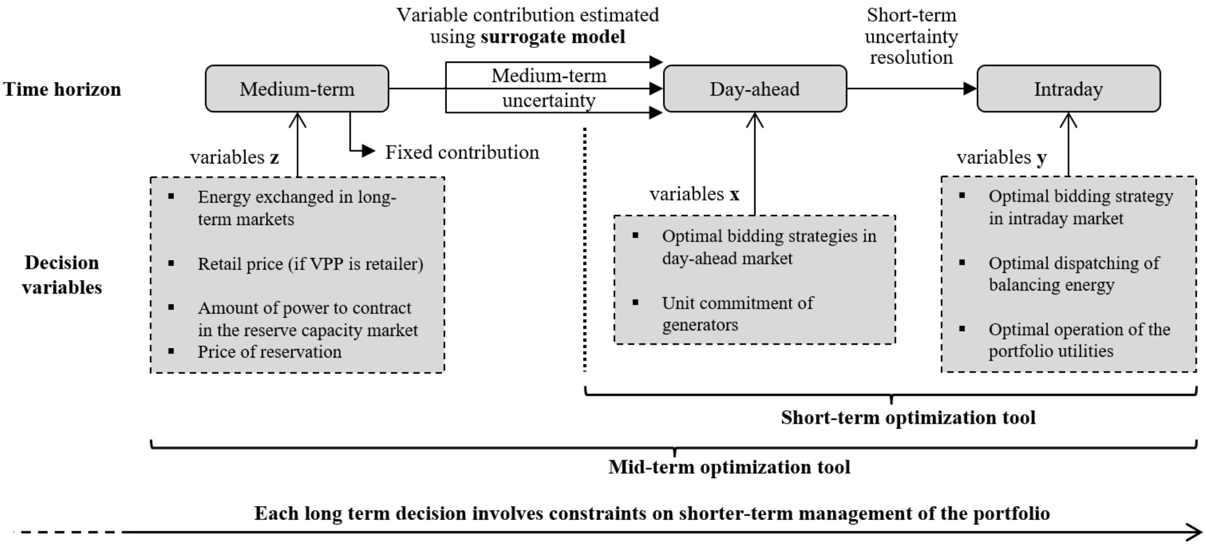


Figure 6.1 – Rationale of the medium-term decision making tool.

In the mid-term (week-ahead), the VPP has to decide about its participation to electricity futures and balancing capacity markets. Practically, decisions for price and volume have to be taken. Hence, as the VPP is a price-maker in the reserve capacity market, nonlinearities are introduced into the formulation due to the product of both variables (price and volume) for computing the profit.

It is nonetheless essential to accurately determine these medium-term decisions **z** since they guarantee a fixed revenue to the VPP over the contractual period. Moreover, they also induce constraints on the operational portfolio management that will impact its daily revenues, because of the necessity to fulfill these longer-term contracts.

Henceforth, failing to accurately model the day-ahead and operational decision-making processes may lead to suboptimal mid-term actions. A key objective of the work is thus to implement a detailed short-term decision tool complementarily to the mid-term one. More particularly, the procurement of balancing services, both in terms of capacity and energy have to be considered and optimized at the portfolio level.

This approach involves to include the amount of balancing energy provided by each portfolio unit for the different products as additional decision variables into the formulation. Moreover, scenarios realistically modeling the needs in terms of real-time balancing energy also have to be considered.

Consequently, as exemplified in the case study of Section 6.6, considering nonlinearities in the objective function as well as the important number of decision variables and stochastic parameters influencing the problem, it is impractical to implement a reliable global optimization tool able to converge towards an optimal solution in a reasonable amount of time without resorting to simplifying assumptions. Hence, in order to ensure tractability of the general problem, a surrogate-based optimization is here privileged. To that end, a surrogate model of tailored complexity is used for modeling the dependence between the expected daily profit and the mid-term decisions. This model is preliminarily constructed offline (first step) in order to be used afterwards in the mid-term optimization tool (second step). This decomposition into two consecutive steps allows a better accuracy and increased controllability of each individual block while efficiently capturing their inter-temporal constraints.

6.2.2 General structure of the proposed formulation

The structure of the decision tool is represented in Figure 6.2.

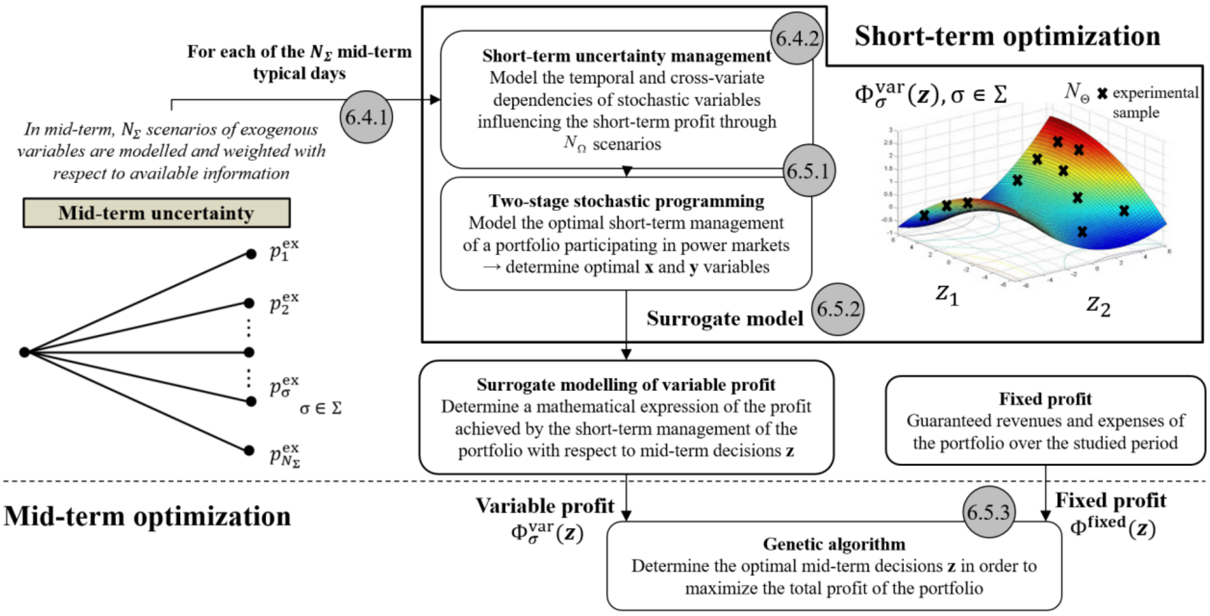


Figure 6.2 – General structure of the mid-term optimization tool whose blocks are fully described in the corresponding subsections of Sections 6.3 and 6.4.

In the mid-term, N_Σ typical days of uncertain variables are created based on the available information (Section 6.4.1). For each of these typical days, the day-ahead bidding strategy under risk and uncertainty of the VPP has to be optimized. In order to efficiently account for stochasticity in the decision-making process, N_Ω scenarios of possible prediction errors need to be simulated (Section 6.4.2).

At this stage, the objective is to construct a surrogate model of the expected short-term profit with regard to mid-term decisions (Section 6.5.2). For this purpose, N_Θ experimental samples fixing mid-term decisions are optimally selected. For each sample, the short-term optimization tool is used for computing the expected profit (Section 6.5.1).

Once a statistically representative model is estimated for every typical day, a genetic algorithm calculates the optimal mid-term decisions for maximizing the total profit over the considered horizon (Section 6.5.3).

Discussion on the transition between typical days

In the proposed framework, the transitions between typical days are not considered. Rather, each day σ is associated with its probability of occurrence p_σ^{ex}, and the results (short-term profit in each day) are then averaged in accordance with these weighted factors p_σ^{ex}. Such an assumption does not lead to significant loss of information, except potentially for energy-constrained technologies. Indeed, as discussed in Section 5.6.1, the conditions may be such that it is preferable to preemptively modify the amount of energy stored at the end of the day (compared to static approaches where the final value is equal to the initial one) to better account for future opportunities. In such situations, it may be interesting to properly consider the temporal organization of the scheduling horizon. However, this issue has not been tackled here, and constitutes one of the most salient perspective of the work.

6.3 Mathematical background

The mid-term optimization is taken with a week-ahead perspective (FCR and aFRR contracts), and the resulting three-stage week-ahead decision procedure is divided into two steps. First, a surrogate model of the variation of the daily profit with respect to mid-term decisions variables is constructed. Then, this model (that can be quickly evaluated) is used in the nonlinear mid-term optimization (solved using a genetic algorithm). This section aims at providing the theoretical background necessary to get more familiar with these mathematical tools. First, the principle of surrogate modeling is introduced, along with the toolbox used in this work. Then, the theory and field of application of genetic algorithms are presented.

6.3.1 Surrogate modeling

The principle of the global surrogate model is to accurately mimic the original system (i.e. short-term scheduling of a portfolio participating in electricity markets) over the entire design space (i.e. operating range of the mid-term decisions). The underlying objective of this modeling procedure is to create a mathematical model that can be safely used instead of the original stochastic scheduling procedure. In other words, the surrogate model is only an intermediate step towards solving a more important problem, and consists in an accurate approximation to replace an expensive reference process (true objective function). The meta-model can be evaluated much quicker than the true objective function, and requires only a one-time computational investment to be constructed.

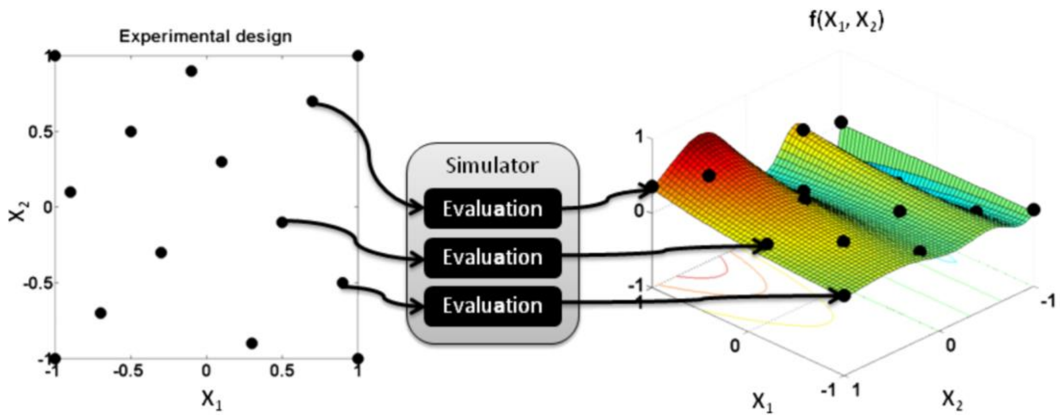


Figure 6.3 – Construction of the surrogate model: a set of data points is estimated by the simulator, and an approximation model (surrogate model) is then fit to these data points [Couckuyt[13]].

The quality of the approximation depends on design choices, mainly consisting in choosing the most efficient data sampling strategy, the right model for the problem at hand (polynomial, kriging, etc.), while adequately tuning the design parameters of the model (hyperparameter optimization) and finding the optimal balance between model accuracy and computational burden [Couckuyt[13]].

Regarding the sampling strategy, it may be interesting to investigate more intelligent policies than the traditional experimental designs such as Latin hypercubes[29] or factorial

[29] Latin hypercube sampling (in contrast with random sampling) aims to evenly distribute the sample points across the design space. To that end, each input distribution (of the multidimensional space) is divided into intervals of equal probability, and one sample is taken from each interval.

designs[30] where the sampling points are all chosen in a single step. The data points should indeed be selected iteratively (based on the information gathered from previous simulations) at locations where the need of new information is the most important [Sugiyama[02]]. This learning process is referred to as active learning (or sequential design, or adaptive sampling).

The structure of the SUMO toolbox [Gorissen[10]] that is used within this work is illustrated in Figure 6.4. First, an initial design (typically a sparse Latin hypercube) is generated and its constitutive data points are evaluated (with the short-term portfolio optimization presented in chapter 4 and reminded in Section 6.5.1). Then, different surrogate models are constructed, and the accuracy of each one is estimated using an error metric (such as sum of squared errors between the sample points and the surrogate model). It should be noted that for interpolation models (such as kriging), for which the response surface automatically passes through all input points, it is mandatory to keep some data points outside the model building phase to serve as a validation set (to compute a representative error metric). Each type of models is characterized by a set of hyperparameters which can be tuned, such as the degree for polynomial models (e.g. second order model), or the smoothness parameters for radial basis function (RBF) models. Then, if none of the stopping criteria is met for any of the different built models (i.e. minimum accuracy reached, maximum number of samples, or maximum run time exceeded), additional samples are selected in the worst regions (i.e. where the model presents the highest error or significant nonlinearities). After each new sample selection, the model is trained and adapted to the new available information in order to obtain an optimal model complexity.

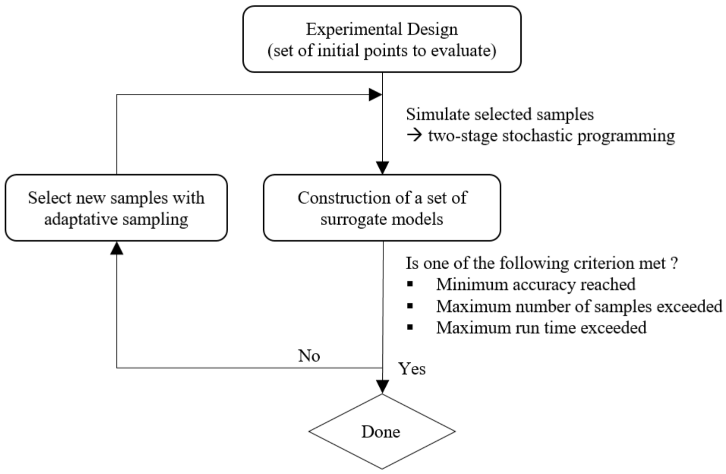


Figure 6.4 – Structure of the surrogate model builder (SUMO toolbox).

Among the different possible types of model, several can be quickly discarded. In this way, neural networks and support vector regression (SVR) are robust but the training phase necessitates numerous inputs (samples), which is not computationally efficient. Then, splines are a form of interpolation that is only valid for two-dimensional problems.

We focus therefore, here, on rational (ratio of two polynomials functions) and Gaussian models (in particular RBF and kriging methods).

[30] Factorial designs represent a sampling strategy where the samples are given by every possible combination of the variables (typically limited to a lower and an upper bound).

6.3.2 Genetic algorithm

There exist many different visions to classify the algorithms to solve optimization problems due to the numerous different characteristics that can be encountered (e.g. multi-objective, stochastic, dynamic, differentiable, mixed-integer, non-linear, etc.). One simple possibility is to divide the optimization algorithms into three categories: exact algorithms, heuristics and metaheuristics.

Exact algorithms are designed in such a way that it is ensured that the optimal solution will be found in a finite amount of time. However, for difficult optimization problems (e.g. large-scale non-linear formulations), this finite simulation time may drastically grow with respect to the problem dimensions. Moreover, for some optimal search procedures such as gradient descent, the algorithm can get trapped into a local optimum (best solution within a neighboring set of possible candidates but not necessarily for all possible solutions).

Heuristics do not provide the guarantee of finding an optimal solution, but usually find a good solution in an amount of time that can be controlled. Heuristics are problem-dependent techniques, i.e. they are adapted to the problem at hand and try to take full advantage of its particularities [Maringer[05]]. Such class of algorithms typically start off with a more or less arbitrary initial solution, iteratively produce new solutions and make the best choice at each stage with the expectation of finding the global optimum. This greedy approach often results in the algorithm converging towards a local optimum.

Meta-heuristics, for their part, are problem-independent algorithms with a higher level perspective that does not take advantage of any specificity of the problem. As such, they can be used as black boxes for any type of problem with nonlinearities, discontinuities in the search space, and noise in the data. They are characterized by an algorithmic mechanism to avoid getting trapped in confined areas of the search space (local optima). For instance, in the simulated annealing technique, we accept temporary deteriorations of the solution so that the solution space can be explored more thoroughly, which improves the probability to reach the global optimum [Sousa[12]].

Based on the previous analysis, for solving our nonlinear mid-term decision procedure, a meta-heuristic (**genetic algorithm**) is selected. Developed in the 1960s in [Holland[62]], genetic algorithms were applied for the first time in 1975 to solve an optimization problem in [De Jong[75]]. At that time, computational capabilities of computers did not allow them to solve real large-scale problems, and it was only in the 1990s, thanks to the advent of informatics and their popularization by [Golberg[89]] that they were revealed to the scientific world. Since then, genetic algorithms are commonly employed as mono- or multi-objective optimization tools to solve issues pertaining to various domains.

The genetic algorithm translates the procedure of **natural selection** where the best individuals are selected for reproduction to create the next generation. First, the **initial population** is obtained with a random choice of a given number of individuals. The fitness of these candidates is evaluated, and the best ones are then selected [Regnier[03]]. Then comes an evolution phase when variation operators are applied to parents so as to create a new set of individuals (children) that will, in turn, be evaluated. In this way, the offspring inherits the genetic material of the parents (the genes of parents are shared through a mechanism known as **crossover**), at the exemption of some possible **mutation** (genes have a low random probability to unexpectedly mutate). The population is repeatedly modified from one generation to the next one so that the population evolves towards optimality.

Practically, four different phases are considered in a genetic algorithm: initial population, selection, crossover, mutation.

Initial population

As represented in Figure 6.5, the algorithm starts with a set of N_i individuals, known as the initial population. From an algorithmic point of view, each individual represents a possible solution of the optimization problem. In this way, each individual is characterized by a chromosome, which is composed of N_x genes (each one corresponding to a particular decision variable).

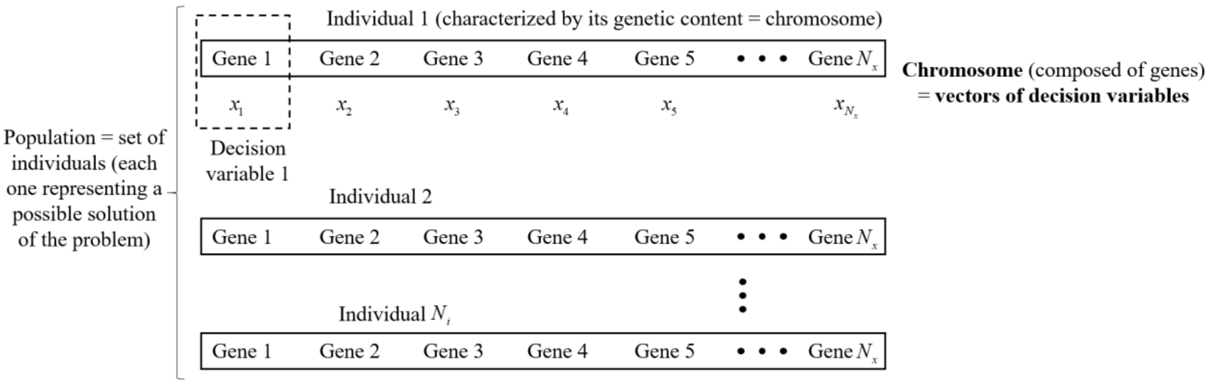


Figure 6.5 – Nomenclature of genetic algorithm.

The fitness of each individual is then evaluated by computing its fitness score (results of the objective function with respect to values of the N_x decision variables, in which penalties can be added in case of violation of optimization constraints so that the algorithm can avoid infeasible solutions).

Selection

The objective of the selection procedure is to duplicate the best individuals and to eliminate the less adapted individuals, while maintaining the size constant. The selected individuals will then reproduce (share their genetic material) to form the next generation. It must be noted that the selection must be able to select the best candidate while maintaining diversity [Regnier[03]]. Different selection procedures are summarized in [Golberg[91], Sareni[99]] but are all based on the fitness score of each individual.

Crossover

The crossover is a random process applied sequentially to pairs of parents (taken randomly) from part of the population that was selected [Sareni[99]]. It consists in exchanging the genetic material of the parents to form two new individuals (children), which allows to explore the search space [Deb[01], Herrera[98]]. Within the objective to keep best individuals, the crossover procedure is not applied to all selected parents.

Mutation

Mutation is a random alteration (generally of the order of 0,001 to 0,01) of some genes from a chromosome. The new value of the mutated gene is randomly chosen within the range of variation associated with the variable [Michalewicz[92]].

The main purpose of the mutation is to maintain diversity among the individuals of the population. Indeed, without such mutations, no new genetic characteristics would appear, and the search procedure would be drawn into local optima [Sareni[99]].

The general structure of the genetic algorithm can be summarized as follows:

Pseudocode
Generate the random initial population
Score each individual of the population by computing its fitness function
REPEAT: At each step, the algorithm uses the individuals in the current generation to create the next population, by relying on the following steps:
 Selection of the parents, based on their fitness
 Children are generated by combining the genes of a pair of parents *(crossover) or by random changes in the chromosome (mutation).*
 Compute fitness function
UNTIL convergence (stopping criteria is met)

6.4 Uncertainty management

In medium-term (even in week-ahead), the accuracy of predictions along the whole considered horizon is questionable due to the significant variability and uncertainty surrounding the variables of interest.

The proposed method to deal with the mid-term uncertainty is summarized in Figure 6.6 and consists in firstly defining N_Σ typical days representing different realizations σ of the exogenous variables of the problem (whose realizations are independent of the other variables of the problem). As previously presented in Table 4.2, these variables are the total consumption and renewable-based generation within the portfolio as well as day-ahead electricity prices. The N_Σ typical days are appropriately weighted in accordance with their probability of occurrence p_σ^{ex}. Each of these typical days is associated with deviation scenarios ω ∊ Ω in order to reliably account for forecast errors.

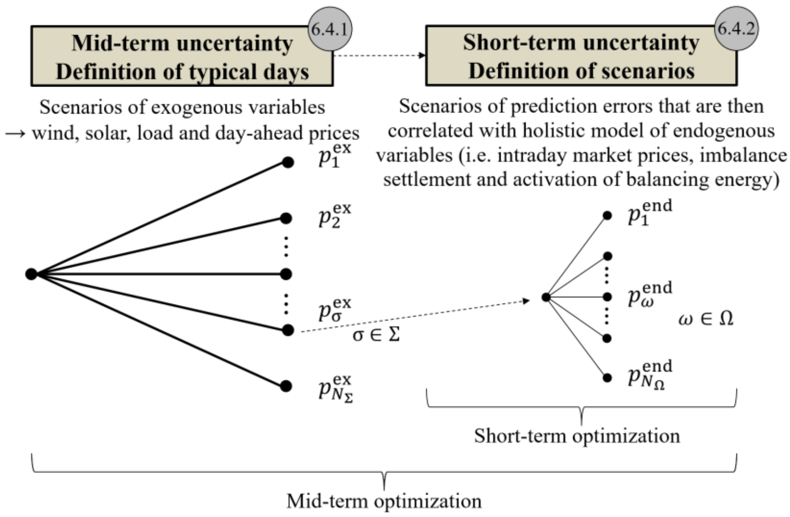


Figure 6.6 – Methodology for modeling mid-term uncertainty.

Then, the scenarios of errors are constructed using historical prediction errors [Exizidis14] and are correlated with the endogenous variables of the problem (using a general probabilistic model encompassing the intricate multivariate dependence structure). The endogenous variables include the Intraday market prices and liquidity, the imbalance settlements, and the amount of power requested in real-time for the different balancing services.

6.4.1 Mid-term uncertainty management

In the mid-term, the purpose consists in modeling the exogenous variables (i.e. wind and solar generation, total load and prices in the day-ahead energy market) whose realizations are independent of other variables of the problem. Since the transitions between consecutive days are neglected, the scenarios can be represented by (independent) statistically representative typical days. To that end, a k-means clustering (with a Euclidean distance) is used on relevant historical realizations, which are composed of days with similar conditions. This procedure allows to properly define $k = N_\Sigma$ typical days with their associated probabilities p_σ^{ex} of occurrence.

6.4.2 Short-term uncertainty management

The methodology is similar to the one presented in Section 4.3, and is summarized in Figure 6.7. The short-term uncertainty relates to the generation of scenarios $\omega \in \Omega$ of the endogenous variables of the problem in the context of the two-stage (day-ahead) stochastic programming.

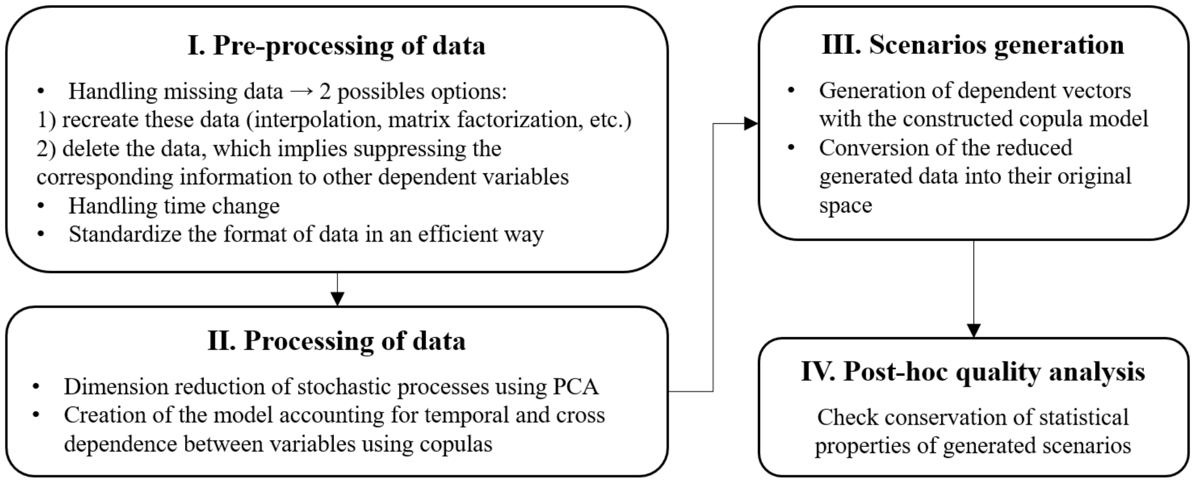


Figure 6.7 – Scenarios generation methodology in the context of stochastic programming using historical data.

First, an important task consists in properly processing (cleaning) the available databases. In this way, it is necessary to handle the missing data by either recreating the missing information (using, for instance, linear interpolation or matrix factorization [Koren09]), or by accepting to lose the information (suppression from the database). Then, it is also important to deal with time changes (i.e. shift from winter time to summer time, and conversely), and to standardize the format of all data in an efficient way to simplify the subsequent processing procedures.

Then, the dimensionality of historical data is reduced by applying principal component analysis (PCA), which converts the original dimensions into linearly uncorrelated variables (called principal components). The transformation iteratively attempts to determine the component that contains the maximum variability in the data, so that, at the end of the procedure, most of the information is contained into a limited number of components (which allows to eliminate many dimensions without much loss of information). Based on this space of reduced-dimensionality, an empirical copula model is constructed.

As explained in Annex B, the copula model can then be easily exploited to generate random vectors (encompassing the statistical information of the historical data), and the inverse PCA is applied to convert the data into their original dimensions. Finally, post-hoc analysis can be performed to analyze the accuracy of the methodology.

6.5 Mathematical formulation

The blocks of the mid-term optimization tool, as represented in Figure 6.2, are described in subsections 6.5.1 to 6.5.3.

6.5.1 Short-term decision procedure

As presented in Section 4.4 (and reminded in Section 5.5.2), two-stage stochastic programming is used as a modeling framework for the day-ahead management of the portfolio. Such a technique aims at maximizing the expected value of the variable component of the profit Φ^{var} on a one day period. In the first stage, facing future uncertainties, the price-taker VPP has to decide on the optimal bidding strategy to adopt in the day-ahead market as well as the unit commitment. More specifically, the schedules of the slow and inflexible plants have to be specified. It represents therefore day-ahead decisions **x** that cannot be modified in the future when the uncertainty set is resolved. The second stage of the model corresponds to hour-ahead operation of the flexible plants (i.e., thermal units and hydro plants with storage) that aim at avoiding portfolio imbalances while providing the power requested for balancing services. The second-stage decisions **y** are therefore scenario-dependent and can be adjusted according to the realization of the stochastic parameters. The problem can be compactly formulated as a mixed-integer linear program (MILP) and the objective function can be expressed as follows:

$$\max_{\mathbf{x},\mathbf{y}} \mathrm{E}\left(\Phi^{\text{variable}}\right) \tag{6.1}$$

$$\mathrm{E}\left(\Phi^{\text{variable}}\right) = \Phi^{\text{var}} = \left(1-\beta\right)\Phi + \beta.\mathrm{Risk} \tag{6.2}$$

Then, all constraints associated with the portfolio management (energy balance equation, cost-optimal allocation of resources, risk-management, etc.) and related to the technical requirements of each technology are fully presented in Section 4.4.

6.5.2 Surrogate model of the variable short-term profit

In the context of mid-term stochastic optimization, it is necessary to compute the variable contribution of the profit a large number of times (for different sets of mid-term decisions **z**), which involve solving the resulting MILP at each iteration. Hence, in order to reduce the simulation time, a surrogate model is used to approximate the expected short-term profit of the VPP with regard to constraints inferred by mid-term decisions. This constitutes indeed a beneficial solution since the number of samples N_{Θ} for building an accurate model is much lower than the number of iterations required for optimization [Queipo05, Simpson08].

However, since the mid-term uncertainty cannot be captured by a single representative day, a different model has thus to be constructed for each of the N_{Σ} mid-term typical days. These mathematical models are therefore determined in pre-processing (i.e. before mid-term optimization) with the purpose to obtain the best trade-off between model accuracy and computational burden for building the models.

With this objective, various models (polynomial and rational functions, as well as Gaussian models) can be studied thanks to the SUMO toolbox, which is a tool designed for adaptive surrogate modeling with sequential design [Gorissen10], as explained in Section 6.3.1.

Illustrative example

The surrogate models for two typical days of July are represented in Figure 6.8, where the variation of the short-term profit with respect to the amount of contracted FCR and aFRR capacity (in MW) is displayed.

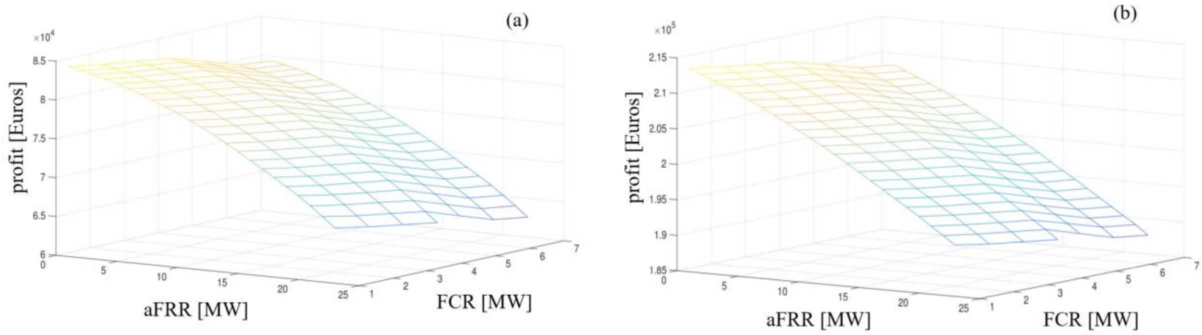


Figure 6.8 – Response surfaces of the short-term profit with respect to the capacity contracted in FCR and aFRR for two typical days of July.

It can be seen that the response surfaces have the same smooth profile, but differ regarding the short-term profit that can be realized (mainly due to the different wind/load conditions). As expected, when more reserve capacity is contracted, the short-term profit is reduced due to the lower amount of flexibility available in short-term to face forecasting errors to take advantage of opportunities in both intraday market and imbalance settlement.

These results highlight the importance that the flexibility providers are remunerated for the availability of the reserve, and illustrate the importance of the mid-term decision process where the optimal trade-off between mid-term and short-term contributions has to be determined. It can nonetheless be observed that the impact of the contracted FCR capacity is less significant than the aFRR, which suggests that the optimization strategy will firstly try to optimize the participation to FCR before offering the remaining flexible capacity to other products (unless the remuneration of aFRR capacity compensates this effect).

Then, the relationship between the number of samples to construct the surrogate model and the modeling error (computed on a test set composed of N_{test} data points that are not used during the training) is presented in Figure 6.9(a), whereas the simulation time is illustrated in Figure 6.9(b). It should be noted that the root mean square error (RMSE) is used as error metric:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_{test}} (y_i - f_i)^2}{N_{test}}} \tag{6.3}$$

where y_i the output of the surrogate model and f_i the actual value computed by the short-term optimization (two-stage stochastic programming).

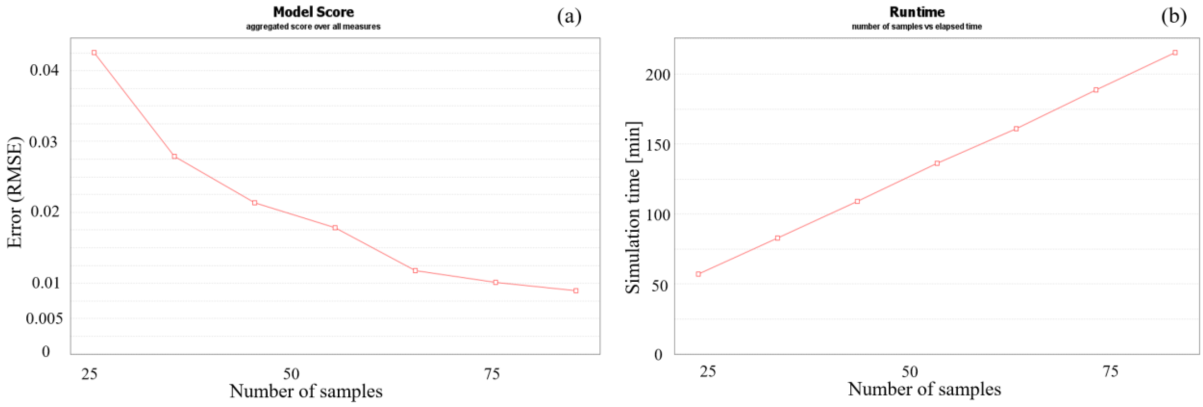


Figure 6.9 – Evolution of the modeling error (a) and simulation time (b) with respect to the number of samples.

6.5.3 Medium-term decision procedure

The mid-term horizon can be characterized by different types of decisions depending on the profile of the portfolio manager. For instance, an electricity retailer, in addition to determine its optimal bidding strategy in forward/futures and balancing services markets, is also eager to define the best selling prices to its potential clients. The decision tool must thus be general and resilient enough to accommodate different objectives.

Generally, the expected total profit Φ_{tot} over the considered period is composed of a fixed and a variable contribution. The fixed component Φ^{fixed} is subject to uncertainty related to the clearing of reserve capacity markets in which the VPP is considered as a price-maker. In this work, relevant historical market data are used to build stepwise curves representing the probabilities $p_{\eta,z}$ of acceptation of offer (for product z) with regard to the bid price. As depicted in Figure 6.10, these curves implicitly model the competition among rival flexibility providers.

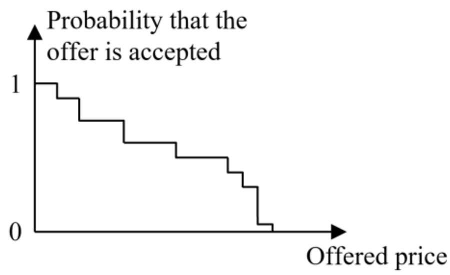


Figure 6.10 – Empirical stepwise curve of offer acceptation probability.

Henceforth, the fixed component can be expressed as follows:

$$\Phi^{\text{fixed}}\left(\mathbf{P}_z,\boldsymbol{\lambda}_z\right)=\sum_{z=1}^{N_z} p_{\eta,z}\left(\boldsymbol{\lambda}_z\right)\mathbf{P}_z \mathrm{T}\boldsymbol{\lambda}_z \tag{6.4}$$

where $\mathbf{P}_z$ and $\boldsymbol{\lambda}_z$ are respectively the power and price proposed for each of the N_Z mid-term contracts (e.g. capacity in the different balancing services and power exchanged in the forward market) and T is the mid-term decision horizon.

The variable component Φ_σ^{var} depends on the VPP policy during the considered period but also on the clearing of mid-term markets since the VPP has to fulfill its concluded contracts. This term is computed using the pre-determined surrogate model. The optimal mid-term profit is thus a trade-off between both fixed and variable contributions (Figure 6.11).

$$\Phi_{tot}\left(\mathbf{z}\right)=\Phi^{\text{fixed}}+\sum_{\sigma=1}^{N_\Sigma} p_\sigma \Phi_\sigma^{\text{var}}\left(\mathbf{z}\right) \tag{6.5}$$

where $\boldsymbol{z}$ is the vector including mid-term decision variables whereas p_σ represent the probabilities of occurrence of the N_Σ mid-term typical days.

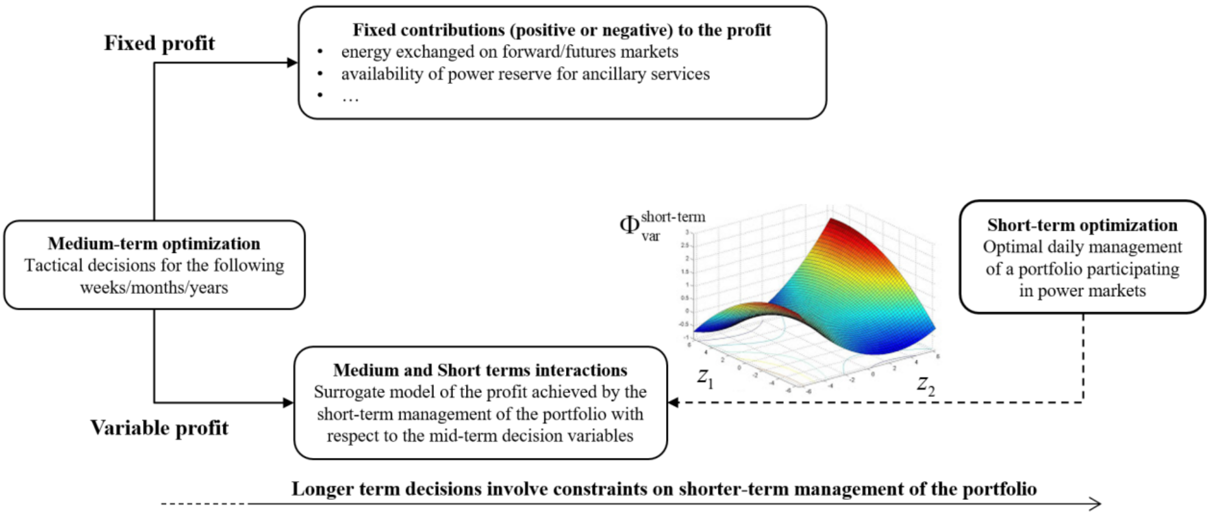


Figure 6.11 – Fixed and variable contributions to the mid-term profit.

The problem consists thus in maximizing the nonlinear multivariate function (6.5) in which there is a mixture between continuous (e.g. offered price for balancing services) and integer (e.g. amount of power in the different balancing products) decision variables. It is solved using a genetic algorithm (GA), which offers the benefit to be free-derivative and is shown to reduce the risk of being involved in a local optimum [Gen00]. Furthermore, in addition to its efficiency in solving complex constrained problems, the heuristic nature of this algorithm is highly valuable for easily adapting the formulation (MILP, MINLP, etc.) to different VPP profiles.

6.6 Case study

The presented methodology is illustrated for the same Belgian electricity retailer than the one presented in Section 4.6. As a reminder, it is composed of 2 medium-sized conventional generation units, a 130 MW slow power plant and a fast unit with an 80 MW capacity. The portfolio is moreover constituted of one PSH station with an output power in both pump and turbine modes of 24 MW (with an energy/power ratio of 5). The VPP has also wind turbines for a total installed power of 20 MW and supply both industrial and residential clients (peak power of 30 MW). Some are equipped with rooftop photovoltaic (PV) units for a total installed power of 5 MW. The electrical energy generated from PV installation distributed among low-voltage networks is treated as a negative load and cannot be curtailed. Finally, the VPP is responsible for optimally operating DR resources from heat pumps (10 MW) and industrial fridges (15 MW).

The results are illustrated for tactical decisions taken by an electricity retailer with a week-ahead perspective. These decisions include the determination of the energy exchanges in futures markets (FM) as well as the quantities and prices of FCR, aFRR and mFRR contracted with the system operator through a pay-as-bid auction (even though the mFRR is currently in month-ahead). The retail price to end-users is considered as constant and is thus not included in the optimization.

The objective is to optimize its week-ahead decisions for a typical week of of July. The results of the implemented tool are then compared with two other approaches in order to highlight the benefits of our methodology.

The short-term operation is characterized by intraday decisions that are made on a 15 minutes basis (96 daily decisions). Hence, when historical data are available with a 1-hour discretization, these data are considered as constant over the 4 constitutive quarters of an hour. Similarly to stochastic models, the short and mid-term optimization procedures are modeled and solved using Matlab.

6.6.1 Stochastic models analysis

The mid-term uncertainty is addressed by defining statistically representative days of wind and solar generation as well as total consumption within the portfolio. In order to capture sufficient information without excessively increasing the size of the decision process, the number of typical days is here fixed to $N_{\Sigma} = 8$. Historical data of hourly wind speed are known for the 5 wind parks and converted into active power using a traditional power curve [Fang[11]]. The data for residential load and their PV generation originates from smart metering (SM) devices installed in some customers. These devices are recording quarter-hourly energy flows that can be used to extrapolate statistical load and PV generation profiles of all clients (even those without SM device) using respectively their yearly electricity metering and PV installed power [Toubeau[16]].

Afterwards, the day-ahead scenarios are generated. The endogenous variables considered in this work are the intraday and imbalance prices (and the liquidity of the Intraday market) as well as the amount of FCR, aFRR and mFRR activated by the system operator. The dimensionality of each variable is first decreased with PCA. The reduced-size variables (i.e. after applying PCA) are determined such as preserving at least 85 % of the information of the original dataset.

The copula model is then used for constructing the scenarios. For limiting the size of the short-term optimization, $N_{\Omega} = 10$ scenarios are here created.

6.6.2 Medium-term optimization tool

The first step consists in constructing the empirical stepwise curve of the acceptation probability of the reserve capacity (Figure 6.8), which was realized by using historical prices (available on the Elia website). The principle is that prices higher than the highest accepted offer have a probability of 0% to be accepted, whereas the lowest price has 100 % of acceptance. The curves are then built by accounting for all offers, and the results for FCR and aFRR are represented in Figure 6.12.

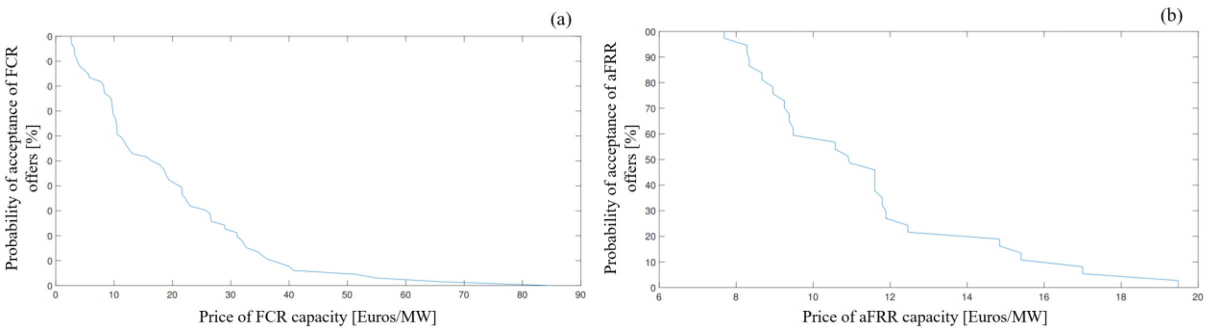


Figure 6.12 – Probability of acceptation of reserve capacity for FCR (a) and aFRR (b).

The second step consists in defining a surrogate model of the short-term portfolio management. A preliminary pragmatic selection of potential models is carried out. For instance, neural networks are discarded since they require too many samples to yield robust results. Likewise, splines are only relevant for two-dimensional problems. Consequently, only rational and Gaussian models are further considered. Concerning Gaussian models, Radial Basis Functions (RBFs) and kriging models present the most interesting features in terms of interpolation capabilities and the study is thus limited to these two approaches. Moreover, it should be emphasized that there is no need to predefine the complexity of rational models since the building algorithm adaptively learns their optimal structure. The different models are then compared and the results for one typical day are shown in Table 6.1. These are focused on the number of samples required to train each model (i.e. determination of its optimal parameters) as well as a post-hoc measure of accuracy. Indeed, in order to obtain an unbiased measure of the generalization capability of the surrogate models, it is important to compute their accuracy on new samples (test set) that are independent from the training data. Here, the performance of the different models is estimated using the coefficient of determination R^2. This coefficient is computed from the sum of squares SS_E of the errors between the experimental data points and the fitting function determined by the regression analysis. Then, this term is normalized by the sum of squares SS_{tot} of the distances between the data points and their mean value. If the model fits the data well, SS_E will be much smaller than SS_{tot} and their ratio will be close to zero. Practically, the R^2 coefficient is calculated as follows:

$$R^2 = 1 - \frac{SS_E}{SS_{tot}} \tag{6.6}$$

The R^2 coefficient is thus included in the [0, 1] interval (assuming that SS_E is always smaller than SS_{tot}) and is equal to one when all the experimental points correspond exactly to those determined by the fitting model. Conversely, if the quality of the model is decreasing, the value of R^2 will drop accordingly. The R^2 value can be interpreted as the proportion of variability for the dependent variables x_i, that can be captured by the mathematical model.

Table 6.1
Comparison of model accuracy on test set.

Model builder	# samples	R² metric
Rational	916	0.91
RBF	NA	NA
Kriging	396	0.90

The RBF model was not able to reach the target accuracy value within the time delay imposed by the learning procedure and no post-hoc analysis was thus performed. Then, one can see that the rational model necessitates an important number of samples to ultimately achieve an accuracy equivalent to the kriging interpolation.

In this work, a kriging model, i.e. interpolation method based on linear regression of surrounding samples, is therefore selected to optimize the parameters of the N_Σ = 8 models.

The quality target was set in such a way that this technique requires simulating N_Θ = 36 points for constructing each of the 8 models. Since the optimal day-ahead bidding strategy necessitates slightly more than 4 minutes, the mean computational time for obtaining a surrogate model for a particular typical day is around 2.5 hours. In order to validate the quality of the different models, R^2 coefficients are computed in post-processing (after determination of the models) with 50 new experimental points randomly generated. The range of the R^2 coefficients obtained for the 8 surrogate models lies between 0.87 and 0.99, which constitutes a good indicator of the performance and reliability of the modeling procedure.

The mid-term optimization is then carried out and the results are provided in Table 6.2. The simulation time is around 1 minute. The quantity range of balancing services are determined by aggregating the capabilities of each portfolio utilities. These capabilities are computed based on the technical requirements (maximum output power and ramping limits) for providing the reserve.

Table 6.2
Operating range of decision variables and their week-ahead optimal values.

	Quantity range	Price range	Optimal quantity	Optimal price
FCR	[0, 6] MW	[18.4, 84.9] €	6 MW	34.6 €
aFRR	[0, 16] MW	[7.7, 19.5] €	12 MW	10.6 €
mFRR	[0, 25] MW	[2.8, 7.5] €	25 MW	2.8 €
FM	[-50, 50] MW	37.5 €	35.8 MW	37.5 €

The results attest the importance of accounting for dependencies between time periods. Indeed, although it seems tempting to reserve as much power reserve as possible in week-ahead for maximizing the fixed revenues, it does not constitute an appropriate strategy for the envisaged VPP. This can be explained by the operational costs of power plants that may be uneconomic with regard to current market prices or by the limited energy of storage utilities that may lead to the impossibility of providing the power requested in real-time. Here, the optimal solution indicates that the portfolio should provide 12 MW of secondary reserve for a maximum technical capacity of 16 MW. However, the maximum amount of primary reserve was selected. This originates from the limited amount of energy required for supplying the reserve. We also notice that the prices for the procurement of balancing services are relatively small (which induces smaller guaranteed fixed revenue) in order to ensure their economic competitiveness in the selection procedure (i.e. pay-as-bid auction).

6.6.3 Comparison of different approaches

Within the objective to illustrate the added value of our surrogate-based optimization approach (method #1), the optimal outcome is compared with the profit that would have been realized with two other approaches. The first one (method #2) consists in a global three-stage stochastic optimization. In order to compare both methods on a sound basis, the formulations are adapted to ensure that both simulations necessitate the same space requirements (16 Go RAM). As expected, the resulting formulation of the three-stage program is far too large to accommodate the same number of scenarios than our methodology. Indeed, the number of scenarios is then equal to $N_{\Sigma} N_{\Omega}$ = 80, which would necessitate more than 15 times the memory allocated for the task. Hence, the number of mid-term typical days (i.e. first-stage stochastic variables) has to be reduced from 8 down to 4 along with the scenarios of endogenous variables (i.e. second-stage stochastic variables) from 10 down to 5. Some technical constraints such as ramping rates of units as well as minimum up/down times are also neglected. Moreover, the nonlinearities are relaxed by imposing a fixed reservation price in the reserve capacity market. We have therefore opted for a risk-averse choice that secures the acceptation of the offers.

Finally, an intuitive but non-optimized perspective (method #3) for mid-term decisions is considered. The portfolio manager, with the principle of due diligence, decides to participate to balancing services in accordance with its full technical capacity with a risk-averse bidding strategy and to sell its expected base generation (i.e. total generation minus total load) in futures market.

The mid-term decisions corresponding to the three methods are given in Table 6.3.

Table 6.3
Decisions taken by the different mid-term strategies.

Methods	R1	R2	R3	FM
#1	6 MW	12 MW	25 MW	-37.5 MW
#2	6 MW	19 MW	30 MW	-50MW
#3	6 MW	20 MW	30 MW	-55 MW

From Table 6.3, it can be observed that the simplifications that were mandatory to make method #2 tractable (i.e. disregarding extreme scenarios or operational inter-temporal constraints) lead to overly optimistic mid-term decisions. Indeed, method #1 leads to the most conservative mid-term decisions, the others strategies seeking instead to maximize the fixed revenues received for providing capacity for balancing reserves.

The short-term optimization tool described in Section 4.4 is then used to compare the short-term portfolio management with regard to these mid-term decisions. The results are presented in Table 6.4. Those include the number of mid-term contract violations (non-provision of the requested balancing reserve) due to operational constraints as well as the expected (fixed and variable) profits over the whole period.

Table 6.4
Comparison of different mid-term strategies.

Methods	# weekly contract violations	Expected variable profit	Expected fixed profit	Expected total profit
#1	0	4.8 10^6 €	1.9 10^5 €	5 10^6 €
#2	15	4.5 10^6 €	2.5 10^5 €	4.8 10^6 €
#3	58	4.1 10^6 €	2.6 10^5 €	4.4 10^6 €

Over the period of interest, the proposed formulation brings an additional profit of respectively $2*10^5$ € and $6*10^5$ € compared to methods #2 and #3 (Table 6.4). This can be explained by the fact that these methods induce violations of longer-term contracts due to limited capacity of some flexibility sources (storage utilities). Such situations have to be absolutely avoided since it incurs severe financial penalties. Furthermore, CPP units providing spinning reserve have to be online, which may be unprofitable in the long run if periods with low prices are too frequent.

Hence, the proposed mid-term strategy, in addition to secure feasible solutions in real-time, is also conferring a substantial extra benefit over the considered week compared to the three-stage and logic-based approaches.

It should be mentioned that the expected short-term profits have to be put into perspective with the important installed power of renewable sources within the studied portfolio that generates substantial incomes by selling energy in sport markets. Indeed, these generators are characterized by very low operational costs. Additionally, the depreciation of the installation is not included in the mid-term optimization since it constitutes a fixed contribution that do not influence the decision process.

6.7 Conclusions and perspectives

This chapter investigates the added value of simulating a detailed short-term operation of the portfolio in the context of mid-term decision procedure in order to reach a global optimum over the studied period. The principle is to couple the time horizons by defining an adequate surrogate model that fulfills therefore two objectives. It indeed allows to account for time-dependent constraints while being computationally efficient in the context of a mid-term decision process. The economic benefit of the procedure has been emphasized for a typical Belgian retailer disposing of its own flexible plants. A major advantage of the proposed approach is that the decision tool implemented for the short-term operation can be used in stand-alone in the context of a day-ahead bidding strategy.

Moreover, within the objective of reducing simulation time, a new method for generating scenarios in the context of two-stage stochastic programming was developed and yields satisfying results. In this regard, an interesting perspective of this work is to include catastrophic scenarios characterizing the potential loss of portfolio components. Such events can indeed significantly influence the longer term decisions for hedging against the resulting losses. Furthermore, including flexibility of end-users installed in distribution systems is also likely to improve the portfolio management.

6.8 Chapter publications

This chapter has led to the following publications:

- J. F. Toubeau, Z. De Grève and F. Vallée, "Medium-Term Multimarket Optimization for Virtual Power Plants: a Stochastic-Based Decision Environment," in *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1399-1410, March 2018.

- M. Gobert, F. Vallée, J. Gmys, N. Melab, J.-F. Toubeau, D. Tuyttens, “Surrogate-Assisted Optimization for Multi-stage Optimal Scheduling of Virtual Power Plants,” in PaCOS - International Workshop on the Synergy of Parallel Computing, Optimization and Simulation (part of HPCS 2019), Dublin (Ireland), Jul 2019.

- J.-F. Toubeau, M. Hupez, V. Klonari, Z. De Grève and F. Vallée, "Statistical Load and Generation Modelling for Long Term Studies of Low Voltage Networks in Presence of Sparse Smart Metering Data", Proc. of the 42nd Annual Conference of IEEE Industrial Electronics Society (IECON), Florence (Italy), Oct. 2016.

CHAPTER 7

CONCLUSIONS AND PERSPECTIVES

Our society is currently undergoing an important **energy transition** with as main objective to rely on a low-carbon (and nuclear-free) energy system. In Europe, it progressively results in the electrification of both transportation and heating sectors, with an electrical system that mainly depends on renewable-based generation.

In parallel, the **European liberalization of the electricity sector** that occurred in the nineties has introduced competition at both production and supply levels, whereas the transmission and distribution of electricity remained regulated monopolies[31]. However, due to the specificities of the electrical grid that has to be continuously balanced (generation must be equal to consumption at any time) and whose power flows cannot be easily controlled, it resulted in a complex environment involving many agents with potentially conflicting objectives. Moreover, power systems are subject to an increased share of uncertainties that become more complex to understand (due to intricate dependencies between electrical and market data), and more dynamic.

In light of these challenges, our work was focused on three main complementary questions with the objective of addressing the decision-making problem faced by market players. In particular, we firstly put our research interest on improving the quality of probabilistic forecasts so as to reduce the uncertainty space that has to be processed in the subsequent optimization tools. We then focused on the modeling framework of these decision tools. In particular, we have developed a new formulation that is able to adequately valorize (i.e. extract the full economic potential) emerging technologies that are typically characterized by strong energy limitations. Specifically, this work has placed emphasis on underground pumped-hydro storage stations, new environmental-friendly solutions that are associated with important dynamic and nonlinear constraints. Thirdly, we focused on how we can efficiently couple the considerations from different time horizons with conflicting objectives so as to achieve the best planning policy over a longer term perspective.

[31] The objective of the liberalization was to foster the energy transition through expected smaller electrical energy costs, and stimulated investment in renewable generation.

In this final chapter, we firstly summarize the developments and findings presented throughout the dissertation, with a particular focus on the salient contributions and conclusions. As a second stage, we formulate some suggestions for future research.

7.1 Concluding remarks

In the current context of power systems (increase of uncertainties regarding both load and generation in a system that needs to keep the continuous energy balance while ensuring the safe and efficient operation in each part of the network), both long-term and operational planning stages are more and more affected by forecasting errors. In order to properly consider these uncertainties (to properly feed and guide subsequent optimization tools), different modeling techniques (to properly characterize mid and long-term uncertainties) and forecasting tools (typically for day-ahead to close to real-time considerations) therefore need to be developed. Even though reaching oracle tools (able to perfectly predict the future) do not seem realistic, there is currently still room for improvement.

Regarding day-ahead predictions, Deep Learning techniques, and, in particular, advanced recurrent neural networks such as Long Short Term Memory (LSTM) networks are already able to significantly improve the forecast accuracy of electrical quantities compared to traditional approaches. This is achieved thanks to the ability of such networks to automatically select and propagate through time the most relevant information for the prediction purpose.

In this work, the question of combining the modeling power of deep neural networks with advanced memory structures able to control the flow of past relevant information is investigated in Chapter 3. In particular, the objective is to take advantage of the specificities of the day-ahead operational planning to design prediction tools with tailored architectural variations that improve their performance. Practically, since the predictions are needed simultaneously for each time step of the scheduling horizon, LSTM networks are combined with a bidirectional processing of data.

Although the method can provide prediction intervals and densities, it is here extended with the aim to provide predictive scenarios. Practically, the tool relies on a copula-based sampling of the multivariate forecasted distribution so as to generate time trajectories that mimic actual time and cross-variable dependencies. This work has focused on the aggregated demand and renewable generation (which suits the market player perspective).

The results demonstrated that **the proposed methodology yields accurate, calibrated forecast distributions learned from the historical dataset, and that the generated scenarios enable to increase the economic profit of energy aggregators participating in electricity markets**.

However, even if the actors of the electricity sector can rely on more reliable forecasts, the situation is such that there is an increasing need of flexibility (i.e. ability to adjust generation/consumption within a short time frame based on an external signal) in power systems. This flexibility can be efficiently provided by pumped storage hydro (PSH) units, and new solutions such as underground stations are consequently emerging. Specifically, this thesis was conducted in the framework of the ***Smartwater*** project (funded by the Walloon Region), which aims at evaluating the feasibility of the rehabilitation of end-of-life quarries and mines into small to medium-sized PSH stations connected to the distribution grid.

In this dissertation, it is assumed that the storage units are operated by market actors who are considering all opportunities to maximize their profit. It should be noted that, if the market is properly designed, the decisions that maximizes the revenues of the market participants are those that help the most the system (e.g. the real-time energy imbalances in the Elia system have significantly decreased since the introduction of the single pricing scheme for the imbalance settlement).

The operation of small PSH units are significantly constrained by their limited energy capacity (typically 4 to 6 hours at full power to completely exhaust the storage capacity), and their economic potential is thus fully leveraged when included within a larger portfolio, i.e. aggregation of assets centrally optimized in order to maximize the total profit. In chapter 4, a methodology for the day-ahead scheduling of an electrical portfolio is built with the aim of incorporating any type of electricity generation, consumption or source of flexibility (storage technologies, demand response, etc.). By considering all market opportunities, i.e. both energy markets (day-ahead, intraday, imbalance settlement) and ancillary services (including spinning and non-spinning reserves), the proposed scenario-based formulation targets the cost-optimal allocation of assets (so as to maximize the total revenues of market participants).

With the proposed methodology, it has been confirmed that **the portfolio effect (aggregation of technologies) results in a more efficient use of assets due to complementarities between the different technologies**. In this respect, it has been shown that a dynamic allocation of reserves (i.e. when the contribution of each unit can vary over time) fosters the participation in ancillary services, which results in higher economic value of the global portfolio. Furthermore, it is observed that **neglecting the real-time activation of operating reserves can lead to conservative solutions that do not fully exploit the potential of available resources**.

Then, a particular focus is given in chapter 5 to the operation of these small to medium-sized units is governed by multiple nonlinearities arising from the geometry of the basins, the head effects (forbidden zones and performance curves in turbine and pump modes) as well as the groundwater exchanges between reservoirs and their surrounding aquifers.

The work was firstly devoted to understand the specificities of underground pump-storage hydro stations, and the resulting limitations of strategies that are currently used for the short-term scheduling of traditional facilities. However, even convexification or linearization procedures are intrinsically onerous computationally, which prevents them to consider the proper dynamics of the system (with an optimization step lower than 1 minute). Moreover, these techniques are associated with modeling approximations that may lead to infeasible solutions.

There is therefore a need of new tailored decision tools, and a hybrid procedure is here developed for tackling this complex and realistic problem in a reasonable amount of time. The procedure consists in the sequential operation of an optimization tool and a simulation model, both included into a control loop ensuring the convergence towards a feasible and optimal solution. The simulation model enables to faithfully represent the complex behaviors of underground PSH plants as well as their associated dynamics. Indeed, the proposed method takes as inputs realistic models coming from partners specialized in electro-mechanics (for operation of hydraulic and electrical machines) and hydro-geology (for water exchanges between reservoirs and surrounding aquifers). This ensures the practical feasibility of the scheduling, while improving the economic valorization of the flexibility, thereby compensating the lack of global optimality guarantee of the procedure.

It should be mentioned that the **grids fees and taxes associated with both consumption and generation modes of PSH stations had been neglected because of the current Belgian context in which these contributions are so important that the optimal solution for the storage is to stay offline during the whole day** (no opportunity for making profit even if the installation costs of the units are negligible).

Then, the results demonstrate that accurately **considering these nonlinear effects is essential to extract the full economic potential of such underground stations**, provided that the solution lies near to the global optimum. It has also been noticed that considering forbidden zones significantly increase the simulation time, and the implementation advanced modeling techniques to alleviate the burden of these constraints can be highly valuable in the future.

Finally, with the current organization of electricity markets (in which the procurement of reserve capacity is carried out in week-ahead), an important care has to be given to the mid-term perspective. Indeed, an important way of valorizing flexibility originates from the participation in ancillary services. At this stage, the challenge is **how to efficiently handle medium-term multi-stage (potentially stochastic and nonlinear) decision procedures in a computationally efficient way**. To that end, a surrogate modeling approach is presented in chapter 6 (consisting in learning as a pre-processing task the relationship between the first-stage decision variables and their impact on the objective function with respect to the optimal operation of the latter decision stages), and offers an elegant and promising solution.

The outcomes shows that the proposed surrogate-based mid-term strategy, in addition to secure feasible solutions in real-time, is also conferring a substantial extra benefit over the considered month compared to the more traditional approaches.

7.2 Perspectives for future research

The forecasting tools developed in this thesis have all the characteristics to be successfully applied to very short-term application (e.g. for the subsequent optimal control of wind turbines), provided that some architectural adjustments are made to fit the specificities of the task (by taking into account the strong relationship with past realizations). In the same vein, it will be interesting to see if such models have the potential to break down the barrier of achieving acceptable accuracy levels for longer term prediction horizons such as a week-ahead perspective.

Moreover, whereas aggregated predictions are of interest for market players eager to efficiently operate their portfolio, the system operator can be more interested in local contributions so as to be able to perform adequate load-flow studies (determining the active and reactive power flows in each line of an interconnected system as well as voltage levels at each node) to guarantee the stable operation of its network. In this context, accounting for space dependencies is crucial. Indeed, complex correlation patterns exist between renewable productions/demands located in different network areas. For instance, a link exists between solar productions in a given geographical zone, which can also be true for wind productions depending on the landscape configuration. The demand patterns, beyond the correlation between end-users with similar consumption habits, may also be influenced by the solar production, in order to increase the self-consumption of households/districts/microgrids, depending on demand response policies that have been established by the local authorities.

Generally, **handling and exploiting the high dimensionality which arises from the wide geographical span of the grids, as well as from the heterogeneous data sources, is a key challenge in data analytics for modern electricity grids.**

In general, the optimization tools developed in this thesis enable to estimate the economic potential of PSH units (or any other technologies or aggregation of assets) in the short and medium term perspectives. Such information can also useful for longer-term studies to estimate the period before return on investment (if any).

Accounting for such time dependencies in a multi-stage procedure can be beneficial for complex tasks such as the operational planning of transmission grids consisting in undertaking preventive (change of network topology, adjust the tap changer position of a transformer, etc.) as well as corrective (P-Q control, etc.) actions at a minimal cost in order to guarantee a continuous, secure and reliable electricity supply. This decision-making process involves different embedded time horizons, going from week-ahead towards real-time that need to be jointly considered.

The implemented tools can also be used by regulators in order to assess the profitability of different kind of portfolios (composed only of PSH units or mainly based on renewable generation, etc.) with respect to future evolution of the regulatory framework (removal of grid fees for energy offtakes for storage devices, removal of financial subsidies to renewable generation, etc.). They can in this way determine if the market design in well aligned with their targets (carbon-free energy mix that ensure the stable and efficient operation of the electrical system) before being put into operation.

In this way, the exploitation of the tools has highlighted that, in the current Belgian regulatory framework, underground pump-storage units are completely unprofitable. Due to the grid fees and taxes associated with both energy injections and offtakes, such units cannot even generate operational profits, and it is therefore even less possible to recover investment costs. In the absence of such costs, however, the outcomes have shown that underground PSH stations can offer valuable services in an economic way.

Moreover, the procurement of balancing services is expected to move towards a day-ahead sourcing in order to leverage all flexible resources distributed throughout the system (technology-neutral market). In this context, only the scheduling horizon is impacted as the sequentiality of decisions of market players remains unaltered (procurement of balancing capacity – day-ahead market – procurement of balancing energy). The complexity of the problem keeps its three-stage stochastic structure, which lies at the frontier of game theory (due to the “pay-as-bid” system where the market player is remunerated at its offered price), resulting in a nonlinear formulation. The surrogate-based optimization tool developed with a mid-term perspective can then be transposed in case of short-term procurement of balancing services (with the same benefit of alleviating the computational burden of the decision procedure). The approach can nonetheless still be improved regarding the temporal dependencies. Indeed, in the proposed approach, the mid-term uncertainty is represented by representative typical days since the mid-term horizon is actually made up of different days characterized by the same decision process (which allows to decouple/decompose the problem). However, in such an approach, the transition between consecutive days is neglected (each one being optimized independently of the others), which necessitates to define assumptions about the operation of energy-dependent technologies (e.g. in this work, it was imposed that the energy stored in pumped-storage hydro units is the same at the start and at the end of the day), potentially resulting in suboptimal solutions.

Finally, by working extensively on the short-term operational planning, we noticed that the intraday scheduling, which defines the actual positions of the unit following the day-ahead decisions, is relatively neglected in the literature. However, this temporal horizon is important since the objective is not only to ensure adherence to the pre-defined bidding strategy in previous horizons (regarding both electrical energy and ancillary services), but also to take advantage of (predictable) system imbalances

REFERENCES

[Abreu12] L. V. L. Abreu, M. E. Khodayar, M. Shahidehpour and L. Wu, “Risk-Constrained Coordination of Cascaded Hydro Units with Variable Wind Power Generation,” *IEEE Trans. Sust. Energy,* vol. 3, no. 3, pp. 359–368, Jul. 2012.

[Akaike59] H. Akaike, “On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method,” Ann. Inst. Statist. Math. Tokyo, vol. 11, pp. 1-17, 1959.

[Al-Awami11] A. T. Al-Awami, M. A. El-Sharkawi, “Coordinated trading of wind and thermal energy,” *IEEE Trans. Sust. Energy,* vol. 2, no. 3, pp. 277–287, July 2011.

[Alvarado15] R. Alvarado, A. Niemann, T. Wortberg, “Underground pumped-storage hydroelectricity using existing coal mining infrastructure,” in E-proceedings of the 36th IAHR world congress, The Hague, 2015.

[Alvarez18] G. E. Alvarez, M.G. Marcovecchio, P.A. Aguirre, “Security-Constrained Unit Commitment Problem including thermal and pumped storage units: An MILP formulation by the application of linear approximations techniques,” in *Electric Power Systems Research,* vol. 154, pp. 67-74, Jan. 2018.

[Amin Tajeddini14] M. Amin Tajeddini, A. Rahimi-Kian and A. Soroudi, “Risk averse optimal operation of a virtual power plant using two stage stochastic programming,” *Energy*, vol. 73, pp. 958-967, Aug. 2014.

[Archer07] C. Archer, M. Jacobson, “Supplying baseload power and reducing transmission requirements by interconnecting wind farms,” *Journal of applied Meteorology & Climatology*, 46(11), 2007.

[Ardizzon14] G. Ardizzon, G. Cavazzini, and G. Pavesi, “A new generation of small hydro and pumped-hydro power plants: advances and future challenges,” Renewable and Sustainable Energy Reviews, vol. 31, pp. 746–761, 2014.

[Argonne13] Argonne national laboratory, “Modeling ternary pumped storage units,” 2013.

[Artiba98] A. Artiba, V.V. Emelyanov, S.I. Iassinovski, “Introduction to Intelligent Simulation: The RAO language,” Operation Research & Decision Theory, Springer, 1998.

[Ausubel10] L.M. Ausubel, P. Cramton, “Using forward markets to improve electricity market design”, Utilities Policy, vol. 18, pp. 195-200, 2010.

[Baringo13]L. Baringo and A. J. Conejo, "Strategic Offering for a Wind Power Producer," in *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4645-4654, Nov. 2013.

[Barot08] H. Barot, and K. Bhattacharya, “Security Coordinated Maintenance Scheduling in Deregulation Based on Genco Contribution to Unserved Energy,” *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1871–1882, Nov. 2008.

[Barzilai88] J. Barzilai and J. M. Borwein, “Two-Point Step Size Gradient Methods,” IMA Journal of Numerical Analysis, vol. 8, pp. 141-148, 1988.

[Baslis09] C.G. Baslis, S.E. Papadakis and A.G. Bakirtzis, "Simulation of Optimal Medium-Term Hydro-Thermal System Operation by Grid Computing," *IEEE Trans. Power Syst.*, vol.24, no.3, pp.1208-1217, Aug. 2009.

[Baslis11] C. G. Baslis and A. G. Bakirtzis, “Mid-Term Stochastic Scheduling of a Price-Maker Hydro Producer With Pumped Storage,” *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 1856–1865, Nov. 2011

[Bathurst02] G. N. Bathurst, J.Weatherill, and G. Strbac, “Trading wind generation in short term energy markets,” *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 782–789, Aug. 2002.

[Bedford01] T. Bedford, and R. M. Cooke, “Probability density decomposition for conditionally dependent random variables modeled by vines,” Annals of Mathematics and Artificial Intelligence 32, pp. 245–268, 2001.

[Belpex15] Belpex, “Indirect Participation on the Belpex Spot Market”, 2015.

[Belpex18] Belpex, “Products specifications”, webpage https://www.belpex.be/trading/product-specification/ last consulted the 1st March 2018.

[Bengio15] S. Bengio, V. Oriol, N. Jaitly, and N. Shazeer. “Scheduled sampling for sequence prediction with recurrent neural networks.” In Advances in Neural Information Processing Systems, pp. 1171-1179, 2015.

[Beringer04] N. Beringer, “Human language acquisition in a machine learning task,” Proc. ICSLP, 2004.

[Bian15] Q. Bian, H. Xin, Z. Wang, D. Gan and K. P. Wong, "Distributionally Robust Solution to the Reserve Scheduling Problem With Partial Information of Wind Power," in *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2822-2823, Sept. 2015.

[Birge97] J. R. Birge and F. Louveaux, “Introduction to stochastic programming,” Springer-Verlag, New York, 1997.

[Bixby00] R. Bixby, M. Fenelon, Z. Gu, E. Rothberg, and R. Wunderling, “MIP: theory and practice– closing the gap,” in *System Modelling and Optimization: Methods, Theory and*

Applications, M. J. D. Powell and S. Scholtes, Eds. Boston: Kluwer Academic Publishers, vol. 174, pp. 19–49, 2000.

[Bixby[07]] R. Bixby and E. Rothberg, "Progress in computational mixed integer programming—A look back from the other side of the tipping point," *Annals of Operations Research*, vol. 149, no. 1, pp. 37–41, Jan. 2007.

[Borghetti[08]] A. Borghetti, C. D'Ambrosio, A. Lodi and S. Martello, "An MILP Approach for Short-Term Hydro Scheduling and Unit Commitment With Head-Dependent Reservoir," in *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1115-1124, Aug. 2008.

[Botterud[12]] A. Botterud *et al.*, "Wind Power Trading Under Uncertainty in LMP Markets," in *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 894-903, May 2012.

[Brechmann[13]] E. C. Brechmann and U. Schepsmeier, "Modeling Dependence with C- and D-Vine Copulas: The R Packatge CDVine," *J. Stat. Softw.*, vol. 52, no. 3, pp. 1–27, 2013.

[Bremnes[04]] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," Wind Energy, vol. 7, no. 1, pp. 47–54, 2004.

[Brijs[17]] T. Brijs, "Electricity storage participation and modeling in short-term electricity markets," PhD dissertation, KU Leuven, 2017.

[Broomhead[88]] D. Broomhead and D. Lowe, "Multivariate Functional Interpolation and Adaptive Networks," Complex Systems, vol. 2, pp. 321–355, 1988.

[Bruninx[16]] K. Bruninx, "Improved modeling of unit commitment decisions under uncertainty," PhD dissertation, KU Leuven, 2016.

[Cabero[05]] J. Cabero, A. Baillo, S. Cerisola, M. Ventosa, A. Garcia-Alcalde, F. Peran and G. Relano, "A Medium-Term Integrated Risk Management Model for a Hydrothermal Generation Company," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1379–1388, Aug. 2005.

[Carrion[06]] M. Carrion and J. M. Arroyo, "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1371–1378, May 2006.

[Castronuovo[04]] E. D. Castronuovo and J. A. P. Lopes, "On the optimization of the daily operation of wind-hydro power plant," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1599–1606, Aug. 2004.

[Catalao[09]] J. P. S. Catalao, S. J. P. S. Mariano, V. M. F. Mendes and L. A. F. M. Ferreira, "Scheduling of Head-Sensitive Cascaded Hydro Systems: A Nonlinear Approach," in *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 337-346, Feb. 2009.

[Chang[01]] G. W. Chang, M. Aganagic, J. G. Waight, J. Medina, T. Burton, S. Reeves, and M. Christoforidis, "Experiences with mixed integer linear programming based approaches on short-term hydro scheduling," *IEEE Trans. Power Syst.*, vol. 16, pp. 743–749, Nov. 2001.

[Chazarra[17]] M. Chazarra, J. I. Pérez-Díaz and J. García-González, "Optimal Energy and Reserve Scheduling of Pumped-Storage Power Plants Considering Hydraulic Short-Circuit Operation," in *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 344-353, Jan. 2017.

[Chazarra[18]] M. Chazarra, J. I. Pérez-Díaz and J. García-González, "Optimal Joint Energy and Secondary Regulation Reserve Hourly Scheduling of Variable Speed Pumped Storage Hydropower Plants," in *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 103-115, Jan. 2018.

[Chen[17]] C. H. Chen, N. Chen and P. B. Luh, "Head Dependence of Pump-Storage-Unit Model Applied to Generation Scheduling," in *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2869-2877, July 2017.

[Chollete[09]] L. Chollete, A. Heinen and A. Valdesogo, "Modeling international financial returns with multivariate regime switching copula", Journal of Financial Econometrics 7, pp. 437–480, 2009.

[Chung[15]] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Gated Feedback Recurrent Neural Networks," in Proceedings of the 32nd ICML, pp. 2067–2075, Lille, France, 2015.
[Clastres[10]] C. Clastres, T. T. Ha Pham, F. Wurtz, and S. Bacha, "Ancillary services and optimal household energy management with PV production," *Energy*, vol. 35, no. 1, pp. 55–64, Jan. 2010.

[CEW[17]] Clean Energy Wire, "Germany's greenhouse gas emissions and climate targets," factsheet climate and CO2, 2017.

[Conejo[02,a]] A. J. Conejo, J. M. Arroyo, J. Contreras and F. A. Villamor, "Self-scheduling of a hydro producer in a pool-based electricity market," in *IEEE Trans. Power Syst.*, vol. 17, no. 4, pp. 1265-1272, Nov. 2002.

[Conejo[02,b]] A. J. Conejo, F. J. Nogales and J. M. Arroyo, "Price-taker bidding strategy under price uncertainty," in *IEEE Trans. Power Syst.*, vol. 17, no. 4, pp. 1081-1088, Nov 2002.

[Conejo[10]] A. J. Conejo, M. Carrion, and J. M. Morales, "Decision Making Under Uncertainty in Electricity Markets," in International Series in Operations Research & Management Science. New York: Springer, 2010.

[Connolly[11]] D. Connolly, H. Lund, P. Finn, B.V. Mathiesen, M. Leahy, "Practical operation strategies for pumped hydroelectric energy storage (PHES) utilizing electricity price arbitrage," Energy Policy, vol. 39, no. 7, pp. 4189–4196, 2011.

[Couckuyt[13]] I. Couckuyt, D. Gorissen, K. Crombecq, D. Deschrijver and T. Dhaene, "The SUMO toolbox: A tool for automatic regression modeling and active learning," *2013 Africon*, Pointe-Aux-Piments, pp. 1-4, 2013.

[Dashti[16]] H. Dashti, A. J. Conejo, R. Jiang and Jianhui Wang, "Weekly Two-Stage Robust Generation Scheduling for Hydrothermal Power Systems," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4554-4564, Nov. 2016.

[Deane[13]] J. P. Deane, E. J. McKeogh and B. P. O. Gallachoir, "Derivation of Intertemporal Targets for Large Pumped Hydro Energy Storage With Stochastic Optimization," in *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2147-2155, Aug. 2013.

[Deb[01]] K. Deb, “Multi-Objective Optimization using Evolutionary Algorithms,” John Wiley and Sons, New York, 2001.

[De Jong[75]] K. A. De Jong, “The Analysis of the Behavior of a Class of Genetic Adaptive System,” *PhD thesis*, University of Michigan, 1975.

[De Ladurantaye[07]] D. De Ladurantaye, M. Gendreau, J. Potvin, “Strategic bidding for price-taker hydroelectricity producers,” *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2187–2203, 2007.

[Ding[12]] H. Ding, Z. Hu and Y. Song, “Stochastic optimization of the daily operation of wind farm and pumped-hydro-storage plant,” *Renewable Energy*, vol. 48, pp. 571-578, 2012.

[Diniz[08]] A. L. Diniz and M. E. P. Maceira, "A Four-Dimensional Model of Hydro Generation for the Short-Term Hydrothermal Dispatch Problem Considering Head and Spillage Effects," in *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1298-1308, Aug. 2008.

[Dupacová[03]] J. Dupacová, N. Gröwe-Kuska, and W. Römisch, “Scenario reduction in stochastic programming: An approach using probability metrics,” Math. Program. Series A, vol. 3, pp. 493–511, 2003.

[ECR[15]] European Commission Regulation, "Electricity Market Regulators, EU countries and stakeholders discuss EU Internal Electricity Market," 2015.

[ECR[17]] European Commission Regulation, “Establishing a guideline on electricity balancing,” 2017.

[ECR[18]] European Commission Regulation, “Europe 2020 targets: statistics and indicators for Belgium,” 2018.

[EEX[15]] EEX, “Phelix Power Futures,” 2015.

[EIA[18]] EIA U.S. Energy Information Administration, “How much carbon dioxide is produced when different fuels are burned?” 2018.

[Elia[13]] Elia, “Balancing, ULG ELEC0018-1 Marché de l’énergie,” 2013.

[Elia[14]] Elia, “The strategic reserve – A mechanism to cover structural shortages in generation,” 2014.

[Elia[15]] Elia, “Implementation of day-ahead flow-based market coupling in the CWE region,” 2015.

[Elia[17,a]] Elia, webpage http://www.elia.be/en/grid-data/power-generation/generating-facilities last consulted the 28st December 2017.

[Elia17,b] Elia, "Electricity scenarios for Belgium towards 2050: Elia's quantified study on the energy transition in 2030 and 2040," November 2017.

[Elia18] Elia, webpage http://www.elia.be/en/grid-data/electrical-losses-fed-transm-system last consulted the 1st March 2018.

[Elman90] J. L. Elman, "Finding Structure in Time," Cognitive Science, vol. 14, pp. 179–211, 1990.

[ENTSOE15] ENTSOE, "Survey on ancillary services procurement, balancing market design 2014," January 2015.

[ENTSOE18] ENTSOE website, last consulted 3th March 2018.

[EPRS14] European Parliamentary Research Service, "EU climate and energy policies post-2020: Energy security, competitiveness and decarbonisation," 24 March 2014.

[EPRS16] European Parliamentary Research Service, "Understanding Electricity Markets in the EU," November 2016.

[EUCO14] European Council, "2030 climate and energy policy framework," Brussels, 24 October 2014.

[Exizidis14] L. Exizidis, Z. De Grève, F. Vallée and J. Lobry, "Construction and Reduction of Aggregated Scenarios Trees for the Day-Ahead Prediction of Wind Speed," in IEEE Benelux Young Researcher Symposium, April 2014.

[Fang11] L-B. Fang and J-D. Cai, "Reliability assessment of microgrid using sequential Monte Carlo simulation", Journal of Electronic Science & Technology, vol. 9, no. 1, 31–34, March 2011.

[Felder10] M. Felder, A. Kaifel and A. Graves, "Wind Power Prediction Using Mixture Density Recurrent Neural Networks," European Wind Energy Conference and Exhibition, April 2010.

[Flunkert17] V. Flunkert, D. Salinas, and J. Gasthaus, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," arXiv preprint arXiv:1704.04110, 2017.

[FROnT18] Fair RHC Options and Trade, "Heating and cooling in figures," webpage http://www.front-rhc.eu/about/countries/europe/ last consulted the 26th March 2018.

[García-González08] J. García-González, R. M. R. de la Muela, L. M. Santos, and A. Mateo González, "Stochastic joint optimization of wind generation and pumped-storage units in an electricity market," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 460–468, May 2008.

[Gen00] M. Gen and R. Cheng, "Genetic Algorithm and Engineering Optimization," New York: Wiley, 2000.

[Gers02] F. Gers, N. Schraudolph, and J. Schmidhuber, "Learning Precise Timing with LSTM Recurrent Networks," Journal of Machine Learning Research, vol. 3, pp. 115– 143, 2002.

[Glasserman04] P. Glasserman, “Monte Carlo methods in financial engineering,” Springer, Vol. 53, 2004.

[Golberg89] D. E. Golberg, « Genetic Algorithms in search, optimization and machine learning », Ready, MA: Addison-Wesley, 1989.

[Golberg91] D. E. Golberg and K. Deb, « A Comparison of Selection Schemes Used in Genetic Algorithms », *Foundations of Genetic Algorithms 1*, FOGA-1, 1991.

[Gorissen10] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene and K. Crombecq, “A surrogate modelling and adaptive sampling toolbox for computer based design,” Journal of Machine Learning Research, vol. 11, pp. 2051-2055, July 2010.

[Graves05,a] A. Graves and J. Schmidhuber, “Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures,” Neural Networks, Vol. 18, no. 5-6, pp. 602-610, 2005.

[Graves05,b] A. Graves, N. Beringer, and J. Schmidhuber, “Rapid retraining on speech data with LSTM recurrent networks,” Technical Report IDSIA-09-05, 2005.

[Graves12] A. Graves, “Supervised Sequence Labelling with Recurrent Neural Networks,” Textbook, Studies in Computational Intelligence, Springer, 2012.

[Graves13] A. Graves, N. Jaitly and A. Mohamed, “Hybrid Speech Recognition with Deep Bidirectional LSTM**,**” ASRU, Olomouc, Czech Republic, 2013.

[Graves16] A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwinska, ..., D. Hassabis, “Hybrid computing using a neural network with dynamic external memory**,**” Nature, 2016.

[Gröwe-Kuska03] N. Gröwe-Kuska, H. Heitsch, and W. Römisch, “Scenario reduction and scenario tree construction for power management problems,” in Proc. IEEE Power Tech Conf., Bologna, Italy, Jun. 2003, vol. 3, pp. 23–26.

[Guan99] X. Guan, A. Svoboda, and C.-A. Li, “Scheduling hydro power systems with restricted operating zones and discharge ramping constraints,” *IEEE Trans. Power Syst.*, vol. 14, pp. 126–131, Feb. 1999.

[Habibollahzadeh86] H. Habibollahzadeh and J. A. Bubenko, “Applications of decomposition techniques to short-term operation planning of hydro thermal power systems,” *IEEE Trans. Power Syst.*, vol. PWRS-1, pp. 41–47, Feb. 1986.

[Hatami01] A. Hatami, H. Seifi and M. Kazem Sheikh-El-Eslami, “A Stochastic-Based Decision-Making Framework for an Electricity Retailer: Time-of-Use Pricing and Electricity Portfolio Optimization,” *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 1808–1816, Nov. 2001.

[Hellmers16] A. Hellmers, M. Zugno, A. Skajaa and J. M. Morales, "Operational Strategies for a Portfolio of Wind Farms and CHP Plants in a Two-Price Balancing Market," in *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2182-2191, May 2016.

[Helseth16] A. Helseth, M. Fodstad, B. Mo, "Optimal Medium-Term Hydropower Scheduling Considering Energy and Reserve Capacity Markets," *IEEE Trans. Sust. Energy*, vol. 7, no. 3, pp. 934–942, July 2016.

[Herrera98] F. Herrera, M. Lozano and J. L. Verdegay, "Tackling real-coded genetic algorithms: Operators and Tools for Behavioural Analysis", Artificial Intelligence Review, Vol. 12, No. 4, pp. 265-319, 1998.

[Hinton06] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006.

[Hochreiter97] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[Hochreiter01] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," In *A Field Guide to Dynamical Recurrent Neural Networks*, 2001.

[Holland62] J. H. Holland, "Outline for a logical theory of adaptive systems," *Journal of the Association of the Computing Machinery*, Vol. 9, No. 3, pp. 297-314, July 1962.

[Hornik89] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators," Neural Networks, vol. 2, no. 5, pp. 359–366, 1989.

[Ioffe15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in ICML, 2015.

[Jaeger01] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks," Technical Report GMD Report 148, German National Research Center for Information Technology, 2001.

[Jaeger07] H. Jaeger, "Discovering multiscale dynamical features with hierarchical echo state networks," Technical report 10, School of Engineering and Science, Jacobs University, 2007.

[Jordan90] M. I. Jordan, "Attractor dynamics and parallelism in a connectionist sequential machine," IEEE Press, pp. 112–127, 1990.

[Karangelos12] E. Karangelos and F. Bouffard, "Towards Full Integration of Demand-Side Resources in Joint Forward Energy/Reserve Electricity Markets, " in *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 280-289, Feb. 2012.

[Karim18] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate LSTM-FCNs for time series classification," arXiv preprint arXiv:1801.04503, 2018.

[Kaur15] D. Kaur, T. T. Lie, N. K. C. Nair and B. Valles, "Wind Speed Forecasting Using Hybrid Wavelet Transform-ARMA Techniques," AIMS Energy, vol. 3, pp. 13-24, 2015.

[Kavousi-Fard16] A. Kavousi-Fard, A. Khosravi and S. Nahavandi, "A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting," in *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 18-26, Jan. 2016.

[Kazempour09] S. J. Kazempour, M. P. Moghaddam, M. R. Haghifam, and G. R. Yousefi, “Risk-constrained dynamic self-scheduling of a pumped-storage plant in the energy and ancillary service markets,” Energy Conversion and Management, vol. 50, no. 5, pp. 1368–1375, May 2009.

[Kazempour12] S. J. Kazempour and A. J. Conejo, "Strategic Generation Investment Under Uncertainty Via Benders Decomposition," in *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 424-432, Feb. 2012.

[Kazempour15] S. J. Kazempour, A. J. Conejo and C. Ruiz, "Strategic Bidding for a Large Consumer," in *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 848-856, March 2015.

[Khodayar13,a] M. E. Khodayar and M. Shahidehpour, “Stochastic Price-Based Coordination of Intrahour Wind Energy and Storage in a Generation Company” *IEEE Trans. Sust. Energy*, vol. 4, no. 3, pp. 554–562, Jul. 2013.

[Khodayar13,b] M. E. Khodayar, M. Shahidehpour and L. Wu, “Enhancing the Dispatchability of Variable Wind Generation by Coordination With Pumped-Storage Hydro Units in Stochastic Power Systems,” *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2808–2818, Aug. 2013.

[Khosravi10] A. Khosravi, S. Nahavandi, and D. Creighton, “Construction of optimal prediction intervals for load forecasting problem,” in *IEEE Trans. Power Syst.*, vol. 25, pp. 1496–1503, 2010.

[Koch11] T. Koch, T. Achterberg, E. Andersen, O. Bastert, T. Berthold, R. E. Bixby, E. Danna, G. Gamrath, A. M. Gleixner, S. Heinz, A. Lodi, H. Mittelmann, T. Ralphs, D. Salvagnin, D. E. Steffy, and K. Wolter, “MIPLIB 2010,” *Mathematical Programming Computation*, vol. 3, no. 2, pp. 103–163, Jun. 2011.

[Koenker78] R. Koenker, and B. Jr Gilbert “Regression quantiles.” Econometrica: journal of the Econometric Society, pp. 33-50, 1978.

[Koistinen91] P. Koistinen and L. Holmström, “Kernel Regression and Backpropagation Training with Noise,” in J. E. Moody, S. J. Hanson, and R. Lippmann, editors, Advances in Neural Information Processing Systems, vol. 4, pp. 1033–1039. Morgan Kaufmann, 1991.

[Koren09] Y. Koren, R. M. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” IEEE Computer, 42(8):30–37, 2009.

[Lamb16] A. M. Lamb, A. Goyal, Y. Zhang, S. Zhang, A. C. Courville, and Y. Bengio. “Professor forcing: A new algorithm for training recurrent networks.” in Advances In Neural Information Processing Systems, pp. 4601-4609, 2016.

[Law00] A. M. Law, W. D. Kelton. “Simulation modeling and analysis,” third ed. New York: McGraw-Hill, 2000.

[LeCun98] Y. LeCun, L. Bottou, G. Orr, and K. Muller, “Efficient BackProp,” in G. Orr and M. K., editors, Neural Networks: Tricks of the trade. Springer, 1998.

[LeCun[15]] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

[Liang[11]] J. Liang, S. Grijalva and R. G. Harley, "Increased Wind Revenue and System Security by Trading Wind Power in Energy and Regulation Reserve Markets," in *IEEE Trans. on Sust. Energy*, vol. 2, no. 3, pp. 340-347, July 2011.

[Luenberger[84]] D. Luenberger, *Linear and Nonlinear Programming*. Reading, MA, USA: Addison-Wesley, 1984.

[Luenberger[16]] D. Luenberger and Y. Ye, “Linear and Nonlinear Programming,” International Series in Operation Research & Management Science, Springer, 2016.

[Lukosevicius[12]] M. Lukosevicius, “A practical guide to applying echo state networks,” In Neural Networks: Tricks of the Trade, volume 7700 of Lecture Notes in Computer Science. Springer, 2nd edition, 2012.

[Ma[13]] X.-Y. Ma, Y.-Z. Sun, and H.-L. Fang, “Scenario generation of wind power based on statistical uncertainty and variability,” in *IEEE Trans. Sust. Energy*, vol. 4, no. 4, pp. 894–904, Oct. 2013.

[Maass[02]] W. Maass, T. Natschläger, and H. Markram, “Real-time computing without stable states: a new framework for neural computation based on perturbations,” Neural Computation, vol. 14, no. 11, pp. 2531– 2560, 2002.

[Maringer[05]] D. Maringer, “Heuristic optimization, Portfolio Management with Heuristic Optimization,” Springer, 2005.

[Martinez[12]] J. J. Martinez, P. Kjaer, P. Rodriguez, and R. Teodorescu, “Design and analysis of a slope voltage control for a DFIG wind power plant,” *IEEE Trans. Energy Conversion*, vol. 27, no. 1, pp. 11–20, 2012.

[Martins[14]] L. S. A. Martins, A. T. Azevedo, and S. Soares, “Nonlinear Medium-Term Hydro-Thermal Scheduling With Transmission Constraints,” *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1623–1633, Jul. 2014.

[Mashhour[11]] E. Mashhour and S. M. Moghaddas-Tafreshi, "Bidding Strategy of Virtual Power Plant for Participating in Energy and Spinning Reserve Markets—Part I: Problem Formulation," in *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 949-956, May 2011.

[Matevosyan[06]] J. Matevosyan, and L. Söder,”Minimization of Imbalance Cost Trading Wind Power on the Short-Term Power Market,” *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1396–1404, Aug. 2006.

[McCulloch[88]] W. S. McCulloch and W. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” MIT Press, pp. 15–27, 1988.

[Mercier[17,a]] T. Mercier, J. Jomaux, E. De Jaeger, M. Olivier, “Provision of primary frequency control with variable-speed pumped-storage hydropower,” in IEEE Powertech, Manchester, 2017.

[Mercier[17,b]] T. Mercier, E. De Jaeger, "Smartwater WP3. Control of variable-speed pumped storage," 2017.

[Michalewicz[92]] Z. Michalewicz, "Genetic Algorithms + Data Structures = Evolution Programs," Springer-Verlag, Berlin, 1992.

[Morales[10,a]] J. M. Morales, R. Minguez and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," Applied Energy, vol. 87, pp. 843-855, 2010.

[Morales[10,b]] J. M. Morales, A. J. Conejo and J. Pérez-Ruiz, "Short-Term Trading for a Wind Power Producer," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 554–564, Feb. 2010.

[Morales-Espana[13]] G. Morales-Espana, J. M. Latorre, and A. Ramos, "Tight and Compact MILP Formulation of Start-Up and Shut-Down Ramping in Unit Commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1288–1296, 2013

[Morales-Espana[14]] G. A. Morales-Espana, "Unit Commitment Computational Performance, System Representation and Wind Uncertainty Management," Ph.D. dissertation, Universidad Pontificia Comillas, 2014.

[Morren[06]] J. Morren, S. W. H. de Haan, W. L. Kling and J. A. Ferreira, "Wind turbines emulating inertia and supporting primary frequency control," in *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 433-434, Feb. 2006.

[Murray[94]] A. F. Murray, and P. J. Edwards, "Enhanced MLP Performance and Fault Tolerance Resulting from Synaptic Weight Noise During Training," *IEEE Trans. on Neural Networks*, vol. 5, pp. 792–802, 1994.

[Nelsen[98]] R.B. Nelsen, *An introduction to copulas*, New-York: Springer, 1998.

[Nemhauser[99]] G. L. Nemhauser and L. A. Wolsey, "*Integer and combinatorial optimization*," New York: John Wiley and Sons, 1999.

[Nielsen[06]] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," Wind Energy, vol. 9, no. 1–2, pp. 95–108, 2006.

[Nilsson[98]] O. Nilsson, L. Söder, and D. Sjelvgren, "Integer modeling of spinning reserve requirements in short term scheduling of hydro systems," *IEEE Trans. Power Syst.*, vol. 13, pp. 959–964, Aug. 1998.

[Ostrowski[12]] J. Ostrowski, M. F. Anjos, and A. Vannelli, "Tight mixed integer linear programming formulations for the unit commitment problem," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 39–46, Feb. 2012.

[Pandzic[13,a]] H. Pandžić, I. Kuzle and T. Capuder, "Virtual power plant mid-term dispatch optimization," Applied Energy, vol. 101, pp. 134-141, Jan. 2013.

[Pandzic[13,b]] H. Pandžić, J. M. Morales, A. J. Conejo, and I. Kuzle, “Offering model for a virtual power plant based on stochastic programming,” Appl. Energy, vol. 105, pp. 282–292, 2013.

[Pannatier[10]] Y. Pannatier, “Optimisation des stratégies de réglage d’une installation de pompage-turbinage à vitesse variable,” Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2010.

[Papaefthymiou[08]] G. Papaefthymiou and P. Pinson, “Modeling of spatial dependence in wind power uncertainty,” Proceedings of the international conference on probabilistic methods applied to power systems, Rincon (Puerto Rico), 2008.

[Papavasiliou[11]] A. Papavasiliou, S. S. Oren and R. P. O'Neill, "Reserve Requirements for Wind Power Integration: A Scenario-Based Stochastic Programming Framework," in *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2197-2206, Nov. 2011.

[Peng[14]] Y. Peng, M. Lei, J.-B. Li, and X.-Y. Peng, “A novel hybridization of echo state networks and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting,” Neural Comput. Appl., vol. 24, pp. 883–890, 2014.

[Pérez-Diaz[15]] J. I. Pérez-Diaz, M. Chazarra, J. Garcia-Gonzalez, G. Cavazzini and A. Stoppato, “Trends and challenges in the operation of pumped-storage hydropower plants,” Renewable and Sustainable Energy Reviews, vol. 44, pp. 767-784, 2015.

[Pfeifer[03]] D. Pfeifer, and J. Neslehova. “Modeling dependence in finance and insurance: the copula approach,” Blätter der DGVFM, vol. 26, no. 2, pp. 177–191, 2003.

[Pinson[07]] P. Pinson, C. Chevallier and G. N. Kariniotakis, "Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power," in *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1148-1156, Aug. 2007.

[Pinson[08]] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, B. Klöckl, “From probabilistic forecasts to statistical scenarios of short-term wind power production,” Wind Energy, vol. 12, no. 1, pp. 51-62, 2008.

[Pinson[10]] P. Pinson and G. Kariniotakis, "Conditional Prediction Intervals of Wind Power Generation," in *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845-1856, Nov. 2010.

[Pinson[17]] P. Pinson, “Uncertainties – what are we talking about? Current challenges related to uncertainties,” in DTU Summer School *Uncertainty in Electricity Markets and System Operation*, July 2016.

[Plaut[86]] D. C. Plaut, S. J. Nowlan, and G. E. Hinton, “Experiments on Learning by BackPropagation,” Technical Report CMU–CS–86–126, Carnegie–Mellon University, 1986.

[Pourmousavi[14]] S. A. Pourmousavi, S. N. Patrick and M. H. Nehrir, "Real-Time Demand Response Through Aggregate Electric Water Heaters for Load Shifting and Balancing Wind Generation," in *IEEE Trans. on Smart Grid*, vol. 5, no. 2, pp. 769-778, March 2014.

[Plazas[05]] M.A. Plazas, A.J. Conejo and F.J. Prieto, “Multimarket optimal bidding for a power producer,” *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2041–2050, Nov. 2005.

[Ponrajah98] R. A. Ponrajah, J. Witherspoon, and F. D. Galiana, “Systems to optimize conversion efficiencies at Ontario Hydro’s hydroelectric plants,” *IEEE Trans. Power Syst.*, vol. 13, pp. 1044–1050, Aug. 1998.

[Poulain18] A. Poulain, J.-R. de Dreuzy, P. Goderniaux, “Pump Hydro Energy Storage systems (PHES) in groundwater flooded quarries,” Journal of Hydrology, available online 27 February 2018.

[Pousinho13] H. M. I. Pousinho, J. Contreras, A. G. Bakirtzis and J. P. S. Catalão, “Risk-Constrained Scheduling and Offering Strategies of a Price-Maker Hydro Producer Under Uncertainty,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1879–1887, May 2013.

[Pujades17] E. Pujades, P. Orban, S. Bodeux, P. Archambeau, S. Erpicum and A. Dassargues, “Underground pumped storage hydropower plants using open pit mines: How do groundwater exchanges influence the efficiency?” in Applied Energy, vol. 190, pp. 135-146, 2017.

[Quan14] H. Quan, D. Srinivasan and A. Khosravi, "Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals," in *IEEE Trans. on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 303-315, Feb. 2014.

[Quan15] H. Quan, D. Srinivasan and A. Khosravi, "Incorporating Wind Power Forecast Uncertainties Into Stochastic Unit Commitment Using Neural Network-Based Prediction Intervals," in *IEEE Trans. Neural Networks and Learning Systems*, vol. 26, no. 9, pp. 2123-2135, Sept. 2015.

[Queipo05] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan and P. K. Tucher, “Surrogate-based Analysis and Optimization,” Progress in Aerospace Sciences, vol. 41, pp. 1-28, 2005.

[Rajan05] D. Rajan, S. Takriti, and Y. Heights, “IBM Research Report Minimum Up / Down Polytopes of the Unit Commitment Problem with Start-Up Costs,” tech. rep., IBM Research Division, 2005.

[Regnier03] J. Régnier, “Conception de système hétérogènes en génie électrique par optimisation évolutionnaire multicritère,” PhD thesis, Institut National Polytechnique de Toulouse, 2003.

[Renner18] R. H. Renner and D. Van Hertem, "Potential of wind farms connected to HVDC grid to provide DC ancillary services," in *IEEE Transactions on Sustainable Energy*, in press.

[Robinson87] A. J. Robinson and F. Fallside, “The Utility Driven Dynamic Error Propagation Network,” Technical Report CUED/F-INFENG/TR.1, Cambridge University Engineering Department, 1987.

[Rockafellar00] R. T. Rockafellar and S. Uryasev, “Optimization of conditional value-at-risk,” *J. Risk.*, vol. 2, pp. 21–41, 2000.

[Römisch03] W. Römisch and H. Heitsch, “Scenario reduction algorithms in stochastic programming,” Computational Optimization and Applications, vol. 24, pp. 187–206, 2003.

[Rosenblatt63] F. Rosenblatt. "Principles of Neurodynamics," Spartan, New York, 1963.

[Rothberg03] E. Rothberg, "The CPLEX library: Presolve and cutting planes," in *4th Max-Planck Advanced Course on the Foundations of Computer Science (ADFOCS)*, Saarbrücken, Germany, pp. 1–17, Sep. 2003.

[Rumelhart86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," MIT Press, pp. 318–362, 1986.

[Sánchez de la Nieta13] A. A. Sánchez de la Nieta, J. Contreras and J. I. Muñoz, "Optimal coordinated wind-hydro bidding strategies in day-ahead markets," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 798–809, May 2013.

[Sánchez de la Nieta14] A. A. Sánchez de la Nieta, J. Contreras, J. I. Muñoz, and M. O'Malley, "Modeling the Impact of a Wind Power Producer as a Price-Maker," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2723–2732, Nov. 2014.

[Sareni99] B. Sareni, "Méthodes d'optimisation multimodales associées à la modélisation numérique en électromagnétisme," *PhD thesis,* École Centrale de Lyon, 1999.

[Seto15] S. Seto, W. Zhang, and Y. Zhou, "Multivariate time series classification using dynamic time warping template selection for human activity recognition," IEEE Symposium Series on Computational Intelligence, pp. 1399–1406, 2015.

[Schäfer 05] O. Schäfer, "A joint declaration by the European renewable energy sector to urge for a directive on heating and cooling from renewable energy sources," EREC Edinburgh, 6th Interparliamentary Forum on Renewable Energy and Energy Efficiency, 07 October 2005.

[Shrivastava15] N. A. Shrivastava, A. Khosravi and B. K. Panigrahi, "Prediction Interval Estimation of Electricity Prices Using PSO-Tuned Support Vector Machines," in *IEEE Trans. on Industrial Informatics*, vol. 11, no. 2, pp. 322-331, April 2015.

[Simpson08] T. W. Simpson, V. Toropov, V. Balabanov and F. A. C. Viana, "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: a Review of How Far We Have Come – or Not," AIAA Paper 2008-5802, 2008.

[Sklar59] A. Sklar, "Fonctions de repartition à n dimensions et leurs marges", Publ. Inst. Statist. Paris, vol. 8, pp. 229-231, 1959.

[Smartwater18] Smartwater – PHES energy storage using end-of-life quarries and mines – Belgium Wallonia specific use-cases, webpage http://www.gridinnovation-on-line.eu/articles/library/smartwater---phes-energy-storage-using-end-of-life-quarries-and-mines---belgium-wallonia-specific-use-cases.kl last consulted on 04 April 2018.

[Srivastava14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting. J. Machine Learning Res. vol. 15, pp. 1929–1958, 2014.

[Schuster97] M. Schuster and K. K. Paliwal, “Bidirectional Recurrent Neural Networks,” *IEEE Trans. on Signal Processing*, vol. 45, pp. 2673–2681, 1997.

[Soltani02] S. Soltani, “On the use of wavelet decomposition for time series prediction,” in Neurocomput., vol. 48, pp. 267–277, Oct. 2002.

[Sousa12] T. Sousa, H. Morais, Z. Vale, P. Faria, J. Soares, “Intelligent Energy Resource Management Considering Vehicle-to-Grid: A Simulated Annealing Approach,” *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 535–542, Mar. 2012.

[Steinwart11] I. Steinwart, A. Christmann, “Estimating conditional quantiles with the help of the pinball loss,” Bernoulli, vol. 17, no. 1, pp. 211–225, 2011.

[Strelen07] J. C. Strelen, F. Nassaj, “Analysis and generation of random vectors with copulas”, Proceedings of the 2007 Winter Simulation Conference, pp. 488-496, 2007.

[Strelen09] J. C. Strelen, “Tools for dependent simulation input with copulas”, SimuTools, p. 30, 2009.

[Sugiartawan17] P. Sugiartawan, R. Pulungan, A. K. Sari, “Prediction by a Hybrid of Wavelet Transform and Long-Short-Term-Memory Neural Network,” International Journal of Advanced Computer Science and Applications, vol. 8, no. 2, 2017.

[Sugiyama02] M. Sugiyama and H. Ogawa, “Release from active learning/model selection dilemma: optimizing sample points and models at the same time,” in Neural Networks, 2002. IJCNN ’02. Proceedings of the 2002 International Joint Conference on, vol. 3, pp. 2917– 2922, 12-17 May 2002.

[Sun17] X. A. Sun, A. Lorca, “Robust Optimization in Electric Power System Operation,” in Integration of Large-Scale Renewable Energy into Bulk Power Systems, Springer, 2017.

[Swissgrid10] Swissgrid, “Overview of ancillary services,” 2010.

[Takeuchi06] I. Takeuchi, Q. V. Le, T. Sears, and A. J. Smola. “Nonparametric quantile estimation,” Journal of Machine Learning Research, 7, 2006.

[Tastu13] J. Tastu and P. Pinson, “Space-time scenarios of wind power generation produced using a Gaussian copula with parametrized precision matrix,” tech. rep., Technical University of Denmark, 2013.

[Thatte13] A. A. Thatte, L. Xie, D. E. Viassolo and S. Singh, “Risk Measure Based Robust Bidding Strategy for Arbitrage Using a Wind Farm and Energy Storage," in *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2191-2199, Dec. 2013.

[TenneT17] TenneT, “Market Review 2016 – Electricity market insights”, 2017.

[Tong13] B. Tong, Q. Zhai and X. Guan, "An MILP Based Formulation for Short-Term Hydro Generation Scheduling With Analysis of the Linearization Effects on Solution Feasibility," in *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3588-3599, Nov. 2013.

[Toubeau[16]] J.-F. Toubeau, M. Hupez, V. Klonari, Z. De Grève and F. Vallée, "Statistical Load and Generation Modelling for Long Term Studies of Low Voltage Networks in Presence of Sparse Smart Metering Data", Proc. of the 42nd Annual Conference of IEEE Industrial Electronics Society (IECON), Florence (Italy), Oct. 2016.

[Toubeau[17]] J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, “Improved Day-Ahead Predictions of Load and Renewable Generation by Optimally Exploiting Multi-Scale Dependencies”, in Proc. of the *7th IEEE conference on Innovative Smart Grid Technologies (ISGT ASIA)*, 5 pages, Dec. 2017.

[Toubeau[18]] J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, “Deep Learning-based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets,” under review.

[Ugedo[06]] A. Ugedo, E. Lobato, A. Franco, L. Rouco, J. Fernández-Caro, J. Chofre, “Strategic bidding in sequential electricity markets,” IEE Proc – Gen. Transm. Distrib., no.153, vol. 4, pp. 431–42, 2006.

[United Nations[15]] United Nations, “Paris Agreement,” 2015.

[Van den Bergh[15,a]] K. Van den Bergh, D. Couckuyt, E. Delarue, and W. D’haeseleer, “Redispatching in an interconnected electricity system with high renewables penetration,” *Electr. Power Syst. Res.*, vol. 127, no. 10, pp. 64–72, 2015.

[Van den Bergh[15,b]] K. Van den Bergh and E. Delarue, “Cycling of conventional power plants: Technical limits and actual costs,” Energy Conversion and Management, vol. 97, pp. 70 – 77, Jun. 2015.

[Van de Vyver[14]] J. Van de Vyver, T. L. Vandoorn, J. D. M. De Kooning, B. Meersman, and L. Vandevelde, “Provision of ancillary services with variable speed wind turbines,” in IEEE Young Researchers Symposium Proceedings. EESA, pp. 1–6, Apr. 2014.

[van Stiphout[17]] A. van Stiphout, K. De Vos and G. Deconinck, "The Impact of Operating Reserves on Investment Planning of Renewable Power Systems," in *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 378-388, Jan. 2017.

[Vargas-Serrano[17]] A. Vargas-Serrano, A. Hamann, S. Hedtke, C. M. Franck and G. Hug, "Economic benefit analysis of retrofitting a fixed-speed pumped storage hydropower plant with an adjustable-speed machine," *2017 IEEE Manchester PowerTech*, Manchester, 2017, pp. 1-6.

[Vermaak[98]] J. Vermaak and E.C. Botha, “Recurrent Neural Networks for Short-Term Load Forecasting,” in *IEEE Trans. Power Syst.*, vol. 13, no. 1, pp. 126-132, Feb. 1998.

[Verstraeten[10]] D. Verstraeten, “Reservoir Computing: computation with dynamical systems,” PhD thesis, 2010.

[Wan[14]] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong and K. P. Wong, "Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine," in *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033-1044, May 2014.

[Wan17] C. Wan; J. Lin; Y. Song; Z. Xu; G. Yang, "Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach," in *IEEE Trans. Power Syst.* , vol. 32, no. 3, pp. 2471-2472, May 2017.

[Wei16] W. Wei, F. Liu and S. Mei, "Distributionally Robust Co-Optimization of Energy and Reserve Dispatch," in *IEEE Trans. Sust. Energy*, vol. 7, no. 1, pp. 289-300, Jan. 2016.

[Weiss12] G. N.F. Weiss, and M. Scheer "Smooth nonparametric bernstein vine copulas," Available at SSRN 2154458, 2012.

[Wen17] R. Wen, K. Torkkola, and B. Narayanaswamy, "A Multi-Horizon Quantile Recurrent Forecaster," arXiv preprint arXiv:1711.11053, 2017.

[Werbos90] P. Werbos, "Backpropagation Through Time: What It Does and How to Do It," Proceedings of the IEEE, vol. 78, no. 10, pp. 1550 – 1560, 1990.

[Williams95] R. J. Williams and D. Zipser, "Gradient-Based Learning Algorithms for Recurrent Networks and Their Computational Complexity," in Y. Chauvin and D. E. Rumelhart, editors, Back-propagation: Theory, Architectures and Applications, pp. 433–486. Lawrence Erlbaum Publishers, 1995.

[Wolsey98] L.A. Wolsey, "*Integer Programming*," Wiley-Interscience, 1998.

[Wu12] L. Wu, M. Shahidehpour and Z. Li, "Comparison of Scenario-Based and Interval Optimization Approaches to Stochastic SCUC," in *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 913-921, May 2012.

[Wu13] H. Wu, M. Shahidehpour and M. E. Khodayar, "Hourly Demand Response in Day-Ahead Scheduling Considering Generating Unit Ramping Cost," in *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2446-2454, Aug. 2013.

[Wu14] H. Wu, M. Shahidehpour, Z. Li and W. Tian, "Chance-Constrained Day-Ahead Scheduling in Stochastic Power System Operation," in *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1583-1591, July 2014.

[Wyffels13] F. Wyffels, "Sequence Generation with Reservoir Computing Systems," PhD thesis, 2013.

[Xi99] E. Xi, X. Guan and R. Li, "Scheduling hydrothermal power systems with cascaded and head-dependent reservoirs," in *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1127-1132, Aug 1999.

[Yu05] C.-N. Yu, A. I. Cohen and B. Danai, "Multi-interval optimization for real-time power system scheduling in the Ontario electricity market," *IEEE Power Engineering Society General Meeting, 2005*, 2005, pp. 2717-2723 Vol. 3.

[Zapata14] J. Zapata, J. Vandewalle and W. D'haeseleer, "A comparative study of imbalance reduction strategies for virtual power plant operation," Applied Thermal Engineering, vol. 71, no. 2, pp. 847–857, Oct. 2014.

[Zhang12] Q. F. Zhang, B. Smith, W. Zhang, “Small Hydropower Cost Reference Model,” Final project report from Oak Ridge National Laboratory, 2012.

[Zhang14] Y. Zhang, J. Wang and X. Wang, “Review on probabilistic forecasting of wind power generation,” Renewable and Sustainable Energy Reviews, vol. 32, pp. 255-270, April, 2014.

[Zhao08] J. H. Zhao, Z. Y. Dong, Z. Xu, and K. P. Wong, “A statistical approach for interval forecasting of the electricity price,” in *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 267–276, 2008.

[Zugno13] M. Zugno, J. M. Morales, P. Pinson and H. Madsen, "Pool Strategy of a Price-Maker Wind Power Producer," in *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3440-3450, Aug. 2013.

ANNEX A

LIST OF PUBLICATIONS

A.1 Publications related to the thesis

Peer-review journal articles

1. J. F. Toubeau, Z. De Grève and F. Vallée, “Medium-Term Multimarket Optimization for Virtual Power Plants: A Stochastic-Based Decision Environment,” in *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1399-1410, March 2018.

2. J.-F. Toubeau, S. Iassinovski, E. Jean, J.-Y. Parfait, J. Bottieau, Z. De Grève, and F. Vallée, “A Nonlinear Hybrid Approach for the Scheduling of Merchant Underground Pumped Hydro Energy Storage,” in *IET Generation, Transmission & Distribution*, in press.

3. J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, “Deep Learning-based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets,” in *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203-1215, March 2019.

4. J.-F. Toubeau, Z. De Grève, P. Goderniaux, F. Vallée and K. Bruninx, "Chance-Constrained Scheduling of Underground Pumped Hydro Energy Storage in Presence of Model Uncertainties," in *IEEE Transactions on Sustainable Energy*, in press.

Peer-review conference papers

1. J.-F. Toubeau, J. Bottieau, F. Vallée and Z. De Grève, “Improved Day-Ahead Predictions of Load and Renewable Generation by Optimally Exploiting Multi-Scale Dependencies,” in IEEE Innovative Smart Grid Technologies, Auckland, New-Zealand, 2017.

2. J.-F. Toubeau, Z. De Grève, F. Vallée, “Technical Impacts on Distribution Systems of Medium-Sized Storage Plants Participating in Energy and Power Reserve Markets,” in 24th International Conference & Exhibition on Electricity Distribution, CIRED 2017, Glasgow, Scotland.

3. J. Bottieau, F. Vallée, Z. De Grève and J.-F. Toubeau, “Leveraging Provision of Frequency Regulation Services from Wind Generation by Improving Day-Ahead

Predictions using LSTM Neural Networks," in IEEE Energycon, Limassol, Chyprus, 2018.

A.2 Publications in Electrical Engineering

Peer-review journal articles

1. F. Vallée, F. Moutier, V. Klonari, J.-F. Toubeau, F. Lecron, Z. De Grève, and J. Lobry, "On the Correlation between Prosumers in Probabilistic Analysis of Low Voltage Distribution Systems," in International Review of Electrical Engineering, vol. 11, pp. 277-283, 2016.

2. V. Klonari, J.-F. Toubeau, T. Vandoorn, B. Meersman, Z. De Grève, J. Lobry, and F. Vallée, "Probabilistic Framework for Evaluating Droop Control of Photovoltaic Inverters," in Electric Power Systems Research, vol. 129, pp. 1-9, December 2015.

3. V. Klonari, J.-F. Toubeau, Z. De Grève, O. Durieux, J. Lobry, and F. Vallée, "Probabilistic Simulation Framework of the Voltage Profile in Balanced and Unbalanced Low Voltage Networks," in International Journal of Electrical Power & Energy Systems, vol. 82, pp. 439-451, November 2016.

4. J.-F. Toubeau, F. Vallée, Z. De Grève, J. Lobry, "A New Approach Based on the Experimental Design Method for the Improvement of the Operational Efficiency of Medium Voltage Distribution Networks " in International Journal of Electrical Power & Energy Systems, vol. 66, pp. 116-124, March 2015.

5. F. Vallée, J.-F. Toubeau, Z. De Grève, J. Lobry, "Consideration of Extreme Wind Geographical Correlation Scenarios in Reliability Assessment Studies Using Sequential Monte Carlo Simulation" in International Review of Electrical Engineering, vol. 9, no. 6, pp. 1148-1153, 2014.

Book Chapters

1. V. Klonari, J.-F. Toubeau, J. Lobry, and F. Vallée, "Estimating the Photovoltaic Hosting Capacity of a Low Voltage Feeder Using Smart Meters Measurements," in Smart Metering Technology and Services - Inspirations for Energy Utilities, Intech, 978-953-51-2452-8, 2016.

2. J.-F. Toubeau, V. Klonari, J. Lobry, Z. De Grève, F. Vallée, "Planning Tools for the Integration of Renewable Energy Sources into Low and Medium Voltage Distribution Grids," in Renewable Energy- Utilisation and System Integration, Intech, 978-953-51-2408-5, 2016.

3. M. Hupez, J.-F. Toubeau, Z. De Grève, F. Vallée, "SARMA Time Series for Microscopic Electrical Load Modeling," in Contribution to Statistics – Advances in Time Series Analysis & Forecasting," Editions Springer, 2017.

Peer-review conference papers

1. F. Vallée, M. Hupez, J.-F. Toubeau, Z. De Grève, "Dealing with Sparse Smart Metering Data in Techno-Economic Analysis of Low Voltage Networks," in 24th International Conference & Exhibition on Electricity Distribution, CIRED 2017, Glasgow, Scotland.

2. J.-F. Toubeau, M. Hupez, V. Klonari, Z. De Grève, F. Vallée, "Statistical Load and Generation Modelling for Long Term Studies of Low Voltage Networks in Presence of Sparse Smart Metering Data" in 42nd Annual Conference of Industrial Electronics Society, IEEE IECON 2016, Florence, Italy.

3. M. Hupez, J.-F. Toubeau, Z. De Grève, F. Vallée, "On the Application of SARMA Time Series to Microscopic Electrical Load Modeling" in International Work-Conference on Time Series Analysis, ITISE 2016, Granada, Spain.

4. V. Klonari, J.-F. Toubeau, J. Lobry, F. Vallée, "Photovoltaic Integration in a Smart City Power Distribution: A Probabilistic PV Hosting Capacity Assessment Based on Smart Metering Data" in 5th International Conference on Smart Cities and Green ICT Systems, SMARTGREENS 2016, Rome, Italy.

5. Z. De Grève, J. Vanstals, J.-F. Toubeau, F. Geth, P. C. Ramasmawy, S. Rapoport, F. Vallée, "Impact of the Geographical Correlation between Wind Speed Time Series on Reliability Indices in Power System Studies" in IEEE International Energy Conference, ENERGYCON 2016, Leuven, Belgium.

6. F. Vallée, J.-F. Toubeau, Z. De Grève, J. Lobry, "Adjusted LOLE and EENS indices for the Consideration of Load Excess Transfer in Power Systems Adequacy Studies" in XIV International Conference on Electrical, Electronics and Power Engineering, ICEEPE'16, London, United Kingdom.

7. F. Vallée, F. Moutier, V. Klonari, J.-F. Toubeau, F. Lecron, Z. De Grève, J. Lobry, "Clustering of Photovoltaic Generation for the Consideration of Time Changing Geographical Correlation in Probabilistic Analysis of Low Voltage Distribution Systems" in 5th International Workshop on Integration of Solar Power into Power Systems, Brussels, Belgium.

8. P. Rousseaux, J.-F. Toubeau, Z. De Grève, F. Vallée, M. Glavic, T. Van Cutsem, "A New Formulation of State Estimation in Distribution Systems Including Demand and Generation States" in IEEE PowerTech 2015 International Conference, Eindhoven, The Netherlands.

9. J.-F. Toubeau, A. Brito Gonçalves de Sà, Z. De Grève, O. Durieux, F. Vallée, J. Lobry, "Optimal Positioning and Pre-Sizing of Storage Devices for the Improvement of MV Distribution Grid Operation" in 23rd International Conference & Exhibition on Electricity Distribution, CIRED 2015, Lyon, France.

10. V. Klonari, J.-F. Toubeau, Z. De Grève, O. Durieux, J. Lobry, F. Vallée, "Probabilistic Analysis Tool of the Voltage Profile in Low Voltage Grids" in 23rd International Conference & Exhibition on Electricity Distribution, CIRED 2015, Lyon, France.

11. V. Klonari, J.-F. Toubeau, Z. De Grève, T. Vandoorn, B. Meersman, F. Vallée, J. Lobry, "Probabilistic Assessment of PDC/Vg Droop Control of PV Inverters" in 23rd

International Conference & Exhibition on Electricity Distribution, CIRED 2015, Lyon, France.

12. V. Klonari, J.-F. Toubeau, Z. De Grève, J. Lobry, F. Vallée, "Probabilistic Analysis of Low Voltage Networks with Distributed Photovoltaic Generation Sources: Case Study in Belgium" in International Workshop on Photovoltaic Cell Systems, IWPCS 2015, Saïda, Maroc.

13. F. Vallée, J.-F. Toubeau, Z. De Grève, J. Lobry, "Impact of Wind Geographical Correlation in Reliability Assessment Studies Using Sequential Monte Carlo Simulations" in International Conference on Renewable Energies and Power Quality, ICREPQ'15, La Coruna, Espagne.

14. J.-F. Toubeau, V. Klonari, Z. De Grève, J. Lobry, F. Vallée, "Probabilistic Study of the Impact on the Network Equipment of Changing Load Profiles in Modern Low Voltage Grids" in International Conference on Renewable Energies and Power Quality, ICREPQ'15, La Coruna, Espagne.

15. J.-F. Toubeau, J. Lobry, F. Vallée, Z. De Grève, "Optimal Allocation Process of Voltage Control Devices and Operational Management of the Voltage in Distribution Systems Using the Experimental Design Method," in IEEE International Energy Conference, Energycon 2014, Dubrovnik, Croatia.

16. J.-F. Toubeau, F. Vallée, Z. De Grève, J. Lobry "A New Approach for the Voltage Regulation in Medium Voltage Networks Using the Experimental Design Method," in 7^{th} Young Researchers Symposium in Electric Power Engineering, Ghent, Belgium, 2014.

A.3 Publications outside Electrical Engineering

Peer-review journal articles

1. C. Liefferinckx, C. Minsart, J.-F. Toubeau, A. Cremer, L. Amininejad, E. Quertinmont, et al. "Infliximab Trough Levels at Induction to Predict Treatment Failure During Maintenance," Inflammatory Bowel Disease, vol. 23, no. 8, pp. 1371-81, 2017.

Peer-review conference papers

1. C. Liefferinckx, C. Minsart, A. Cremer, L. Amininejad, J.-F. Toubeau, E. Quertinmont, J. Devière, A. van Gossum, D. Franchimont, "Infliximab trough level measured during treatment induction may be predictive of the loss of response to IFX during treatment maintenance in Inflammatory Bowel Disease patients: A longitudinal observational retrospective study," in DWW 2016, San Diego; ECCO 2016, Amsterdam; BWGE 2016, Brussels.

2. C. Liefferinckx, C. Minsart, J.-F. Toubeau, A. Cremer, L. Amininejad, E. Quertinmont, J. Deviere, A. Gils, A. Van Gossum, D. Franchimont, "Trough levels (TLs) at induction: Impact on long term response when re-initiating infliximab," ECCO 2017, Barcelona; DWW 2017, Chicago.

ANNEX B

COPULA

B.1 Introduction

The term copula has a Latin origin that means "*connection*". Copula is an emerging statistical concept whose popularity originates from Sklar's theorem stating that any continuous multivariate distribution can be expressed as a simple (copula) function of its one-dimensional constitutive marginal [Sklar[59]]. Copulas provides thus convenient way to compute multivariate distributions by containing the whole inter-dependence structure between different variables.

A valuable asset of copulas is that it allows decoupling the individual marginals from their dependence structure. In this way, two different set of variables with the same dependence structure are thus characterized by the same copula model but can have different multivariate distribution if there are discrepancies in their marginal distributions.

Overall, the use of copulas allows substituting the difficult task of identifying a multivariate distribution by performing two simpler tasks. The first one consists in appropriately modelling the marginal distributions of each variable and the second is to estimate the copula, which summarizes the whole dependence structure. Hence, copulas enable generating vectors with any specific dependence structure, and not only the linear correlation such as traditional methods.

B.2 Definition

A copula is a function that joins (couples) a multivariable distribution to its one-dimensional marginal distribution. In fact, a copula is a *D*-dimensional distribution function with univariate uniform margins restricted to the unit *D*-cube $[0, 1]^D$ [Pfeifer[03]].

Sklar theorem

Let $\boldsymbol{X} = (X_1, \ldots, X_D)$ be a random vector, and H be a *D*-dimensional distribution function with marginal distributions $F_1, \ldots, F_D$. The Sklar's theorem states that the multivariate distribution *H* can be expressed as a function *C* on the unit *D*-cube $[0, 1]^D$ of its marginal, such that:

$$H(\mathbf{x}) = C\left(F_1(x_1),...,F_D(x_D)\right) \tag{B.1}$$

The function C is called a copula and is defined on the set $Range(F_1)\times...\times Range(F_D)$. Hence, if $Range(F_d)$ is continuous on [0, 1] ∀ $d = 1,\ldots,D$, then the copula is unique. In short, the copula is invariant under strictly increasing transformation of the margins.

B.3 Generation of random vector

The most interesting part is the converse of Sklar's theorem (Figure B.1), which states that it is possible to link any group of univariate distributions with a copula defining the inter-dependence of the variables in order to define a valid multivariate distribution.

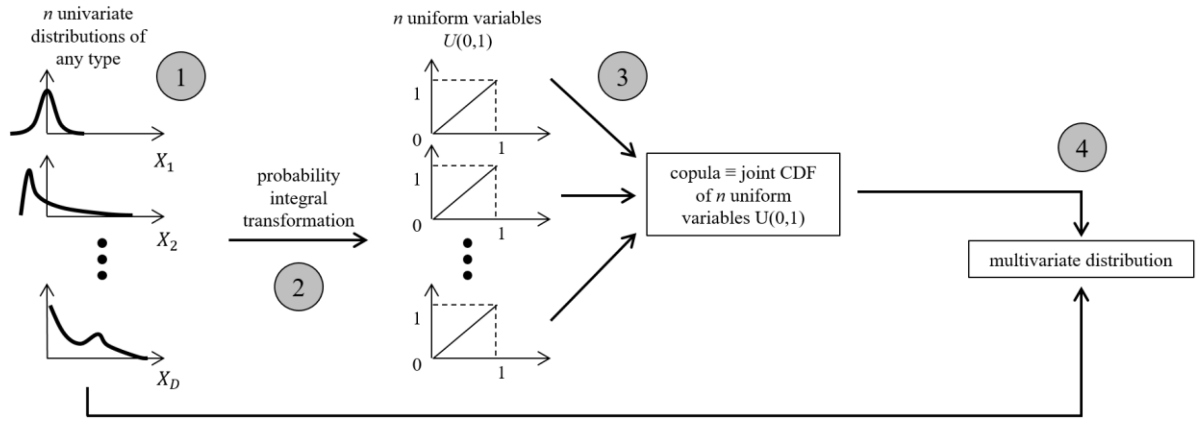


Figure B.1 – Illustration of the converse of Sklar's theorem.

Together with the marginal distributions that encompass the information of the variables taken individually, copulas allow accurately modeling dependent random variates. In this way, if $F_d(x_d)$, $d = 1,\ldots,D$ are continuous distribution functions, then C is unique and given by:

$$C(\mathbf{u}) = H\left(F_1^{-1}(u_1),...,F_D^{-1}(u_D)\right) \tag{B.2}$$

where $u_d = F_d(x_d)$, $d = 1,\ldots,D$ are the probability integral transformations of the margins.

The **probability integral transform** states that if a continuous random variable $\boldsymbol{X}$ has a cumulative distribution function F_X, then the variable $Y = F_X(X)$ is uniformly distributed on [0,1]. The theorem is illustrated in Figure B.1 for a Weibull distribution with shape and scale parameters respectively equal to 5 and 2.

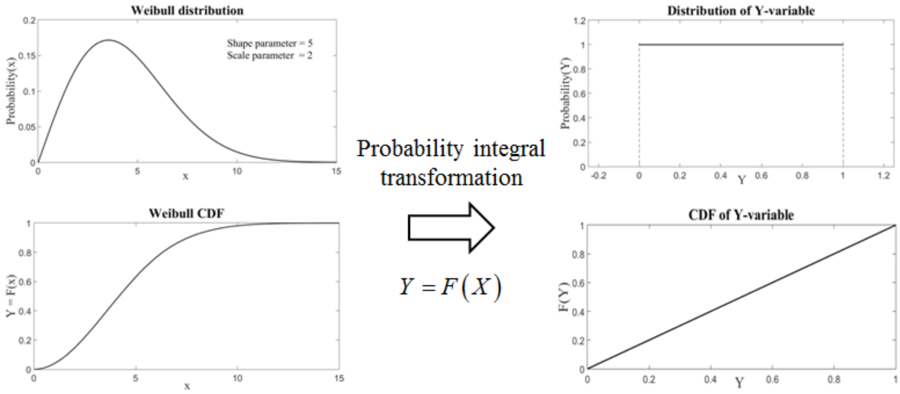


Figure B.2 – Illustration of the probability integral transform.

It should be noted that $F_d^{-1}(u_d)$ is computed using the inverse transform sampling which is illustrated in figure 3, as follows:

$$F_d^{-1}(u_d) = \begin{cases} \inf\{x \mid F_d(x) \geq u_d\} & u_d > 0 \\ \sup\{x \mid F_d(x) = 0\} & u_d = 0 \end{cases} \tag{B.3}$$

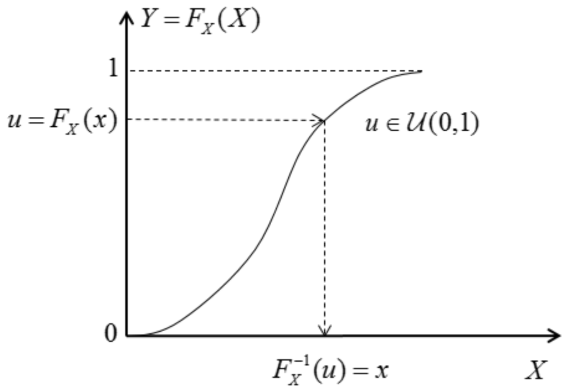


Figure B.3 – Illustration of the inverse transform sampling.

B.4 Determination of copulas

The estimation of copulas can be achieved parametrically by assuming models for both the copulas and the marginal. In this way, several parametric models of copulas are developed such as the Gaussian copula or the Archimedean copulas [Nelsen[98]]. In such cases, the considered copula is fitted to the data and the parameters are computed using a maximum likelihood approach. This method can be computationally cumbersome for high-dimensional distributions because the marginal and copula parameters must be estimated together, and is not used in practice. Then, semiparametric estimations are also of common use and specify a parametric copula with empirical marginal.

However, it should be mentioned that standard parametric copulas are typically suitable for bivariate models but suffer from rather inflexible structures that are difficult to generalize for higher-dimensional distribution. Moreover, such models do not allow to consider different dependency structures between the different pairs of variables, which makes them unsuitable for high-dimensional problem [Brechmann[13]]. Consequently, parametric copulas are not used for accurately modelling the dependence among a large number of variables.

To that end, vine copulas were developed as the enable the extension to arbitrary dimensions [Chollete[09]]. The principle is to represent a density $f(x_1, \dots, x_D)$ as a product of pair copula densities and marginal densities. In other words, the dependency structure is determined by decomposing a multivariate probability density into a cascade of bivariate copulas, where each pair-copula can be chosen independently from the others. Statistical inference (maximum-likelihood, Bayesian approach ...) is used to fit the bivariate copulas to the empirical data, which allows a great flexibility in dependence modeling. The decomposition is therefore not unique and a graphical structure called regular vine structure has been introduced in [Bedford[01]] in order to help organize them.

Although there is a rich variety of copula families, such vine copulas intrinsically generates a succession of approximations due to the need of assigning each pair-copulas to an existing parametric family, which may be problematic in high-dimensional problems. Henceforth, along with those studies, full nonparametric approaches were also investigated. These present many assets if the number of available data for constructing the empirical model is sufficient. Indeed, a nonparametric estimation of copula treats both the copula and the marginal parameter-free, and therefore offer a greater generality allowing to represent any type of dependence